Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 24 (2022), Issue 8 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Viewpoint

Research Letter

Retraction

Review

Methodologies Used to Study the Feasibility, Usability, Efficacy, and Effectiveness of Social Robots For Elderly Adults: Scoping Review

Aysan Mahmoudi Asl[1,2], MSc; Mauricio Molinari Ulate[1,2], MSc; Manuel Franco Martin[2,3], MD; Henriëtte van der Roest[4], PhD

[1]Department of Research and Development, Iberian Institute of Research in Psycho-Sciences, INTRAS Foundation, Zamora, Spain

[2]Psycho-Sciences Research Group, Salamanca Biomedical Research Institute, Salamanca University, Salamanca, Spain

[3]Psychiatry and Mental Health Service, Assistance Complex of Zamora, Zamora, Spain

[4]Department on Aging, Netherlands Institute of Mental Health and Addiction, Trimbos Insititute, Utrecht, Netherlands

Corresponding Author:
Aysan Mahmoudi Asl, MSc
Department of Research and Development
Iberian Institute of Research in Psycho-Sciences
INTRAS Foundation
C. de la Hiniesta, 137
Zamora, 49024
Spain
Phone: 34 604543985
Email: amahmoudi@ides.es

Abstract

Background: New research fields to design social robots for older people are emerging. By providing support with communication and social interaction, these robots aim to increase quality of life. Because of the decline in functioning due to cognitive impairment in older people, social robots are regarded as promising, especially for people with dementia. Although study outcomes are hopeful, the quality of studies on the effectiveness of social robots for the elderly is still low due to many methodological limitations.

Objective: We aimed to review the methodologies used thus far in studies evaluating the feasibility, usability, efficacy, and effectiveness of social robots in clinical and social settings for elderly people, including persons with dementia.

Methods: Dedicated search strings were developed. Searches in MEDLINE (PubMed), Web of Science, PsycInfo, and CINAHL were performed on August 13, 2020.

Results: In the 33 included papers, 23 different social robots were investigated for their feasibility, usability, efficacy, and effectiveness. A total of 8 (24.2%) studies included elderly persons in the community, 9 (27.3%) included long-term care facility residents, and 16 (48.5%) included people with dementia. Most of the studies had a single aim, of which 7 (21.2%) focused on efficacy and 7 (21.2%) focused on effectiveness. Moreover, forms of randomized controlled trials were the most applied designs. Feasibility and usability were often studied together in mixed methods or experimental designs and were most often studied in individual interventions. Feasibility was often assessed with the Unified Theory of the Acceptance and Use of Technology model. Efficacy and effectiveness studies used a range of psychosocial and cognitive outcome measures. However, the included studies failed to find significant improvements in quality of life, depression, and cognition.

Conclusions: This study identified several shortcomings in methodologies used to evaluate social robots, resulting in ambivalent study findings. To improve the quality of these types of studies, efficacy/effectiveness studies will benefit from appropriate randomized controlled trial designs with large sample sizes and individual intervention sessions. Experimental designs might work best for feasibility and usability studies. For each of the 3 goals (efficacy/effectiveness, feasibility, and usability) we also recommend a mixed method of data collection. Multiple interaction sessions running for at least 1 month might aid researchers in drawing significant results and prove the real long-term impact of social robots.

(J Med Internet Res 2022;24(8):e37434) doi:10.2196/37434

KEYWORDS

aged; dementia; social robots; pet-bots; community settings; long-term care; methods; scoping review

Introduction

In the next few decades, we expect the global population to age due to a combination of high life expectancy, low birth rates, and the baby boomer generation entering their senior years. By 2030, 33% of the population in Western Europe will be over 60 years of age [1]. Dementia is one of the most common neurodegenerative diseases that affects 50 million older people around the world, and it is projected to reach 155 million in 2050 [2].

Dementia is characterized by deterioration in memory, cognition, behavior, and ability to perform everyday activities [3]. It is estimated that approximately one-third of people with dementia live alone [4]. They experience unmet needs because of living alone and are at a higher risk of faster deterioration. In addition, people with dementia who live alone are considered at a higher risk of medication use problems, falls and injuries, inadequate self-care, trouble with activities of daily living, and reduced social networks [5-8].

In the past decades, technological advances coincided with the great health challenge of aging societies [9]. New research fields in assistive technology are dedicated to designing social robots for older adults with or without cognitive impairment to promote their quality of life (QoL) through communication and social interactions [10]. Social robots are intended to provide and facilitate social contact, psychosocial and cognitive stimulation, and have the potential to support elderly people to maintain their autonomy and independence and enhance their well-being [11].

Socially assistive robots (SARs) can be grouped into 2 main categories based on their feature and function: (1) service robots, and (2) companion robots [12]. The main task of these robots is to establish any form of interaction and communication. This function can be performed by SARs in multiple manners, such as through touch sensors, cameras, (robotic) body movements, tablet interfaces, and sound and speech systems. Within the subgroup of the companion robots, humanoid robots like Pepper and Nao provide users with advanced applications that provide leisure activities (music, photos, and games), cognitive and physical stimulation activities, and assistance with mental or physical tasks. Pet robots, such as PARO, AIBO, and NeCoro as substitutes for pets and companion animals are intended to provide emotional and physiological stimulation, have calming effects, and lead to mood improvements [13].

For the successful implementation and large-scale uptake of social robots or any other psychosocial intervention, their feasibility, usability, and cost-effectiveness should be perceived as good by the end users (people with dementia and healthy older adults), clinicians, and other stakeholders (eg, health care insurers and policy makers). The Monitoring and Evaluating Digital Health Interventions framework recommends evaluating 4 factors to integrate and implement a digital health intervention: (1) feasibility, to assesses whether the digital health system works as intended in a given context; (2) usability, to assess whether the digital health system can be used as intended by users; (3) efficacy, to assess whether the digital health intervention can achieve the intended results in a research (controlled) setting, and (4) effectiveness, to assess whether the digital health intervention can achieve the intended results in a nonresearch (uncontrolled) setting [14].

Despite the rising interest in social robots after the COVID-19 pandemic, there is limited evidence on the effectiveness of social robots in elderly care. The methodological quality of studies on the effectiveness of social robots in elderly adults is still low, and inappropriate study designs, samples, form, duration of interventions, and data collection methods have affected the strength of study outcomes [12].

Currently, there is no state-of-the-art proof of concept for study designs to evaluate the use of social robots for elderly people. Since the degenerative nature of dementia can cause methodological challenges, specific attention should be paid to studies that include people with dementia. To determine what the appropriate research methods are to study feasibility, usability, efficacy, and effectiveness, this article aims to review the methodologies used thus far in studies with social robots in clinical and social settings with elderly people to pave the way for future researchers in this field.

Methods

Protocol Registration

The protocol of this scoping review was registered in Open Science Framework (OSF) [15].

Search Strategy

Searches were conducted on August 13, 2020, in MEDLINE (PubMed), Web of Science, PsycInfo, and CINAHL databases. No time window was applied. Three search strings covering the topics “social robots,” “community setting,” and “elderly people” were constructed. For each database, reverent subject headings were adapted. For MEDLINE, we used the following strings and keywords: ((robotics[MeSH Terms] OR robot*)) AND ((humanoid OR companion OR social* OR “socially assistive” OR interact* OR android)), (((((((aging[MeSH Terms]) OR (aged[MeSH Terms])) OR (elderly[MeSH Terms])) OR (vulnerable population[MeSH Terms])) OR (senior)) OR (ageing)) OR (geriatric)) OR (old*), and (((((community health service[MeSH Terms]) OR (social support[MeSH Terms])) OR (residential facilities[MeSH Terms])) OR (independent living[MeSH Terms])) OR (social support[MeSH Major Topic])) OR (“community dwelling” OR “home dwelling” OR “care home” OR “in-home” OR “at home” OR “home-based” OR “home setting” OR “nursing home” OR home).

Selection Criteria

Publications potentially eligible for inclusion had to study a social robot that was physically embedded in an experimental or clinical study in people aged 65 or above. Studies were excluded if they were (1) conducted in an acute care setting;

(2) conference abstracts, case studies, dissertations, books, or review papers; (3) published in a language other than English or Spanish; (4) solely reporting a robot design, development process, or theoretical models (5) stakeholder opinions on robots without any interaction; (6) involved in the implementation of new hardware or software or an assessment tool on a robot (such as assessing a fall detection sensor); and (7) involving telepresence robots with only video call features.

Selection Procedure

After the removal of duplicates, 2 reviewers (authors AM and MM) independently applied the inclusion and exclusion criteria in 3 steps, starting with screening titles, abstracts, and then full texts. The selections were compared, and in case of disagreement, discussed by the 2 reviewers. In cases where no consensus could be reached, a third reviewer was consulted (author HR).

Data Extraction

The literature was mapped according to the following areas of interest: author and country, robot name, the aim of the robot, aim of the study, type of outcome measure, study design, study sample, study setting, methodology of data collection, interaction scenario, relevant outcome measures, measurement instruments, results, and reported limitations of the study. Information was synthesized descriptively, and findings were narratively summarized according to the areas of interest.

The quality of the included articles was appraised independently by 2 authors (AM and MM), through the quality assessment of digital health interventions within the Monitoring and Evaluating Digital Health Interventions framework established by the World Health Organization (WHO) [14]. The tool includes a list of methodological study criteria comprising (1) 23 essential criteria for all types of studies and (2) essential criteria for qualitative and quantitative studies (3 criteria each). The extent to which criteria were met by studies was rated by 3 independent researchers on a 3-point scale (0= poor, 1= fair, 2= good). We calculated the percentage of agreement between the ratings (Multimedia Appendix 1). We also applied this framework to categorize the studies according to their aims in 4 categories: (1) feasibility, (2) usability, (3) efficacy, and (4) effectiveness.

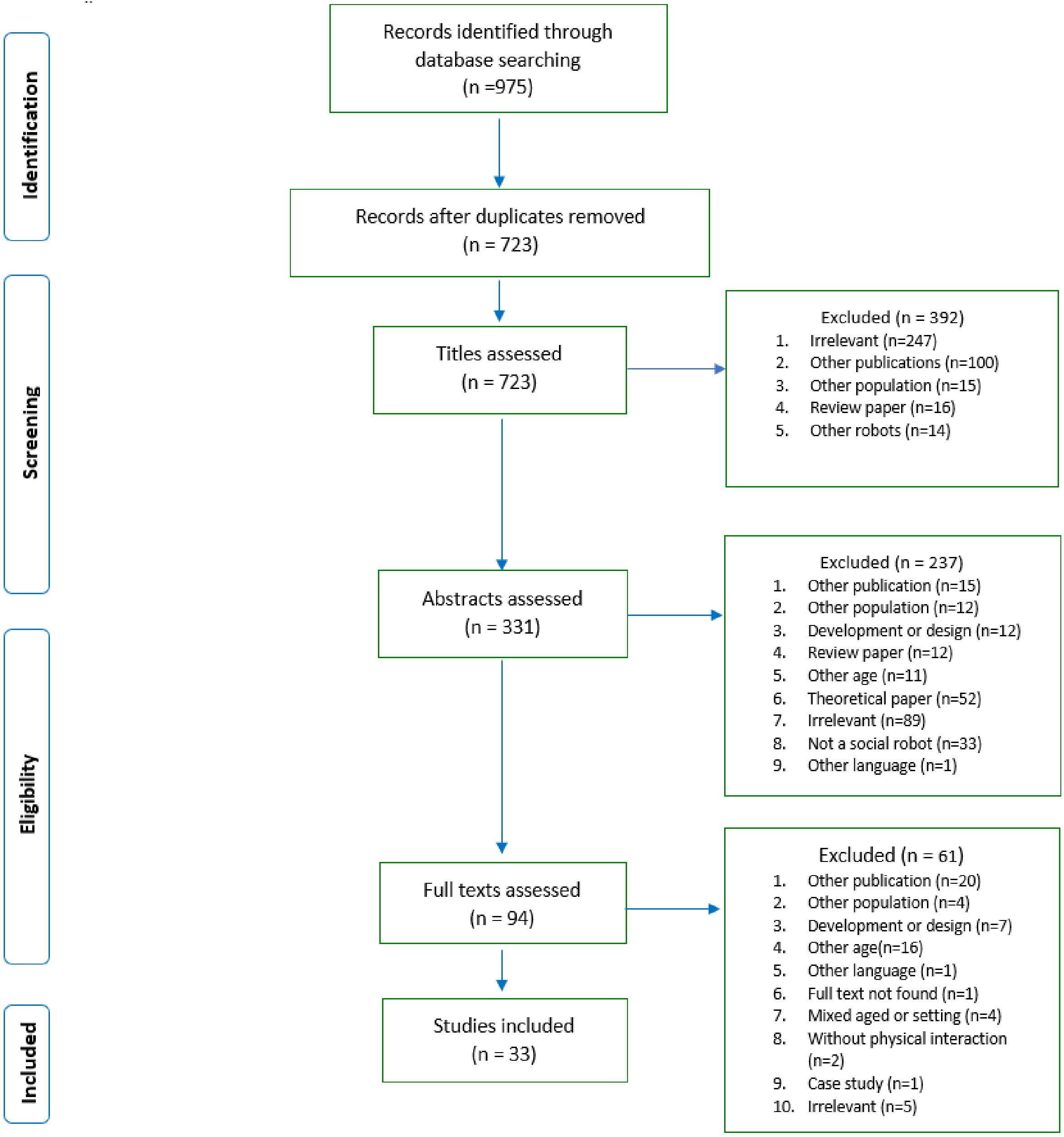
Results

General Findings

The search resulted in a total of 723 individual publications. After the screening process, 33 articles met the selection criteria (Figure 1). In 33 papers [11,16-45], 23 different social robots were evaluated among elderly adults and people with dementia in 13 different countries. Moreover, 19 studies specifically evaluated either feasibility, usability, efficacy, or effectiveness and were considered as single aim studies. The remaining studies (n=14) had multiple aims, evaluating 2 or 3 of the aforementioned study aims. Overall, feasibility was studied in 17 (51.5%) studies, usability in 13 (39.3%), effectiveness in 12 (36.3%), and efficacy in 10 (30.3%).

The quality appraisal identified that primary and secondary outcomes were clearly defined in all studies. Additionally, the methods of data collection were described well, but the eligibility of the participants was not reported in 12 (36.4%) papers. Moreover, 12 out of 33 (36.4%) papers did not present a clear description of the study design. Multimedia Appendices 2 and 3 show a summary of the characteristics, methodologies, and outcomes of the included studies.

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



Robots

Of the 23 different social robots, PARO (n=10, 30.3%), Nao (n=5, 15.1%), AIBO (n=2, 6%), and Hobbit (n=4, 12.1%) were the most often investigated.

Participants and Settings

Of the 33 identified studies, 16 (48.4%) focused on people with dementia [11,18,23,24,27,29-35,38,39,43,44], 9 (27.3%) were performed in samples of residents of long-term care facilities whose cognitive status was not mentioned [17,20,22,26,28,36,40,41,45], and the remaining 8 (24.2%) focused on elderly people living in the community whose cognitive status was also not clearly revealed [16,19,22,25,37,42,46,47]. Moreover, 4 (12.1%) studies additionally recruited care staff [22,24,26,38]. The age range of older adults was 65-98 years. Of the included studies, 3 (9.1%) did not report the number of participants [17,20,26]. The sample sizes used in the studies ranged from 5 to 139.

The social robots were studied in long-term facilities (n=17, 51.5%), private households (n=8, 24.2%), and laboratory settings (n=4, 12.1%). Additional settings were based in a care organization (n=1, 3%), a daycare center for dementia (n=1, 3%), and a health service facility (n=1, 3%). Four (12.1%) studies investigated the robots in 2 different settings [22,33,43,46].

Study Aims, Designs, and Outcome Measures

Single Aim Studies

Of the included studies, 3 (9.1%) focused solely on feasibility, using quasi-experimental designs [22,27,46], and 1 (3%) explicitly focused on usability in a private home setting [37]. Additionally, 7 (21.2%) studies aimed at studying efficacy [31-35,38,44], of which 2 (28.5%) applied a form of randomized controlled trial (RCT) design, 1 (14.3%) randomized crossover design, 1 (14.3%) pretest-posttest design, and the other 3 (42.8%) a form of quasi-experimental design. The effectiveness of the robots was explicitly studied in 8 (24.2%) articles [17,20,24,36,40,41,43,45] using randomized designs, with 1 (12.5%) RCT, 2 (25%) blocked RCTs, 2 (25%) quasi-experimental designs, 1 (12.5%) pretest-posttest design, 1 (12.5%) cross-sectional, and 1 (12.5%) qualitative study. The impact of robots was evaluated on QoL (n=6, 18.2%), mood and depression (n=6, 18.2%), behavioral (n=6, 18.2%) and neuropsychiatric symptoms (n=2, 6.1%), emotions and affect (n=5, 15.2%), cognition (n=4, 12,1%), engagement (n=8, 24.2%), participation and social interaction (n=8, 24.2%), care burden (n=1, 3%), loneliness (n=1, 3%), and physiological indicators (n=3, 9.1%). The sample size for the 22 studies with effectiveness/efficacy aims ranged from 11 to 139 participants, and 7 (31.8%) of these studies included samples of over 40 participants.

Multiple Aim Studies

A total of 14 (42.4%) studies had multiple aims. Of these, 7 (50%) focused on feasibility and usability [11,16,19,22,25,28,44], of which 3 (42.9%) applied a mixed methods design, and the remaining 4 (57.1%) applied either an experimental design or a field trial. Meanwhile, 3 (9.1%) focused on feasibility, usability, and effectiveness, and all applied a mixed methods design [26,29,39]. Additionally, 1 (3%) study investigated feasibility and efficacy and applied an experimental design [30], 1 (3%) focused on feasibility, usability, and efficacy using a pretest-posttest design [18], and 1 (3%) assessed the feasibility and effectiveness of the robot, applying a nonrandomized controlled trial design [22].

Study Aims and Settings

Only 5 (27.8%) of the 18 studies aiming to evaluate feasibility and/or usability were performed in nursing home settings; 5 (27.8%) were performed in laboratory settings, and the remaining 8 (44.4%) were performed in private households. In 7 (38.9%) of the 18 studies, people with dementia and those with cognitive impairment were included. In the remaining 11 (61.1%) studies, the cognitive status of the participants was not clearly indicated.

Of the 22 (66.7%) studies that focused on efficacy or effectiveness, all but 4 (81.8%) [29,33,39,47] were performed in long-term care settings. These 4 (18.2%) were performed in private households and a daycare facility. Of these studies, 13 (59.1%) included cognitively impaired samples, only 1 (4.5%) study included community-dwelling elderly persons without disclosing their cognitive status, and the remaining 8 (36.4%) included long-term care residents.

Study Interventions

Interaction between study participants and social robots was mostly investigated during individual sessions (n=18, 54.5%). In 12 (36.4%) studies, interactions were studied in group sessions. Only 3 (9.1%) studies applied both individual and group interactions [11,20,33], while 1 (3%) demonstrated the task performance of the robot without any close interaction with study participants [19]. Feasibility and/or usability (n=11, 33.3%) were mostly studied in individual settings (n = 8, 72.7%); 1 (9.1%) study was performed in a group, and 1 (9.1%) was applied to both individual and group settings. Individual (n=7, 21.2%) and group settings (n=6, 18.2%) were used most often to study efficacy and effectiveness (n=15, 45.5%); 2 (6.1%) studies applied the intervention individually and in a group. In studies with multiple aims, 4 (28.6%) individual and 2 (14.3%) group setting interventions were found. In 2 (6.1%) studies, social robots were available in the residents' lounge in nursing homes, and participants were free to interact with the social robots during scheduled time slots [22,45]. In 8 (24.2%) studies, the robots were installed in participants' private homes for a duration of 5 days to 3 months [16,27,29,33,37,39,44,47].

A total of 9 (27.3%) studies executed 1 or 2 interactive sessions [11,19,22-24,28,30,32,34], of which 6 (66.7%) investigated the usability and feasibility of the robot, 1 (11.1%) investigated effectiveness, and 2 (22.2%) investigated efficacy. Most of the studies conducted more than 2 interactive sessions: 5 (15.2%) studied feasibility and/or usability, 12 (36.4%) studied efficacy or effectiveness, and 8 (24.2%) were multiple aim studies. The interactive sessions ran from 10 to 90 minutes a day for a maximum of 4 months.

Data Collection

We identified 4 methods of data collection: (1) questionnaires (n=26, 78.8%), (2) observations (physical and videotape) (n=19, 57.6%), (3) interviews (n=13, 39.4%), and physiological measurements (n=3, 9.1%). Figures 2 and 3 show the data collection methods and the responsible administrator of data for the identified data collection methods, respectively.

Figure 2. Used methods of data collection for single and multiple aim studies.

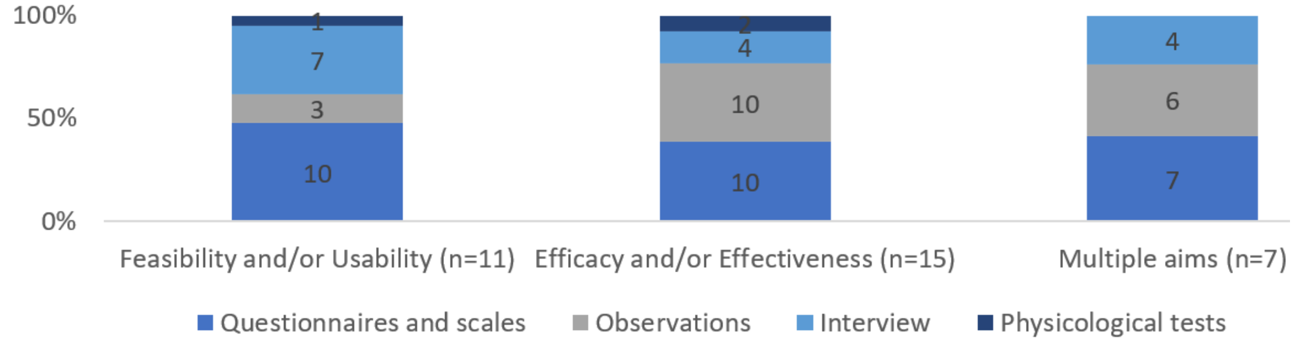
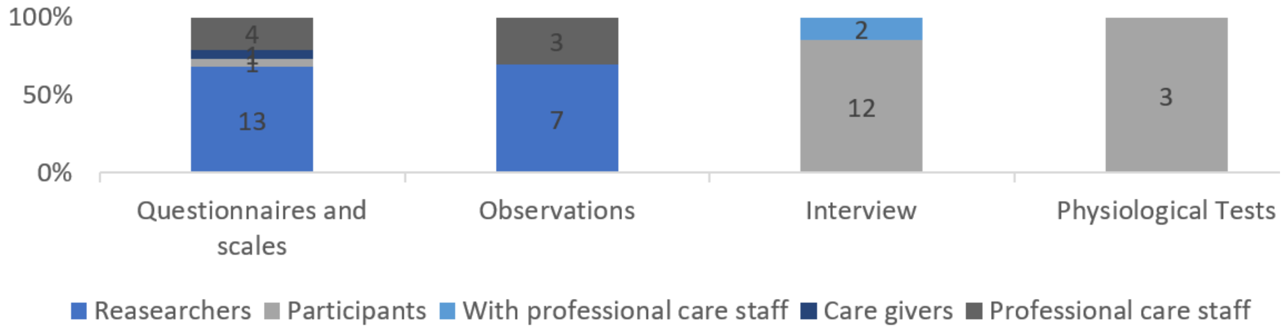


Figure 3. Responsible administrator of data for identified methods of data collection.



Measurement Instruments

Outcomes regarding feasibility were assessed with the Unified Theory of the Acceptance and Use of Technology (UTAUT) model adapted by Heerink [11,19,23,29,30,46,48], the Robot User Acceptance Scale, the Robot Attitude Scale, the Mind Perception Scale [21,47], and the Negative Attitudes Toward Robots Scale [16,37]. Of the included studies, 5 (15.2%) utilized questionnaires regarding robot functions and acceptance that were specifically developed for the study [19,22,23,25,46].

Studies exploring usability applied the System Usability Scale [22,25,28], a modification of the Usefulness, Satisfaction, and Ease of Use [26] scale, and the Technology Usage Inventory [39]. Two (6.1%) qualitative studies performed conversation and video analysis [27,28] to extract statements on acceptability and usability.

Efficacy and effectiveness outcomes were evaluated by a wide range of neuropsychosocial measurement instruments: (1) mood: Geriatric Depression Screening [49], the Cornell Scale For Depression in Dementia [50], Apparent Emotion Rating Instrument [51], University of California, Los Angeles (UCLA) Loneliness Scale [52], Observed Emotion Rating Scale [53]; (2) cognition: Montreal Cognitive Assessment [54], Mini-Mental State Examination [55]; (3) QoL: QoL Alzheimer Disease, Dementia QoL Questionnaire [56], QoL in Late-Stage Dementia [57]; (4) behavior: Neuropsychiatric Inventory [58], Gottfries-Bråne-Steen Scale [59], Apathy Evaluation Scale [60], Cohen-Mansfield Agitation Inventory [61], Apathy Inventory [62], Apathy Scale for Institutionalized People With Dementia Nursing Home Version [63]; (5) Participation and Interaction: Activity Participation Scale [40], and Assessments of Communication and Interaction Skills [64].

Among the studies applying questionnaires, 5 (19.2%) indirectly collected data via care staff and informal caregivers, and 13 (50%) directly collected data via the researchers.

Study Outcomes

Concerning social robots' feasibility outcomes, almost all studies (n=16, 94.1%) deemed social robots acceptable. Nevertheless, 1 (5.9%) study reported mixed results on acceptability by care staff [22], and 1 (5.9%) did not find any significant results on quantitative measurements for acceptability, but qualitative results were positive [16]. In 3 (17.6%) studies, the perceived agency [21,47] and perceived enjoyment [46] were found to decrease over time.

The reported usability (n=12, 36.4%) was overall positive, except in 2 (16.7%) studies in which the usability was negatively affected by technical issues or lack of robustness of the robots [28,37]. Only 3 (9.1%) studies assessed affordability for Hobbit and Nao, in which the participants did not consider the social robots affordable and were skeptical of buying them [25,28,37].

Most of the findings endorse the use of social robots by older adults. Improvements were mostly found in emotion and mood [20,31,33-35,38,44], engagement [24,29,30], and participation and social interaction [20,31-33,40,45]. Increased job satisfaction of staff [22], self-report pain reduction [38], and improved global psychiatric symptoms [43] were the other positive study outcomes. There were findings of reduced challenging behavior [20,31,34], sense of loneliness [17], and stress levels [45]. However, dementia symptoms like agitation and other problematic symptoms did not improve in 1 (3%) study [33].

Meanwhile, 4 (12.1%) studies did not find a significant impact on QoL [19,21,43,47]. Only 1 (3%) study found a moderate-to-large positive effect on QoL of people with dementia [35]. Social robot interventions also failed to significantly improve depression [18,21,43,47], perceived social support [18], medication adherence [47], and cognition [24]. There were mixed results regarding physiological measures, such as urine tests measuring stress levels and blood pressure [33,45]. No author declared a proven negative effect of social robots on older adults.

Reported Study Limitations

Of the 33 studies, 7 (21.2%) did not report any study limitations [16,17,23,25,30,42,46]. A wide range of limitations was reported, and the most common barrier considered in 17 (51.5%) studies was the small sample size, which was mostly reported for efficacy and effectiveness studies. In the feasibility and/or usability studies, the limitations were mainly attributed to technical problems or interaction difficulties. The use of unvalidated questionnaires, homogeneity in the sex of the study sample, inadequate observation, and short duration of interventions were reported as other limitations in general.

Quality Appraisal

The inter-rater agreement for the quality appraisal was 86.1%. Reports of the description of study design, bias, and enrollment procedure were mostly rated as "fair." In most of the articles, the sampling methods, confounding factors, missing data in quantitative studies, and reflexivity of data interpretation in qualitative studies were poorly reported. Other criteria were mostly rated as "good" (Multimedia Appendix 1). The quality appraisal revealed unclear descriptions or insufficient details in 5 (15.2%) studies, especially those in disciplines other than health research [25,30,42,44,45].

Discussion

Principal Findings

The results of this scoping review revealed a variety of applied study methods in studies with social robots concerning study design, sample size, study setting, method of data collection, interaction scenario (the sequence, duration, and setting of the intervention), outcome measures, measurement instruments, study results, and reported limitations. Feasibility and usability were mainly studied on preprototype social robots in laboratory settings. Considering the relatively short history of the use of social robots in psychosocial interventions, it is crucial to determine the main features and functions of the robots to be considered in the design and development phase. Hence, usability, feasibility, and implementation should be strategic research aims. Fully developed robots such as PARO were evaluated in terms of effectiveness in real-world settings. Most of the identified studies aimed to determine the neuropsychosocial impact of social robots on older adults.

For the studies that explicitly fall within a feasibility and/or usability evaluation, researchers applied experimental, mixed method, and field trial designs, mostly applied outside nursing home care settings. This might imply that feasibility and/or usability for persons that are more severely cognitively impaired are currently understudied. Most of the studies verified the acceptability and usability of the robots within single or multiple interactive sessions in individual or group settings, and all these studies reported positive outcomes in varying degrees on the feasibility and/or usability of the social robots. The quantitative and qualitative data were collected mostly through questionnaires and interviews and a few by direct observation. Regarding this point, researchers should consider using the direct observational methodology to capture main factors of the interaction and emotional relationships fostered by robot use. Within the questionnaires and interview questions based on the UTAUT model, some concepts such as trust, anxiety, perceived enjoyment, and social support can change over time [37,46,47]. Therefore, longer use of the robots might reveal these changes and reduce the novelty effect over time [46].

Overall, efficacy and effectiveness studies were conducted on study populations either with cognitive impairment or residing in long-term care facilities. The studies with significant results [17,24,29-31,34-36,45] mostly employed experimental designs including RCTs and quasi-experimental designs with larger sample sizes and longer intervention periods compared to studies showing slight or no improvements. RCTs are likely to be the most appropriate design and a gold standard to confidently demonstrate that a specific intervention has resulted in a change in a process or a health outcome [14]. Biased assessment of outcomes and any confounding effects can be avoidable by large-scale RCTs. However, due to the degenerative character of dementia and personal differences in capacities of people with dementia, difficulties in randomizing subjects often arise [14]. Additionally, when using long study periods, the dropout rate might be high, as participants' cognitive deterioration can hinder their continued participation in the study. On the other hand, when it is not feasible or ethical to conduct an RCT, a quasi-experimental design may serve best for collecting quantitative data. We also recommend randomized controlled block designs in the case of heterogeneous study samples. When, for instance, including people with dementia in studies with long intervention periods, the dementia levels alter. With a randomized controlled block design, some variables in different blocks can be controlled for, or comparable approaches can be applied within the blocks.

Studies targeting the efficacy and effectiveness of the robots delivered interventions diverged in format, duration, dosage, and location. Two (6.1%) studies [32,38] highlighted a need for individual intervention sessions tailored to users' preferences and capacities, and the findings of another study confirmed this approach [65]. Additionally, individualized sessions could omit confounding factors by reducing the chance of interactions with the facilitator or other participants [66]. Group interventions may lower the odds of interaction between potential users and the robot, potentially lowering the effect of the intervention, especially when the intervention is delivered in a noisy setting with many participants [18]. The issues of small sample sizes and short interactions were considered major limitations in studies that failed to find significant results, and they are considered the toughest challenges for researchers in this field [66,67]. These limitations are often cojoined with the study setting. In nursing homes, a larger number of residents are often enrolled in a clinical trial, and the robots are not personalized but must be shared by the entire group. Whereas in private homes, studies are conducted with individuals or dyads, which creates better possibilities for personalization of the robot. Overall, the personalization of the intervention and alleviation of loneliness are 2 advantages of home-based clinical trials. However, there are some challenges to these types of studies, such as the need for several robots, implementation difficulty, and personalization, but it is nevertheless a step in the right direction. We observe a paradox, in that new or experimental robots are employed in research with low numbers of participants, whereas commercially available robots are tested on large study samples. Commercialization allows for better testing and evaluation, which is logical from an economical perspective. However, we urge that before robots are marketed, developers should study the feasibility and usability appropriately in the target group, as well as the effectiveness in a substantial study sample with sufficient power. After bringing the robot to the market, producers should continue to invest in studies to improve their product to tailor it optimally to their users. This should be a joined mission of producers and policy makers to improve the sustainability of health care systems.

Apart from the aforementioned limitations of the studies, some weak aspects of the study designs led to failure of the social robots' impacts. For instance, a mismatch between the studied construct and the main aim of the robot may lead to the poor conclusion that the robot is not efficient. An example of this is the studies on PARO that failed to demonstrate significant results for cognition, as PARO is not developed to stimulate cognitive functioning [31,33,41,43]. Additionally, to capture significant results in constructs such as cognition, QoL, and depression, a long intervention period is necessary because these are constructs that do not change very quickly. In studies with people with dementia, it might also be useful to study stability of these constructs instead of improvement, since it is inherent to the disease that these constructs deteriorate over time. Regarding the broad concept of QoL ranging from physical health to psychological state and social relationships, the application of a suitable QoL measurement instrument that corresponds to the robot's aim should be taken into account.

Implications For Efficacy and Effectiveness Studies

Appropriate RCTs with large sample sizes and individual interaction sessions running for longer than 1 month would serve best for such studies to draw relatively robust and reliable results.

Implications For Feasibility and Usability Studies

The study methods are similar for both aims, so researchers could apply the same design. Experimental designs with mixed methods of data collection are recommended for these studies. Multiple interaction sessions might reveal the changes in feasibility and usability.

Implication For Studies With Multiple Aims

We recommend performing separate studies for multiple aims since the study designs for each aim are comparable.

We gathered further practical recommendations through which future work may address existing shortcomings (Table 1). Regarding the mixed methods of data collection, studies suggest a combination of qualitative and quantitative methods for data collection, which will enable the researcher to capture different details in users' responses and address different aspects of the research question. A mixed methods approach was helpful in studies that could not derive positive results from quantitative data but did from qualitative data [16]. Regarding the difficulties of recruiting many users in case of availability of just a few robots, these mixed methods should be mandatory. Even though we did not find any negative results regarding the intervention dosage, there are shreds of evidence of highly intense intervention resulting in negative effects or exhaustion [18]. Hence, the dose response for specific measures remain an open question for future researchers.

Table 1. Further implications and recommendations for future studies.

Area of consideration	Type of study	Recommendation
Participant and setting	Efficacy/effectiveness	• Gender homogeneity • Different levels of dementia • Realistic environments
Intervention and data collection	Efficacy/effectiveness	• Multiple intervention sessions for longer than 1 month
	Feasibility/usability	• Initialization phase before trial
	All types of studies	• Well-trained observers and professionals • Tailored interventions • Include an intervention facilitator apart from an observer • Consistent observations • Standardized and validated measurement instruments • A client-centered approach to intervention design • A combination of qualitative and quantitative methods of data collection • Observational study when including people with severe dementia
Gap in the existing literature to be filled	Efficacy/effectiveness	• Best response-dosage of intervention for particular measure and participant condition • Characteristics of subjects who benefit most from the social robots

Limitations and Strengths

Although the use of social robots is promising to support people with dementia, we did not include dementia specifically in the search strings, since this scoping review focused on elderly people in general. However, we believe that our search captured most of the studies executed in the field of dementia because many of the identified studies included people with dementia in either mixed or specific study samples. However, some relevant studies on elderly people with dementia may be missing in this review, as well as may studies that are only traceable in databases that were not taken into account in this review. The searches were conducted in scientific databases deemed the most viable to retrieve valid and reliable results for this scoping review. The exclusion of studies focusing only on the development phase of social robots can be considered a limitation of this study. Some information on the evaluation of the feasibility and usability executed in the development stage might have been missed. In addition, studies on telepresence robots were excluded due to their relatively simple features. Compared to pet robots and humanoid robots, telepresence robots are limited in interactions with users, which occur merely through a touch screen, making use of visual and audio stimuli but omitting other sensory stimulation. Although mainly used for medical visits, some telepresence robots might support social functioning. Information on studies performed on these robots might have been missed.

Our study is the first scoping review on the methodologies for studying social robots in elderly people and people with dementia. The existing reviews on this topic mostly focus on design, use, effectiveness, facilitators, and barriers to the implementation of social robots [12,66,67,68-73]. This study might support future researchers to design a research study on social robots in elderly adults and answer some study design queries.

Conclusions

This review narratively synthesizes information on the methodology of studying social robots in elderly adults and people with dementia. Relevant recommendations were formulated, directed for studies with specific aims that may aid future researchers in developing adequate study designs to evaluate social robots, allowing for more reliable information on study outcomes. Our research leads us to the conclusion that more studies with large sample sizes are needed on the effectiveness of social robots in private households of elderly adults and people with dementia to demonstrate the actual usefulness of social robots on delaying institutionalization by improving QoL, cognition, and social contact, and counteracting loneliness. Most of the identified studies focused on usability, and the robots appeared to be favorably accepted in most cases. It is time to encourage investigations in private homes to supplement existing knowledge about the effectiveness of robots and personalization of their functions. We expect that additional research will corroborate the impact of social robots on loneliness, mood, QoL, and social health in people with dementia and the elderly.

Acknowledgments

This research was carried out as part of the Marie Curie Innovative Training Network (ITN) action, H2020-MSCA-ITN-2018, under grant agreement number 813196.

Conflicts of Interest

MFM has participated in clinical trials with Lundbeck, Acadia, Janssen, Otsuka, Angellini, Oxygen, Roche, Servier, Boehringer Ingelheim, and Gw Research Limited.

Multimedia Appendix 1
Quality appraisal of the included studies assessed by two independent reviewers.
[DOCX File , 56 KB - jmir_v24i8e37434_app1.docx]

Multimedia Appendix 2
Characteristics and methodologies of the studies.
[DOCX File , 43 KB - jmir_v24i8e37434_app2.docx]

Multimedia Appendix 3
Outcomes.
[DOCX File , 24 KB - jmir_v24i8e37434_app3.docx]

References

1. World Population Prospects: 2015 Revision. New York, NY: United Nations Department of Economic and Social Affairs; 2015:1-66.
2. Dementia: key facts. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2022-06-17]
3. Wimo A. The global economic impact of dementia. 2010. URL: https://abramsonseniorcare.org/media/1199/observed-emotion-rating-scale.pdf [accessed 2022-06-21]
4. Ebly EM, Hogan DB, Rockwood K. Living alone with dementia. Dement Geriatr Cogn Disord 1999 Nov 4;10(6):541-548. [doi: 10.1159/000017202] [Medline: 10559572]
5. Miranda-Castillo C, Woods B, Orrell M. People with dementia living alone: what are their needs and what kind of support are they receiving? Int Psychogeriatr 2010 Mar 10;22(4):607-617. [doi: 10.1017/s104161021000013x]
6. Webber PA, Fox P, Burnette D. Living alone with Alzheimer's disease: effects on health and social service utilization patterns. Gerontologist 1994 Feb 01;34(1):8-14. [doi: 10.1093/geront/34.1.8] [Medline: 8150314]

7. Meaney A, Croke M, Kirby M. Needs assessment in dementia. Int J Geriatr Psychiatry 2005 Apr;20(4):322-329. [doi: 10.1002/gps.1284] [Medline: 15799078]
8. Wilcox J. Identifying the support needs of people with dementia and older people with mental illness on a Joint Community Team: A preliminary report. J Ment Health 2009 Jul 06;4(2):157-164. [doi: 10.1080/09638239550037695]
9. Perrier E. Positive Disruption: Healthcare, Ageing and Participation in the Age of Technology. Sydney, Australia: The McKell Institute; Sep 2015:2015.
10. Mordoch E, Osterreicher A, Guse L, Roger K, Thompson G. Use of social commitment robots in the care of elderly people with dementia: a literature review. Maturitas 2013 Jan;74(1):14-20. [doi: 10.1016/j.maturitas.2012.10.015] [Medline: 23177981]
11. Fan J, Bian D, Zheng Z, Beuscher L, Newhouse P, Mion L, et al. A robotic coach architecture for elder care (ROCARE) based on multi-user engagement models. IEEE Trans Neural Syst Rehabil Eng 2017 Aug;25(8):1153-1163. [doi: 10.1109/tnsre.2016.2608791]
12. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP. Scoping review on the use of socially assistive robot technology in elderly care. BMJ Open 2018 Feb 12;8(2):e018815 [FREE Full text] [doi: 10.1136/bmjopen-2017-018815] [Medline: 29440212]
13. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc 2013 Sep;14(9):661-667. [doi: 10.1016/j.jamda.2013.02.007] [Medline: 23545466]
14. Mitchell K. Interventions. In: Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Cambridge, UK: Cambridge University Press; 2010:15-22.
15. OSF. 2021. URL: https://osf.io/vpyg2/ [accessed 2022-02-14]
16. Bajones M, Fischinger D, Weiss A, Puente P, Wolf D, Vincze M, et al. Results of field trials with a mobile service robot for older adults in 16 private households. J Hum-Robot Interact 2020 Feb 07;9(2):1-27. [doi: 10.1145/3368554]
17. Banks MR, Willoughby LM, Banks WA. Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. J Am Med Dir Assoc 2008 Mar;9(3):173-177. [doi: 10.1016/j.jamda.2007.11.007] [Medline: 18294600]
18. Barrett E, Burke M, Whelan S, Santorelli A, Oliveira BL, Cavallo F, et al. Evaluation of a companion robot for individuals with dementia: quantitative findings of the MARIO project in an Irish residential care setting. J Gerontol Nurs 2019 Jul 01;45(7):36-45. [doi: 10.3928/00989134-20190531-01] [Medline: 31237660]
19. Beer J, Prakash A, Smarr CA, Chen TL, Hawkins K, Nguyen H, et al. Older users' acceptance of an assistive robot: attitudinal changes following brief exposure. Gerontechnology 2017 Mar 29;16(1):21-36 [FREE Full text] [doi: 10.4017/gt.2017.16.1.003.00] [Medline: 31178671]
20. Birks M, Bodak M, Barlas J, Harwood J, Pether M. Robotic seals as therapeutic tools in an aged care facility: a qualitative study. J Aging Res 2016;2016:8569602-8569607 [FREE Full text] [doi: 10.1155/2016/8569602] [Medline: 27990301]
21. Broadbent E, Kerse N, Peri K, Robinson H, Jayawardena C, Kuo T, et al. Benefits and problems of health-care robots in aged care settings: A comparison trial. Australas J Ageing 2016 Mar;35(1):23-29. [doi: 10.1111/ajag.12190] [Medline: 26364706]
22. Cavallo F, Esposito R, Limosani R, Manzi A, Bevilacqua R, Felici E, et al. Robotic services acceptance in smart environments with older adults: user satisfaction and acceptability study. J Med Internet Res 2018 Sep 21;20(9):e264 [FREE Full text] [doi: 10.2196/jmir.9460] [Medline: 30249588]
23. Cavallo F, Limosani R, Manzi A, Bonaccorsi M, Esposito R, Di Rocco M, et al. Development of a socially believable multi-robot solution from town to home. Cogn Comput 2014 Jul 3;6(4):954-967. [doi: 10.1007/s12559-014-9290-z]
24. Chu M, Khosla R, Khaksar SMS, Nguyen K. Service innovation through social robot engagement to improve dementia care quality. Assist Technol 2016 Apr 11;29(1):8-18. [doi: 10.1080/10400435.2016.1171807]
25. Fischinger D, Einramhof P, Papoutsakis K, Wohlkinger W, Mayer P, Panek P, et al. Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. Rob Auton Syst 2016 Jan;75:60-78. [doi: 10.1016/j.robot.2014.09.029]
26. Huisman C, Kort H. Two-year use of care robot Zora in Dutch nursing homes: an evaluation study. Healthcare (Basel) 2019 Feb 19;7(1):31 [FREE Full text] [doi: 10.3390/healthcare7010031] [Medline: 30791489]
27. Inoue T, Nihei M, Narita T, Onoda M, Ishiwata R, Mamiya I, et al. Field-based development of an information support robot for persons with dementia. TAD 2012 Dec 12;24(4):263-271. [doi: 10.3233/TAD-120357]
28. Olde Keizer RACM, van Velsen L, Moncharmont M, Riche B, Ammour N, Del Signore S, et al. Using socially assistive robots for monitoring and preventing frailty among older adults: a study on usability and user experience challenges. Health Technol 2019 Apr 9;9(4):595-605. [doi: 10.1007/s12553-019-00320-9]
29. Khosla R, Chu M, Khaksar SMS, Nguyen K, Nishida T. Engagement and experience of older people with socially assistive robots in home care. Assist Technol 2021 Mar 04;33(2):57-71. [doi: 10.1080/10400435.2019.1588805] [Medline: 31063044]
30. Khosla R, Nguyen K, Chu M. Human robot engagement and acceptability in residential aged care. Int J Hum-Comput Interact 2016 Dec 27;33(6):510-522. [doi: 10.1080/10447318.2016.1275435]
31. Koh IS, Kang HS. Effects of intervention using PARO on the cognition, emotion, problem behavior, and social interaction of elderly people with dementia. J Korean Acad Community Health Nurs 2018;29(3):300. [doi: 10.12799/jkachn.2018.29.3.300]

32. Kramer SC, Friedmann E, Bernstein PL. Comparison of the effect of human interaction, animal-assisted therapy, and AIBO-assisted therapy on long-term care residents with dementia. Anthrozoös 2015 Apr 28;22(1):43-57. [doi: 10.2752/175303708x390464]
33. Liang A, Piroth I, Robinson H, MacDonald B, Fisher M, Nater UM, et al. A pilot randomized trial of a companion robot for people with dementia living in the community. J Am Med Dir Assoc 2017 Oct 01;18(10):871-878. [doi: 10.1016/j.jamda.2017.05.019] [Medline: 28668664]
34. Libin A, Cohen-Mansfield J. Therapeutic robocat for nursing home residents with dementia: preliminary inquiry. Am J Alzheimers Dis Other Demen 2004 Jun 30;19(2):111-116 [FREE Full text] [doi: 10.1177/153331750401900209] [Medline: 15106392]
35. Moyle W, Cooke M, Beattie E, Jones C, Klein B, Cook G, et al. Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial. J Gerontol Nurs 2013 May;39(5):46-53. [doi: 10.3928/00989134-20130313-03] [Medline: 23506125]
36. Obayashi K, Kodate N, Masuyama S. Enhancing older people’s activity and participation with socially assistive robots: a multicentre quasi-experimental study using the ICF framework. Adv Robot 2018 Oct 14;32(22):1207-1216. [doi: 10.1080/01691864.2018.1528176]
37. Pripfl J, Körtner T, Batko-Klein D, Hebesberger D, Weninger M, Gisinger C. Social service robots to support independent living : Experiences from a field trial. Z Gerontol Geriatr 2016 Jun;49(4):282-287. [doi: 10.1007/s00391-016-1067-4] [Medline: 27220733]
38. Pu L, Moyle W, Jones C. How people with dementia perceive a therapeutic robot called PARO in relation to their pain and mood: A qualitative study. J Clin Nurs 2020 Feb 02;29(3-4):437-446. [doi: 10.1111/jocn.15104] [Medline: 31738463]
39. Schüssler S, Zuschnegg J, Paletta L, Fellner M, Lodron G, Steiner J, et al. Effects of a humanoid socially assistive robot versus tablet training on psychosocial and physical outcomes of persons with dementia: protocol for a mixed methods study. JMIR Res Protoc 2020 Feb 03;9(2):e14927 [FREE Full text] [doi: 10.2196/14927] [Medline: 32022697]
40. Sung H, Chang S, Chin M, Lee W. Robot-assisted therapy for improving social interactions and activity participation among institutionalized older adults: a pilot study. Asia Pac Psychiatry 2015 Mar 01;7(1):1-6. [doi: 10.1111/appy.12131] [Medline: 24692085]
41. Thodberg K, Sørensen L, Videbech P, Poulsen P, Houbak B, Damgaard V, et al. Behavioral responses of nursing home residents to visits from a person with a dog, a robot seal or a toy cat. Anthrozoös 2016 Mar 08;29(1):107-121. [doi: 10.1080/08927936.2015.1089011]
42. Zsiga K, Tóth A, Pilissy T, Péter O, Dénes Z, Fazekas G. Evaluation of a companion robot based on field tests with single older adults in their homes. Assist Technol 2018;30(5):259-266. [doi: 10.1080/10400435.2017.1322158] [Medline: 28628395]
43. Valentí Soler M, Agüera-Ortiz L, Olazarán Rodríguez J, Mendoza Rebolledo C, Pérez Muñoz A, Rodríguez Pérez I, et al. Social robots in advanced dementia. Front Aging Neurosci 2015;7:133 [FREE Full text] [doi: 10.3389/fnagi.2015.00133] [Medline: 26388764]
44. Wada K, Shibata T, Saito T, Tanie K. Robot assisted activity at a health service facility for the aged for ten weeks: An interim report of a long-term experiment. P I MECH ENG I-J SYS 2006 Dec 27;220(8):709-715. [doi: 10.1243/09596518jsce159]
45. Wada K, Shibata T. Social and physiological influences of robot therapy in a care house. IS 2008 May 26;9(2):258-276. [doi: 10.1075/is.9.2.06wad]
46. Torta E, Werner F, Johnson DO, Juola JH, Cuijpers RH, Bazzani M, et al. Evaluation of a small socially-assistive humanoid robot in intelligent homes for the care of the elderly. J Intell Robot Syst 2014;76:57-71. [doi: 10.1007/s10846-013-0019-0]
47. Broadbent E, Peri K, Kerse N, Jayawardena C, Kuo I, Datta C, et al. Robots in older people’s homes to improve medication adherence and quality of life: a randomised cross-over trial. ICSR 2014;8755. [doi: 10.1007/978-3-319-11973-1_7]
48. Heerink M, Kröse B, Evers V, Wielinga B. Assessing acceptance of assistive social agent technology by older adults: the Almere model. Int J of Soc Robotics 2010 Sep 4;2(4):361-375. [doi: 10.1007/s12369-010-0068-5]
49. Yesavage JA, Brink T, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: A preliminary report. J Psychiatr Res 1982 Jan;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4]
50. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell scale for depression in dementia. Biol Psychiatry 1988 Feb;23(3):271-284. [doi: 10.1016/0006-3223(88)90038-8]
51. Snyder M, Ryden M, Shaver P, Wang J, Savik K, Gross C, et al. The apparent emotion rating instrument. Clinical Gerontologist 2008 Oct 11;18(4):17-29. [doi: 10.1300/j018v18n04_03]
52. 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]
53. Lawton MP. Observed Emotion Rating Scale. 1999. URL: https://abramsonseniorcare.org/media/1199/observed-emotion-rating-scale.pdf [accessed 2022-06-21]
54. Nasreddine Z, 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]

55. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
56. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998 May 01;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
57. Weiner MF, Martin-Cook K, Svetlik DA, Saine K, Foster B, Fontaine CS. The quality of life in late-stage dementia (QUALID) scale. J Am Med Dir Assoc 2000;1(3):114-116. [Medline: 12818023]
58. Cummings JL. The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. Neurology 1997 May 01;48(5 Suppl 6):S10-S16. [doi: 10.1212/wnl.48.5_suppl_6.10s] [Medline: 9153155]
59. Bråne G, Gottfries CG, Winblad B. The Gottfries-Bråne-Steen scale: validity, reliability and application in anti-dementia drug trials. Dement Geriatr Cogn Disord 2001;12(1):1-14. [doi: 10.1159/000051230] [Medline: 11125236]
60. Marin RS, Biedrzycki RC, Firinciogullari S. Reliability and validity of the apathy evaluation scale. Psychiatry Research 1991 Aug;38(2):143-162. [doi: 10.1016/0165-1781(91)90040-v]
61. Cohen-Mansfield J. Instruction Manual for the Cohen-Mansfield Agitation Inventory (CMAI). 1991. URL: https://uiowa2.instructure.com/files/22132/download?download_frd=1 [accessed 2022-06-21]
62. Robert P, Clairet S, Benoit M, Koutaich J, Bertogliati C, Tible O, et al. The apathy inventory: assessment of apathy and awareness in Alzheimer's disease, Parkinson's disease and mild cognitive impairment. Int J Geriatr Psychiatry 2002 Dec;17(12):1099-1105. [doi: 10.1002/gps.755] [Medline: 12461757]
63. Agüera-Ortiz L, Gil-Ruiz N, Cruz-Orduña I, Ramos-García I, Osorio RS, Valentí-Soler M, et al. A novel rating scale for the measurement of apathy in institutionalized persons with dementia: the APADEM-NH. Am J Geriatr Psychiatry 2015 Feb;23(2):149-159. [doi: 10.1016/j.jagp.2013.01.079] [Medline: 23871117]
64. Forsyth K, Lai J, Kielhofner G. The Assessment of Communication and Interaction Skills (ACIS): measurement properties. Br J Occup Ther 2016 Nov 05;62(2):69-74. [doi: 10.1177/030802269906200208]
65. Kachouie R, Sedighadeli S, Khosla R, Chu M. Socially assistive robots in elderly care: a mixed-method systematic literature review. Int J Hum-Comput Interact 2014 Apr;30(5):369-393. [doi: 10.1080/10447318.2013.873278]
66. Pu L, Moyle W, Jones C, Todorovic M. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. Gerontologist 2019 Jan 09;59(1):e37-e51. [doi: 10.1093/geront/gny046] [Medline: 29897445]
67. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. J Am Med Dir Assoc 2012 Feb;13(2):114-120.e1. [doi: 10.1016/j.jamda.2010.10.002] [Medline: 21450215]
68. Oliveira R, Arriaga P, Santos FP, Mascarenhas S, Paiva A. Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. Comput Hum Behav 2021 Jan;114:106547. [doi: 10.1016/j.chb.2020.106547]
69. Koh WQ, Ang FXH, Casey D. Impacts of low-cost robotic pets for older adults and people with dementia: scoping review. JMIR Rehabil Assist Technol 2021 Feb 12;8(1):e25340 [FREE Full text] [doi: 10.2196/25340] [Medline: 33497349]
70. Góngora Alonso S, Hamrioui S, de la Torre Díez I, Motta Cruz E, López-Coronado M, Franco M. Social robots for people with aging and dementia: a systematic review of literature. Telemed J E Health 2019 Jul;25(7):533-540. [doi: 10.1089/tmj.2018.0051] [Medline: 30136901]
71. Hirt J, Ballhausen N, Hering A, Kliegel M, Beer T, Meyer G. Social robot interventions for people with dementia: a systematic review on effects and quality of reporting. J Alzheimers Dis 2021;79(2):773-792 [FREE Full text] [doi: 10.3233/JAD-200347] [Medline: 33361589]
72. Ozdemir D, Cibulka J, Stepankova O, Holmerova I. Design and implementation framework of social assistive robotics for people with dementia - a scoping review. Health Technol 2021 Feb 05;11(2):367-378. [doi: 10.1007/s12553-021-00522-0]
73. Koh WQ, Felding SA, Toomey E, Casey D. Barriers and facilitators to the implementation of social robots for older adults and people with dementia: a scoping review protocol. Syst Rev 2021 Feb 05;10(1):49 [FREE Full text] [doi: 10.1186/s13643-021-01598-5] [Medline: 33546772]

Abbreviations

ADL: activities of daily living
ITN: innovative training network
LTCF: long-term care facility
OSF: Open Science Framework
QoL: quality of life
RCT: randomized controlled trial
SAR: socially assistive robot
UCLA: University of California, Los Angeles
UTAUT: Unified Theory of the Acceptance and Use of Technology
WHO: World Health Organization

Edited by R Kukafka; submitted 21.02.22; peer-reviewed by I Guemghar, V Stara; comments to author 28.03.22; revised version received 22.05.22; accepted 08.06.22; published 01.08.22.

Please cite as:
Mahmoudi Asl A, Molinari Ulate M, Franco Martin M, van der Roest H
Methodologies Used to Study the Feasibility, Usability, Efficacy, and Effectiveness of Social Robots For Elderly Adults: Scoping Review
J Med Internet Res 2022;24(8):e37434
URL: https://www.jmir.org/2022/8/e37434
doi:10.2196/37434
PMID:35916695

Review

Digital Health Technologies Enabling Partnerships in Chronic Care Management: Scoping Review

Carolina Wannheden[1], PhD; Matilda Åberg-Wennerholm[1], MSc; Marie Dahlberg[1], MSc; Åsa Revenäs[1,2,3], PhD; Sara Tolf[1], PhD; Elena Eftimovska[1], MD, MSc; Mats Brommels[1], MD, PhD

[1]Medical Management Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden
[2]Division of Physiotherapy, School of Health Care and Social Welfare, Mälardalen University, Västerås, Sweden
[3]Center for Clinical Research, County of Västmanland, Uppsala University, Västerås, Sweden

Corresponding Author:
Carolina Wannheden, PhD
Medical Management Centre
Department of Learning, Informatics, Management and Ethics
Karolinska Institutet
Tomtebodavaegen 18a
Stockholm, 171 77
Sweden
Phone: 46 0852483934
Email: carolina.wannheden@ki.se

Abstract

Background: An increasing number of patients expect and want to play a greater role in their treatment and care decisions. This emphasizes the need to adopt collaborative health care practices, which implies collaboration among interprofessional health care teams and patients, their families, caregivers, and communities. In recent years, digital health technologies that support self-care and collaboration between the community and health care providers (ie, participatory health technologies) have received increasing attention. However, knowledge regarding the features of such technologies that support effective patient-professional partnerships is still limited.

Objective: This study aimed to map and assess published studies on participatory health technologies intended to support partnerships among patients, caregivers, and health care professionals in chronic care, focusing specifically on identifying the main features of these technologies.

Methods: A scoping review covering scientific publications in English between January 2008 and December 2020 was performed. We searched PubMed and Web of Science databases. Peer-reviewed qualitative, quantitative, and mixed methods studies that evaluated digital health technologies for patient-professional partnerships in chronic care settings were included. The data were charted and analyzed thematically. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist was used.

Results: This review included 32 studies, reported in 34 papers. The topic of participatory health technologies experienced a slightly increasing trend across publication years, with most papers originating from the United States and Norway. Diabetes and cardiovascular diseases were the most common conditions addressed. Of the 32 studies, 12 (38%) evaluated the influence of participatory health technologies on partnerships, mostly with positive outcomes, although we also identified how partnership relationships and the nature of collaborative work could be challenged when the roles and expectations between users were unclear. Six common features of participatory health technologies were identified: patient-professional communication, self-monitoring, tailored self-care support, self-care education, care planning, and community forums for peer-to-peer interactions.

Conclusions: Our findings emphasize the importance of clarifying mutual expectations and carefully considering the implications that the introduction of participatory health technologies may have on the work of patients and health care professionals, both individually and in collaboration. A knowledge gap remains regarding the use of participatory health technologies to effectively support patient-professional partnerships in chronic care management.

(J Med Internet Res 2022;24(8):e38980) doi:10.2196/38980

KEYWORDS

participatory health; digital health; eHealth; collaborative care; participatory health informatics; cocare; partnership care management; chronic disease; long-term conditions; mobile phone; scoping review

Introduction

Chronic Care Model

Changes in demographics, disease panorama, and medical technology enabling early diagnoses and effective treatments have led to chronic diseases dominating the disease burden, accounting for alarming increases in health care use and costs [1]. Thirty years ago, Wagner et al [2,3] proposed the chronic care model (CCM), which called for a system of care that goes beyond health care provision by mobilizing supporting resources in the community. The model foresaw that the design and organization of service delivery would need to be adjusted to what was required and made possible by viewing chronic care as a system. Clinical information systems have been identified as important assets, particularly for decision support. In addition, self-management support for *informed and activated patients* has been emphasized [3]. Over the years, the CCM has inspired much of the development of chronic care management, as individuals with chronic conditions need services and support from several providers. Research has shown that more patients expect and want to play greater roles in decisions about their treatment and care and perform self-care more effectively [4-6]. Nevertheless, the transformation of health care practices into effective chronic care systems remains challenging [7].

Collaborative Care Partnerships

The reference to *informed and activated patients* indicates that care arrangements according to the CCM are professionally driven, with health care professionals providing information and guidance to patients [3]. In contrast, patient-centered and person-centered care initiatives, launched at the start of the new millennium, highlighted the need to place patients at the center of their care and to make space for patient preferences in care planning, accomplished through a shared decision-making process [8]. This shift in care philosophy undoubtedly paved the way for patients with chronic conditions to be active rather than activated, which matched the aims of empowering patients and promoting equality in the patient-provider relationship. These movements emphasize the need to adopt collaborative health care practices, which implies collaboration between an interprofessional health care team and patients, their families, caregivers, and communities [9].

Central to this form of collaboration is the acknowledgment of patients as experts in their “experience, feelings, fears, hopes, and desires” [10]. Patient participation in co-design and *slow coproduction* helps strengthen their voices in the design of care services and can lead to improved patient experiences [11]. There is also evidence that when patients express what is important to them and have active roles in designing care, outcomes, including clinical outcomes, will improve [12]. In addition, several randomized controlled trials have shown that engaging patients in symptom monitoring, usually by applying digital technologies, has a positive effect on patient outcomes [13,14]. Symptoms are important not only to alert and guide the diagnostic workup but also to measure treatment effects (especially in severe illnesses) [15]. As patient-professional partnerships based on mutual respect for professional and experiential knowledge can strengthen patients in their self-management and shared decision-making with health care professionals, ultimately leading to improved clinical outcomes, it is worthwhile to study how such a collaborative care partnership can be enhanced.

Participatory Health Technologies

Digital health technologies delivered in real time and in real-world settings offer opportunities to support such partnerships. Participatory health informatics, which emerged as a field around 2008, concerns the use of “information technology as provided through the web, smartphones, or wearables to increase participation of individuals in their care process, and to enable them in self-care and decision-making” [16]. For example, web-based social health networks such as PatientsLikeMe [17] have become powerful tools for patients to share their experiences and learn from each other. Technologies for community support marked the beginning of participatory health technologies, and interest has been increasing in technologies supporting self-care and patient-professional partnerships, which was the focus of this study. In particular, the use of text-based patient-professional communication tools has increased over the past decade [18], supporting self-management and contributing to increased patient participation [19]. Although the use of mobile health apps generally has a positive influence on patient-professional relationships, health care professionals may still be reluctant to use them [20]. In addition, despite these apps’ potential to improve health care delivery to people with chronic conditions, their effects on health outcomes have been found to be inconsistent [19,21]. Thus, there is a need to gain more knowledge about the mechanisms that contribute to effective patient-professional partnerships. Therefore, this study aimed to map and assess published studies on participatory health technologies intended to support partnerships between patients, caregivers, and health care professionals in chronic care, focusing specifically on identifying the main features of these technologies.

Methods

Study Design

A scoping review was considered relevant as our aim was to examine the size, scope, and nature of the available literature on our phenomenon of interest and summarize existing research findings [22]. The review was performed in 5 stages, guided by the Arksey and O’Malley framework [23,24]. A review protocol (available on request) was developed beforehand and continuously updated to ensure consistency and reproducibility. The review team covered multiple areas of relevant expertise, including health informatics, health services research, medical technology management, and medicine. A list of experts in the domains of health care, patient self-care, and digital health was established to be contacted if expert advice was needed. For example, we sought and obtained input on the practical relevance of our research questions.

Stage 1: Identifying the Research Question

The scoping review question was specified by considering the population, intervention, comparison, and outcome [25]. The population of interest was broadly limited to people living with chronic illnesses; the intervention of interest was specified in detail, focusing on digital health technologies that enable partnerships between patients, caregivers, and health care professionals (ie, participatory health technologies). No comparison method was specified, and we aimed to identify all types of outcomes explored in previous studies. We posed the following overarching research question: what is known from the existing literature about participatory health technologies that intend to support partnerships between patients, caregivers, and health care professionals in chronic care? More specific research questions were posed in line with our aim:

- The context of use: At which levels of care are the participatory health technologies used? For which types of chronic conditions are participatory health technologies used? Who are the users of participatory health technologies?
- Evaluation: What study designs are used and what outcomes are measured and reported?
- Features supporting partnerships: What are the main features of participatory health technologies? How do the different features influence partnerships?

Stage 2: Identifying Relevant Studies

The search strategy was developed in consultation with the Karolinska Institutet University Library, following the Peer Review of Electronic Search Strategies guidelines [26]. Search terms were designed to capture papers related to three key concepts: (1) digital health technologies, (2) partnerships between patients, caregivers, and health care professionals, and (3) chronic care management. Searches were performed using the bibliographic databases PubMed and Web of Science, which were considered most relevant in relation to our aim. First, we identified synonymous terms for each key concept and combined them into a search phrase using the Boolean operator OR. We also identified and used relevant Medical Subject Heading terms in PubMed. We then combined the search phrases for the 3 concepts using the Boolean operator AND. The exact search phrases for the 2 databases are presented in Multimedia Appendix 1. Our searches were performed on November 21, 2017, and updated on December 14, 2020. The search results were filtered by language and time span, covering papers in English and Swedish published between January 2008 and December 14, 2020.

Stage 3: Study Selection

The inclusion and exclusion criteria are listed in Table 1. The period of inclusion, from 2008 to 2020, was motivated by the emergence of the term *participatory health informatics* in 2008 [16]. Screening was performed using the open-source platform Rayyan [27]. We specified the labels to be used as reasons for exclusion in the screening process if the inclusion criteria were not met. At the beginning of the screening process, the inclusion criteria were piloted and refined in several iterations until a consensus was reached among all authors. The first screening was performed in late 2017 and early 2018 by EE and MÅW with support from MD, ST, ÅR, and CW; titles and abstracts for each study were screened by at least two of these researchers (blinded). Conflicts were resolved through discussion and, if necessary, by involving the research team. The second screening, following an updated search, was performed in early 2021 by CW and MB who both screened all titles and abstracts and resolved conflicts through discussion. They also screened the reference lists of the included studies to identify additional relevant publications.

Table 1. Inclusion and exclusion criteria.

Criteria	Inclusion	Exclusion
Type of studies	Qualitative, quantitative, and mixed methods studies on the phenomenon published in peer-reviewed journals	Letters, commentaries, editorials, conference abstracts, doctoral theses, or any type of review
Period	January 1, 2008, until December 14, 2020	Before January 1, 2008, and after December 14, 2020
Language	English and Swedish	All other languages
Type of participants	Patients with chronic conditions, defined as a health condition that lasts at least 3 months	Patients who do not have chronic conditions
Phenomenon of interest	Studies that meet all 3 criteria listed below	Studies that do not meet all the 3 criteria listed below
Digital health technology	A digital health technology is defined as software intended for use for *preventive, promotive, curative, rehabilitative, assistive, or palliative care*; this includes categories such as eHealth or mobile health, wearable devices, and telehealth services; the digital health technology should enable processing and exchange of health information between end users using the *internet*	Nondigital services or digital services not specifically intended for medical use; for example, WhatsApp, email, telephone, and SMS text messages are technologies that are not primarily intended for the abovementioned purposes and were thus excluded
Partnership	The digital health technology intends to support collaboration and enables interaction between at least two types of users: patients or caregivers, and health care professionals or allied professionals (eg, pharmacists)	Digital health services for peer-to-peer collaboration between patients or caregivers only, or tools for team collaboration among staff, without patient or caregiver involvement, were excluded; tools that only intended to support self-care or treatment adherence were also excluded
Evaluation	Evaluation results testing the digital health technology in chronic care need to be available	Studies that merely describe the design and development of digital health technologies; evaluation that has not been performed in a real-world setting (eg, heuristic evaluation by experts)

Stage 4: Charting the Data

A data extraction sheet was developed containing bibliometric variables (author, country based on corresponding author affiliation, title, year, and journal), descriptive study variables (study aim, study design, and sample size), and variables based on the research questions (chronic condition, level of care, name and description of the participatory health technology, outcome measures, and evaluation results). All authors were involved in testing and refining the data extraction sheet with a selection of papers. A total of 2 authors per paper extracted and compared these to calibrate the variable definitions and our shared understanding thereof. Thereafter, CW and MB performed the remaining extractions for all the papers.

Stage 5: Collating, Summarizing, and Reporting Results

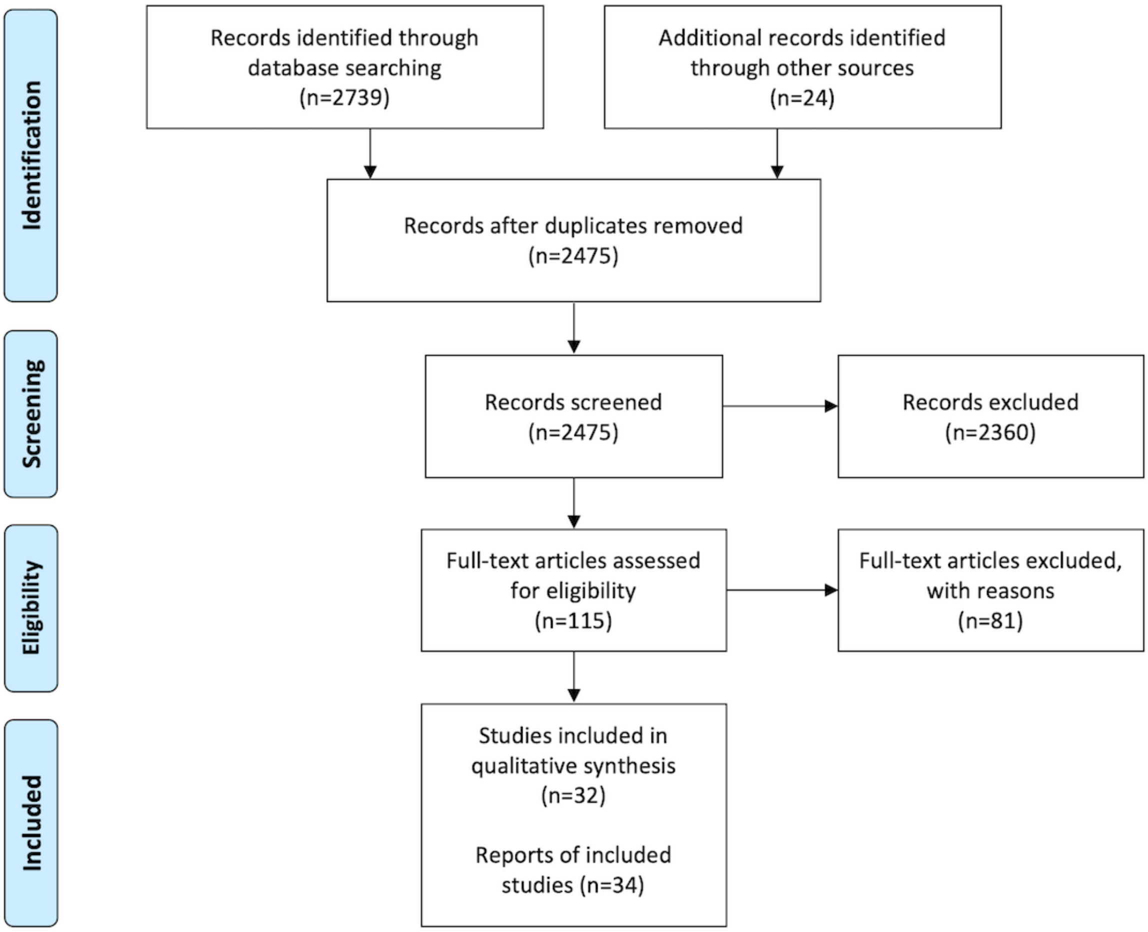
Charted data were condensed and grouped into categories that enabled the classification of the studies based on their study aims, study designs, chronic conditions, outcome measures, and evaluation results. The charted text describing the participatory health technology features and their influences on partnerships was extensive and was, therefore, analyzed separately using a qualitative content analysis process [28]. The charted text was abstracted through text condensation and categorization, which was performed by CW and discussed with MB. Meaning units were identified and coded inductively using a Microsoft Excel spreadsheet. Thereafter, all codes were transferred to the open source FreeMind mind-mapping software [29], where they were grouped into categories and subcategories. After categorizing all charted data, we used the statistical software R (R Foundation for Statistical Computing) [30] to explore descriptive statistics and the ggplot2 package [31] to produce visualizations. We first present a descriptive numerical summary of the included papers and then present an inductive categorization of the main participatory health technology features, supported by illustrative examples. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist for scoping reviews was used for reporting.

Results

Study Selection

Database and manual searches yielded 2763 records (Figure 1); after removing duplicates, the titles and abstracts of 2475 (89.58%) records were screened, and 2360 (85.41%) records that did not meet the inclusion criteria were removed. We read the full texts of 115 papers and excluded 81 (70.4%) for the following reasons: not partnerships (n=67, 58.3%), not digital health technologies (n=11, 9.6%), not used in care (n=1, 0.9%), not evaluated (n=1, 0.9%), and not original research (n=1, 0.9%). The remaining 34 papers, reporting on 32 studies of 30 participatory health technologies, were included in the qualitative synthesis. The characteristics of individual papers are presented in Multimedia Appendix 2 [32-65]. In our presentation of study characteristics, data were consolidated from publications reporting on the e-BP study [33,34] and the MyCyFAPP study [43,44].

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

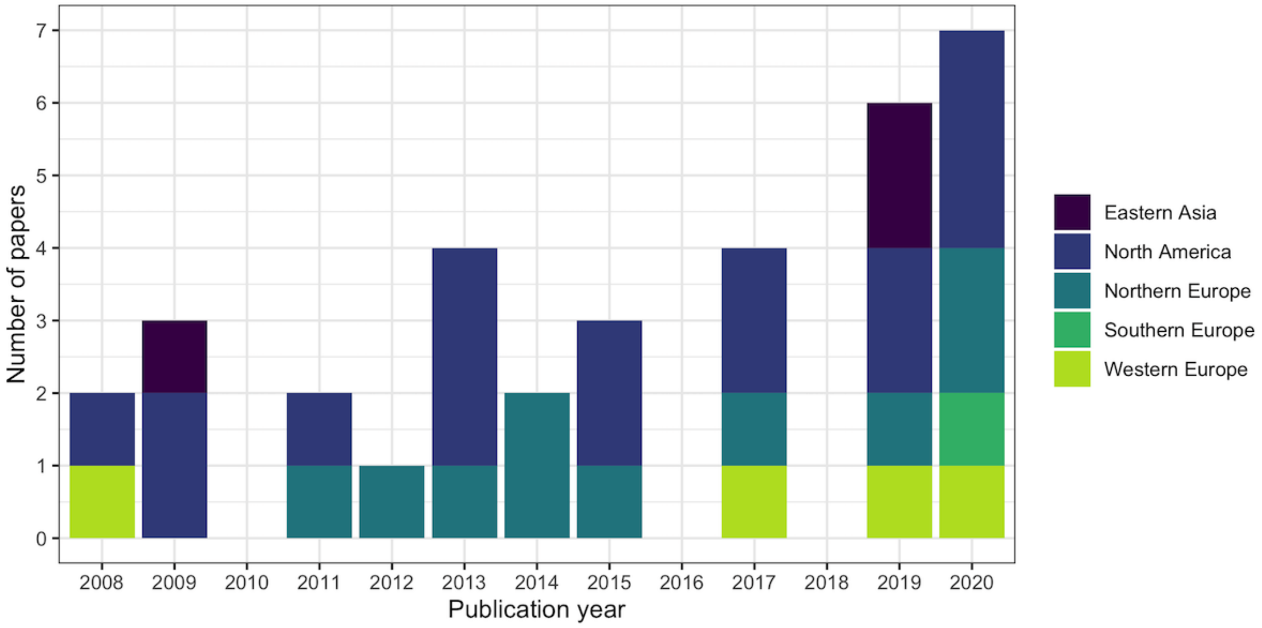


Study Characteristics

Publication Details

The topic of participatory health technologies experienced a slightly increasing trend across the publication years, with most papers being published in the past 2 years of the review period (Figure 2). The papers originated from the United States (15/34, 44%), Norway (7/34, 21%), and China (2/34, 6%), and a single (1/34, 3%) paper each from the following countries: Canada, Denmark, Sweden, the United Kingdom, the Netherlands, Germany, Austria, Switzerland, Spain, and South Korea. They were published in 21 journals, most commonly the *Journal of Medical Internet Research* (5/34, 15%), *JMIR mHealth and uHealth* (5/34, 15%), *Telemedicine and eHealth* (3/34, 9%), and the *International Journal of Medical Informatics* (3/34, 9%).

Figure 2. Number of included papers by year and world region.



Context of Use

The studies were conducted in primary care (18/32, 56%), secondary care (18/32, 56%), and tertiary care (7/32, 22%; Table 2). Approximately one-third (10/32, 31%) of the studies were set across levels of care, and some (3/32, 9%) additionally involved home care, social services, or school health services. Most studies addressed diabetes (8/32, 24%), followed by cardiovascular disease (6/32, 19%). The participatory health technologies were deployed almost exclusively as web applications or websites for health care professionals, whereas mobile deployment was common for patient users (14/32, 44%).

In some studies, participatory health technologies were integrated into electronic health record systems or personal health records [32-38]. In addition to interactions between patients and health care professionals, participatory health technologies supported collaboration with allied professionals [33,34,36,39] or technical staff [40,41]. Caregivers were identified as users in some studies in which the patients were children [37,42-45], cognitively impaired [46], or in palliative care [47]. We identified 3% (1/32) of studies in which patients who did not meet these criteria had the option to invite their families and friends to be users of participatory health technology [48].

Table 2. Context of use.

Characteristic	References
Level of care	
Primary care	[33-40,42,46,48-56]
Secondary care	[32,39,40,42-45,47,49,52,55-63]
Tertiary care	[39,41,43,44,58,60,64,65]
Type of chronic condition	
Metabolic (diabetes)	[35,38,51,53,55,57,64,65]
Cardiovascular	[32-34,36,39,48,61]
Autoimmune	[45,49,58-60]
Pulmonary	[37,48,54]
Cancer	[47,62,63]
Genetic	[43-45]
Immunodeficiency	[40,41]
Psychiatric	[42]
Neurodegenerative	[46]
Unspecified	[50,52]

Evaluation

Of the 32 studies, 13 (41%) were effect studies of participatory health technology use in clinical practice, 8 (25%) were feasibility studies, 7 (22%) explored user experiences, and 4 (13%) reported on the design and implementation of participatory health technologies (Table 3). The study designs included randomized clinical trials (11/32, 34%), quantitative evaluations (8/32, 25%), qualitative evaluations (7/32, 22%), and mixed methods evaluations (6/32, 19%). Of the 32 studies, the sample size was as high as 50 in 11 (34%) studies, 50 to 200 in 14 (44%) studies, and >200 in 7 (22%) studies. The studies evaluated the effects of participatory health technologies on clinical outcomes, including health, well-being, quality of life (17/32, 53%), user experiences (12/32, 38%), and self-management (7/32, 22%). Approximately one-third (12/32, 38%) of studies evaluated effects on partnerships by describing the content, experiences, and nature of collaboration [39,48,51,56]; the distribution of tasks and responsibilities [42]; patient-professional relationships [51,54]; engagements of patients and family caregivers [41,45,46]; and the perceived quality of collaborations [45,46]. Other effects that were evaluated included access to care and waiting times [49], continuity of care [47], and health care costs [34,52]. Most studies reported positive outcomes (22/32, 69%), although they were minor or temporary in some cases [35,62]. A few studies reported mixed results (5/32, 15%) or no change (3/32, 9%).

Table 3. Study designs and outcomes.

Characteristic	References
Study aim	
Design and implementation	[39,40,61,64]
User experiences	[42,46,48,50,51,56,59]
Feasibility	[32,37,45,47,53-55,60]
Effects	[33-36,38,41,43,44,49,52,57,58,62,63,65]
Study design	
Randomized controlled trial	[33-35,37,38,47,57,58,62,63,65]
Quantitative	[36,41,43,45,52,53,55,60,64]
Qualitative	[39,40,42,46,48,50,51]
Mixed methods	[32,44,54,56,59,61]
Sample size	
≤50	[39,42,46,48,50,51,54,55,60,61,64]
51-100	[32,37,38,45,47,56,57]
101-150	[36,41,53]
151-200	[40,43,44,58,63]
>200	[33-35,49,52,59,62,65]
Outcome variables	
Clinical outcomes	[32-38,41,43,44,47,55-58,62-65]
Partnership	[37,39-42,44-46,48,51,54,56]
Self-management	[32,36,44,45,57,62,63]
User experiences	[32,37,44-46,48,50,56,57,59,60,64]
Outcomes	
Positive outcomes	[32-37,39,41,43-46,48,49,53-55,57,59,60,62-65]
Mixed results	[42,47,50,51,56]
No change	[38,52,58]
N/A[a]	[40,61]

[a] N/A: not applicable.

Qualitative Synthesis of Participatory Health Technology Features and Their Influences on Partnerships

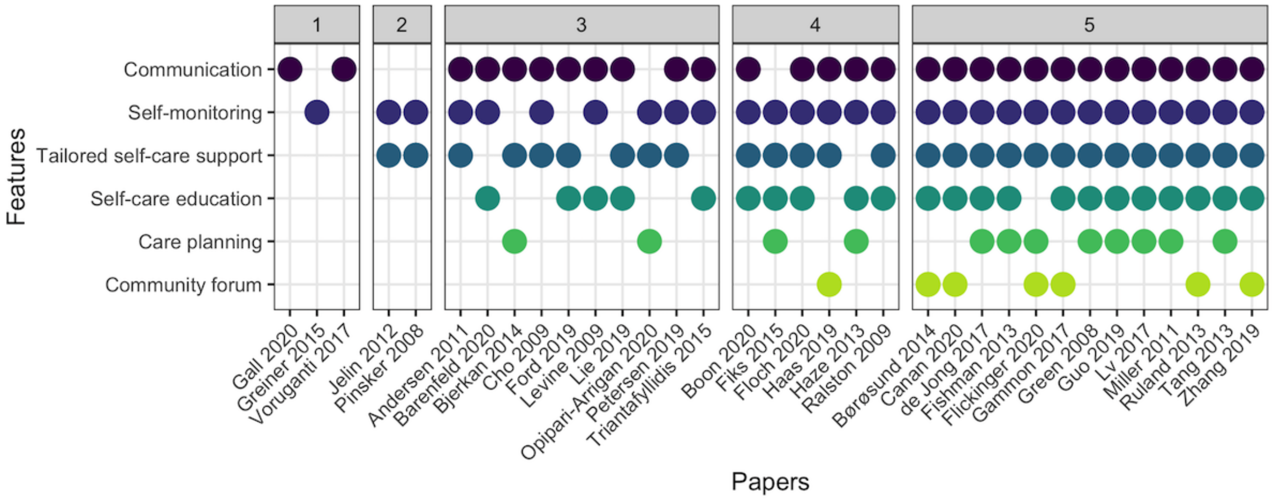
Overview of Features

We identified six main participatory health technology features for enabling partnerships between patients and health care professionals: (1) communication, (2) self-monitoring, (3) tailored self-care support, (4) self-care education, (5) care planning, and (6) community forum (Table 4). Most studies described a combination of these features, often involving the first 3 (Figure 3 [32-65]). In the following sections, we describe the 6 features and their influences on patient-professional partnerships that were discussed in the studies.

Table 4. Thematic analysis of participatory health technology features.

Themes (features)	References
Communication	
Asynchronous message exchange	[32-36,38,40-44,46-49,52-57,60-63]
Audio or video communication	[32,39,48,52]
Unspecified	[65]
Self-monitoring	
Self-measurements of health parameters	[33-36,52,53,57,59,61,65]
Self-assessment of symptoms or problems	[32,37,39,43-45,50,58,62,63]
Self-reported health status or activity	[32,35,38,40,41,44,45,48,50,54-56,59-61,64,65]
Self-reported medication adherence or side effects	[32,35,38,40,44,64]
Diary for personal notes	[48,55,63]
Tailored self-care support	
Personalized goals	[34-37,44,51,55,64]
Medication management	[32-34,36,39,41,49,56,64,65]
Individual feedback	[35,50,51,55,64]
Tailored recommendations	[33-35,37,43-45,49,52,57,62,63]
Alerts and reminders or prompts	[32,36-38,41,42,45,52,57,58,60,64]
Self-care education	
Educational material integrated in participatory health technology	[32,34-37,41,43,44,49,51,53,54,56,60-62,65]
Links to external sources	[32,33,38,48,58,62,63]
Care planning	
Access to a personal care plan	[32,34-37,40,42,54]
Appointments and previsit planning	[33,45,58,60]
Community forum	
Anonymous contributions	[40,41,55,56,62,63]
Health care professional monitored	[55,56,63]
Unspecified	[65]

Figure 3. Illustration of identified participatory health technology features in each of the included papers, grouped by the total number of features (ranging from 1 to 5) [32-65].



Communication

Support for patient-professional communication was a central participatory health technology feature described in 84% (27/32) of studies. Most commonly, communication was facilitated through asynchronous text-based information exchanges between users; however, some studies also implemented audio- or video-based communication. In some studies, team-based communication between multiple users (including patients, caregivers, care team members, and allied health professionals) was also supported, enabling patients and caregivers to communicate with multiple care team members and care team members to interact with each other [46,47,49]. This communication feature contributed to rapport building [40] and improved patient-professional relationships among young and adult patients [44,54,60]. In a study of teenagers living with asthma, health care professionals reported that written communication could lead to more honest and elaborate responses among patients who may be less talkative in face-to-face encounters [54]. Meanwhile, a study that evaluated e-consultations for diabetes self-management support found that asynchronous communication could make patient-nurse relationships more fragile because of the risk of misunderstandings, suggesting that the best option may be a combination of written and face-to-face interactions [51]. Most studies did not describe any constraints in content, time, or word limits for message exchanges. Although one of the studies reported that health care professionals did not experience answering messages as too time consuming [63], other studies reported that tighter communication and follow-up of patients led to greater workloads for health care professionals between consultations [44] and could blur the boundaries between their private and work lives [56]. Various engagements with participatory health technologies among health care professionals and patients led to frustration when expectations were not met, for example, when messages were not answered [42,56].

Self-monitoring

Self-monitoring was also a central feature found in nearly all studies (27/32, 84%). It comprised the use of self-measurement devices to register and report health data, such as blood pressures [33,34,36,61,64], blood glucose levels [35,53,57,64,65], or physical activity [36,52]; self-assessments of symptoms or problems; self-reported health statuses or activities; self-reported medication adherence or side effects; and diaries for personal notes. When health parameters were not measured using external devices, self-monitoring was mostly facilitated through structured data input based on predefined forms. Several benefits regarding patient-provider partnerships were identified: increased patient motivation [50,59], higher perceptions of being recognized and respected by health care professionals [48], and more efficient consultations because of less time being spent on collecting and explaining data [44]. By providing contextual information and their own interpretations of self-monitored data, patients could participate as diagnostic agents in clinical assessments [39].

Tailored Self-care Support

The features for providing tailored self-care support were identified in 78% (25/32) of studies. This entailed support for setting and monitoring the progression toward personalized goals, medication management (eg, personal medication lists and managing refills), individual feedback, tailored recommendations, and alerts and reminders. Health care professionals provided individual feedback to patients on the basis of clinical variables or reported self-assessments and reflections [35,50,51,55,64]. In addition to feedback, tailored recommendations were often provided regarding therapy adjustments, symptom management, self-management activities, self-monitoring, and topics to discuss with clinicians. Recommendations were either automatically generated based on patients' reported data [37,43-45,52,62,63] or individually tailored by health care professionals [34,35,49,57]. The provision of feedback contributed to the development of good relationships and made patients feel understood and addressed, although feedback could also be experienced as challenging for patients [50]. Motivational messages could be both appreciated and experienced as annoying [44]. Alerts were provided to draw attention to patients and health care professionals, generally based on predefined threshold values for clinical parameters [32,36,45,60,64]. In some cases, alerts were also used to inform health care professionals about patients' activities or engagement with participatory health technologies [42,52]. Reminders were used to support medication adherence [41], prompt patients to upload self-monitoring data [37,57,58,64], or remind them of clinical examinations or appointments [38].

Self-care Education

Features for providing self-care education were described in almost two-thirds of the studies (20/32, 62%). Educational material was integrated into the participatory health technologies or provided through links to external sources and often covered both disease-specific information and lifestyle topics, such as nutrition, health and wellness, or smoking cessation [41,43,44,60]. Where educational material was embedded in participatory health technology, contents could be adapted specifically to the target group. For example, in a study of patients with diabetes [53], educational material was adapted to be culturally appropriate to the target group of native communities. Self-care education was sometimes delivered through video clips [35,37,54,61] or could include a toolbox of resources, such as recommended activities, good-to-know texts, and workbooks [56]. Several studies reported improvements in self-management knowledge and self-efficacy [32,43,44,60].

Care Planning

Approximately one-third (11/32, 34%) of the studies described features for participatory health technology–supported care planning. This involved access to planned activities or personal health plans and support for scheduling appointments or planning care visits. In preparation for care visits, patients had opportunities to identify goals, questions, or problems to discuss with their clinicians and provide information about their disease activity by filling in structured data forms [45,58,60]. A study of pediatric patients [45] found that visits and collaborations improved through this preparation. Another study identified shifts in roles and sometimes power transitions from health care professionals to patients and caregivers as they took more responsibility for care planning [42].

Community Forum

Web-based community forums for peer-to-peer interactions with other patients were provided in some studies (7/32, 22%). This functionality appeared in studies published in 2013 or later and only in participatory health technologies that had several other features as well. In most cases (5/7, 71%), the community forums enabled patients to write questions and comments anonymously to protect their integrity. In 43% (3/7) of studies, community forums were monitored by health care professionals who could contribute with answers to posted questions. In a community mental care setting, it was reported that peer support initially established through an anonymous community forum could develop into friendships when combined with café gatherings where service users could meet in real life [56]. Several studies found that patients would visit the web-based community forums to read others' posts more often than to post something themselves [41,56,62].

Discussion

Principal Findings

This scoping review identified and described the characteristics of participatory health technologies supporting patient-professional partnerships in chronic care management evaluated in 32 studies and published in 34 papers. These papers originated almost exclusively from North America and Europe and were published in a variety of journals, mainly in the fields of biomedical informatics or information science but also in the fields of health services research, medicine, and nursing. This reflects the multidisciplinary nature of participatory health informatics in chronic care management. The slight increase in the publication trends may suggest an increasing interest in digital services for participatory medicine in recent years. Notably, the included papers represented high variation in terms of the chronic conditions addressed, the levels of care where the participatory health technologies were used, the study designs, and the sample sizes. Nevertheless, 6 common participatory health technology features could be identified. Most participatory health technologies had features to support patient-professional communication, self-monitoring, and tailored self-care support. More than half of the studies described self-care education features, and approximately one-third discussed features to support care planning. In more recently published studies, the facilitation of peer support through web-based community forums emerged as a new feature. The engagement of caregivers as participatory health technology users was also more common in recent studies, possibly indicating a shift from focusing merely on the patient-professional dyad to a system view of collaborative care, acknowledging the involvement of more stakeholders. Most studies reported positive outcomes, although there were mixed results, highlighting the importance of tailoring participatory health technology implementation and use to individuals' needs and preferences.

Comparison With Prior Work

Our thematic analysis focused on identifying common participatory health technology features and describing the identified influences on patient-professional partnerships. In the following sections, we discuss 3 observations made when interpreting our findings, namely, how participatory health technologies influence roles and relationships, the changing nature of chronic care work, and a shift from intermediation to apomediation.

Changing Roles and Relationships

Overall, our findings are in line with previous studies indicating that the use of eHealth interventions can positively influence patient-professional communication and relationships [20] and also challenge these because of undefined or changed roles [66]. As a previous review has shown [20], the positive influences of participatory health technologies on patient-professional relationships depend on participating actors who meet the expectations and rules of minimal engagement. This was clearly seen in some studies in mental care [42,56], where variation in patient or professional engagement with participatory health technologies could lead to either enhanced or challenged relationships. One of these studies highlighted that health care professionals may need to communicate their personal boundaries to patients; for example, they would only check messages on certain weekdays [56]. However, these kinds of social interaction norms have rarely been made explicit in studies where asynchronous interactions were not structured or constrained, which could lead to a blurring of the boundaries that define the contents, extents, and times of interactions [67,68]. The ethical implications of digital patient-professional communication can be complex and require organizational guidelines to promote good practices in the use of digital communication [67]. Role uncertainty may affect both staff and patients [69], suggesting that the introduction of participatory health technology features that enable unlimited asynchronous interactions or task shifting (eg, the patient takes on tasks traditionally performed by health care professionals) should also involve mutual agreements on the distribution of tasks, roles, and responsibilities between patients and professionals.

Changing Nature of Collaborative Chronic Care Work

In addition to communication support, the most common participatory health technology features we identified were self-monitoring and tailored self-care support, which is comparable with the results of a recent scoping review focusing on the features of web portals for telerehabilitation [70]. These 3 features were often combined, and they have the potential to profoundly influence the nature of collaborative work among patients, caregivers, and health care professionals. Through self-monitoring and self-care, patients take over tasks that were previously performed by health care staff (eg, measuring of vital parameters) or not performed at all (eg, continuous collection of health parameters between consultations). As described in one of the papers in our review [44], patients' self-monitoring also influenced the work of health care professionals in several ways. Consultations could become more efficient as data were collected in real time and made available to both patients and health care professionals before consultations. In contrast, health care professionals had to spend more time between visits responding to questions or providing feedback to their patients (ie, communication and tailored self-care support). This indicates that health care professionals' workloads may increase in some areas but decrease in others, with implications for how their work is organized. Workloads, workflow disruptions, and alignment with clinical processes are among the most common barriers to the adoption of eHealth services [66]. Another study found that when patients' transmission of data replaced physical meetings, the patients could become passively disengaged, resulting in poorer collaborations [39]. Enabling patients to provide contextual information in addition to automated self-measurements contributed to reintroducing them as collaborative partners in diagnostic interpretation; however, the authors questioned whether this could really be labeled as collaboration or merely the transmission of more data. This study clearly problematized the potential issues when self-monitoring merely replaced previous collaborative work. When self-measured data are not interpreted in collaboration with the patient, the partnership may be harmed rather than improved.

Features for care planning provided another example in which the nature of collaborative work could change. For example, care planning enabled patients to influence the agenda for care visits by communicating their personal goals and the questions they wanted to address. One of the studies described a power transition, as patients took more responsibility for their care plans [42]. Altogether, these findings emphasize that the potential implications of participatory health technologies on the nature of collaborative work need to be carefully considered when introducing eHealth services that influence the work of patients and health care professionals in chronic care management.

Moving From Intermediation to Apomediation

With the rise of web-based technologies, referred to as Web 2.0, and similarly, Medicine 2.0, the terms intermediation, disintermediation, and apomediation were introduced [71,72]. Intermediation refers to the selection and delivery of "relevant" health information by an intermediary (eg, health care professionals or a web portal vetted by experts). For example, self-care education and self-care support features that were quite common in this review may be understood as methods of intermediation. By providing patients with relevant self-care information, health care professionals can shift away from the paternalistic model of physician-patient relationships to an interpretive model, where they take on roles as counselors or advisers in individuals' self-care [73]. The provision of self-care education was associated with increased knowledge and self-efficacy, which are resources that individuals can draw on to build their capacities for self-management [74]. It has been suggested that the more knowledgeable and self-efficacious patients become, the less they want to rely on experts (ie, disintermediation), preferring guidance from peers who "stand by" rather than "in between" patients and the knowledge they seek (ie, apomediation) [71]. An example of apomediation is web-based social health networks, which have been integrated as components in the eHealth-enhanced CCM [75]. Although web-based communities marked the beginning of participatory health informatics [16], the integration of social networking features in participatory health technologies intended for patient-professional interactions emerged as a new trend in this study. Our results illustrate that the 3 different types of participatory health technologies that have been previously distinguished (ie, Web 2.0, self-care support, and tools supporting health care provision) are being increasingly combined in multimodal services. This suggests that participatory health technologies may indeed enable a shift toward a more collaborative and networked approach to participatory medicine beyond the patient-professional dyad. We have identified several features to support partnerships in chronic care management; however, the processes of how patient knowledge is shaped and integrated in shared decision-making are still poorly characterized [76]. Future research may reveal how knowledge from web-based health communities, patients, caregivers, and health care professionals can be effectively combined to support patients in their individual self-care and drive quality improvement and collective organizational learning.

Strengths and Limitations

This scoping review has several strengths, including the inclusion of all types of study designs to obtain findings assessed using different methods, a screening method involving multiple researchers, and a qualitative synthesis contributing to new knowledge. The included studies covered various chronic conditions, clinical settings, and study designs. Our search strategy limited the review to papers published in English and Swedish between 2008 and 2020, implying that we may have missed important studies published earlier and in different languages. Furthermore, the inclusion criteria restricted the studies to those that reported the use of software specifically intended for clinical use (ie, excluding the use of email, SMS text messages, or nondigitally supported means of partnership) and had been evaluated in clinical practice. Nevertheless, our findings add new knowledge that contributes to describing the scope and nature of participatory health technology features to support patient-professional partnerships. Only 38% (12/32) of studies evaluated the effects on partnerships, which suggests that a knowledge gap remains regarding the influence of participatory health technologies on the nature of partnerships and how to support collaborative health care practices effectively. As most studies reported positive results, there may also be a publication bias, given that studies of failed eHealth interventions are published less frequently [77].

Conclusions

This scoping review identified participatory health technologies evaluated in studies intending to support partnerships between patients and caregivers and health care professionals in chronic care and qualitatively analyzed the main features of these technologies. A total of 6 common features were identified: patient-professional communication, self-monitoring, tailored self-care support, self-care education, care planning support, and community forums for peer-to-peer interactions. The integration of social networking features for community support in health technologies intended for patient-professional interactions is an emerging trend, which suggests a shift toward a more collaborative and networked approach to participatory medicine beyond the patient-professional dyad. The studies in this review mainly reported positive outcomes; however, we also identified how partnership relationships and the nature of collaborative work could be challenged when roles and expectations between users were unclear. This emphasizes the importance of clarifying mutual expectations and carefully considering the implications that the introduction of participatory health technologies may have on the work of patients and health care professionals, individually and in collaboration. Future research should further explore the mechanisms by which participatory health technologies contribute to the shaping and use of collaborative knowledge to benefit individual patients, patient populations, and organizational learning.

Acknowledgments

The research was funded by the Swedish Research Council for Health, Working Life, and Welfare (Forte) (grants number 2014-4238 and 2017-01451), the latter funding CW. The APC and Marie Dahlberg were additionally funded by the Kamprad Family Foundation for Entrepreneurship, Research, and Charity (grant number 20170012).

Authors' Contributions

CW contributed to conceptualization; methodology; formal analysis; investigation; data curation; writing, reviewing, and editing; and visualization, supervision, project administration, and funding acquisition. MÅW contributed to conceptualization, methodology, investigation, data curation, and reviewing and editing. MD, ÅR, and ST contributed to conceptualization, methodology, investigation, and reviewing and editing. EE contributed to conceptualization, methodology, investigation, data curation, reviewing and editing, supervision, and project administration. MB contributed to conceptualization; investigation; writing of the original draft, reviewing, and editing; supervision; project administration; and funding acquisition. All the authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

MÅW was at the time of the study part-time employed by Philips. However, the company had no influence or financial interests related to the manuscript.

Multimedia Appendix 1
Search strings.
[PDF File (Adobe PDF File), 98 KB - jmir_v24i8e38980_app1.pdf]

Multimedia Appendix 2
Characteristics of included papers.
[PDF File (Adobe PDF File), 161 KB - jmir_v24i8e38980_app2.pdf]

References

1. Atella V, Piano Mortari A, Kopinska J, Belotti F, Lapi F, Cricelli C, et al. Trends in age-related disease burden and healthcare utilization. Aging Cell 2019 Feb;18(1):e12861 [FREE Full text] [doi: 10.1111/acel.12861] [Medline: 30488641]

2. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-544. [Medline: 8941260]
3. 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]
4. Understanding and Using Health Experiences Improving Patient Care. Oxford: Oxford Scholarship Online; 2013.
5. Living Well with Chronic Illness A Call for Public Health Action. Washington, DC: National Academies Press; 2012.
6. deBronkart D. The patient's voice in the emerging era of participatory medicine. Int J Psychiatry Med 2018 Nov;53(5-6):350-360. [doi: 10.1177/0091217418791461] [Medline: 30114957]
7. Wagner EH. Organizing care for patients with chronic illness revisited. Milbank Q 2019 Sep;97(3):659-664 [FREE Full text] [doi: 10.1111/1468-0009.12416] [Medline: 31424130]
8. Håkansson Eklund J, Holmström IK, Kumlin T, Kaminsky E, Skoglund K, Höglander J, et al. "Same same or different?" A review of reviews of person-centered and patient-centered care. Patient Educ Couns 2019 Jan;102(1):3-11. [doi: 10.1016/j.pec.2018.08.029] [Medline: 30201221]
9. World Health Organization. Framework for Action on Interprofessional Education & Collaborative Practice. 2010 Sep 01. URL: https://www.who.int/publications/i/item/framework-for-action-on-interprofessional-education-collaborative-practice [accessed 2022-03-10]
10. Kennedy I. Patients are experts in their own field. BMJ 2003 Jun 14;326(7402):1276-1277 [FREE Full text] [doi: 10.1136/bmj.326.7402.1276] [Medline: 12805126]
11. Goodrich J. Why experience-based co-design improves the patient experience. J Health Design 2018 Mar 22;3(1):84-85. [doi: 10.21853/jhd.2018.45]
12. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, et al. A systematic review of chronic disease management interventions in primary care. BMC Fam Pract 2018 Jan 09;19(1):11 [FREE Full text] [doi: 10.1186/s12875-017-0692-3] [Medline: 29316889]
13. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst 2017 Sep 01;109(9):436. [doi: 10.1093/jnci/djx029] [Medline: 28423407]
14. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
15. Armstrong TS, Bishof AM, Brown PD, Klein M, Taphoorn MJ, Theodore-Oklota C. Determining priority signs and symptoms for use as clinical outcomes assessments in trials including patients with malignant gliomas: panel 1 Report. Neuro Oncol 2016 Mar;18 Suppl 2:ii1-i12 [FREE Full text] [doi: 10.1093/neuonc/nov267] [Medline: 26989127]
16. Denecke K, Gabarron E, Petersen C, Merolli M. Defining participatory health informatics - a scoping review. Inform Health Soc Care 2021 Sep 02;46(3):234-243. [doi: 10.1080/17538157.2021.1883028] [Medline: 33622168]
17. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
18. Voruganti T, Grunfeld E, Makuwaza T, Bender JL. Web-based tools for text-based patient-provider communication in chronic conditions: scoping review. J Med Internet Res 2017 Oct 27;19(10):e366 [FREE Full text] [doi: 10.2196/jmir.7987] [Medline: 29079552]
19. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
20. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019 Jun;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
21. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
22. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
23. 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]
24. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
25. Thomas J, Kneale D, McKenzie J, Brennan S, Bhaumik S. Determining the scope of the review and the questions it will address. In: Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Hoboken, New Jersey, United States: Wiley; 2019.

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. 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]
28. Lindgren B, Lundman B, Graneheim UH. Abstraction and interpretation during the qualitative content analysis process. Int J Nurs Stud 2020 Aug;108:103632. [doi: 10.1016/j.ijnurstu.2020.103632] [Medline: 32505813]
29. Müller J, Polansky D, Novak P, Foltin C, Polivaev D. FreeMind - free software, licensed under GNU General Public License. 2013. URL: http://freemind.sf.net/ [accessed 2021-03-22]
30. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.
31. Wickham H. Ggplot2 Elegant Graphics for Data Analysis. New York, United States: Springer; 2016.
32. Guo X, Gu X, Jiang J, Li H, Duan R, Zhang Y, et al. A hospital-community-family-based telehealth program for patients with chronic heart failure: single-arm, prospective feasibility study. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13229 [FREE Full text] [doi: 10.2196/13229] [Medline: 31833835]
33. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]
34. Fishman PA, Cook AJ, Anderson ML, Ralston JD, Catz SL, Carrell D, et al. Improving BP control through electronic communications: an economic evaluation. Am J Manag Care 2013 Sep;19(9):709-716 [FREE Full text] [Medline: 24304254]
35. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 01;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
36. Lv N, Xiao L, Simmons ML, Rosas LG, Chan A, Entwistle M. Personalized hypertension management using patient-generated health data integrated with electronic health records (empower-h): six-month pre-post study. J Med Internet Res 2017 Sep 19;19(9):e311 [FREE Full text] [doi: 10.2196/jmir.7831] [Medline: 28928111]
37. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, Localio AR, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973 [FREE Full text] [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
38. Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care 2009 Feb;32(2):234-239 [FREE Full text] [doi: 10.2337/dc08-1220] [Medline: 19017773]
39. Andersen T, Bjørn P, Kensing F, Moll J. Designing for collaborative interpretation in telemonitoring: re-introducing patients as diagnostic agents. Int J Med Inform 2011 Aug;80(8):e112-e126. [doi: 10.1016/j.ijmedinf.2010.09.010] [Medline: 21067968]
40. Flickinger TE, Ingersoll K, Swoger S, Grabowski M, Dillingham R. Secure messaging through PositiveLinks: examination of electronic communication in a clinic-affiliated smartphone app for patients living with HIV. Telemed J E Health 2020 Mar;26(3):359-364 [FREE Full text] [doi: 10.1089/tmj.2018.0261] [Medline: 30900961]
41. Canan CE, Waselewski ME, Waldman AL, Reynolds G, Flickinger TE, Cohn WF, et al. Long term impact of PositiveLinks: clinic-deployed mobile technology to improve engagement with HIV care. PLoS One 2020;15(1):e0226870 [FREE Full text] [doi: 10.1371/journal.pone.0226870] [Medline: 31905209]
42. Bjerkan J, Vatne S, Hollingen A. Web-based collaboration in individual care planning challenges the user and the provider roles - toward a power transition in caring relationships. J Multidiscip Healthc 2014;7:561-572 [FREE Full text] [doi: 10.2147/JMDH.S70470] [Medline: 25525367]
43. Boon M, Calvo-Lerma J, Claes I, Havermans T, Asseiceira I, Bulfamante A, et al. Use of a mobile application for self-management of pancreatic enzyme replacement therapy is associated with improved gastro-intestinal related quality of life in children with Cystic Fibrosis. J Cyst Fibros 2020 Jul;19(4):562-568. [doi: 10.1016/j.jcf.2020.04.001] [Medline: 32335023]
44. Floch J, Vilarinho T, Zettl A, Ibanez-Sanchez G, Calvo-Lerma J, Stav E, et al. Users' experiences of a mobile health self-management approach for the treatment of cystic fibrosis: mixed methods study. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e15896 [FREE Full text] [doi: 10.2196/15896] [Medline: 32673237]
45. Opipari-Arrigan L, Dykes DM, Saeed SA, Thakkar S, Burns L, Chini BA, et al. Technology-enabled health care collaboration in pediatric chronic illness: pre-post interventional study for feasibility, acceptability, and clinical impact of an electronic health record-linked platform for patient-clinician partnership. JMIR Mhealth Uhealth 2020 Nov 26;8(11):e11968 [FREE Full text] [doi: 10.2196/11968] [Medline: 33242014]
46. Gall D, Preßler J, Hurtienne J, Latoschik ME. Self-organizing knowledge management might improve the quality of person-centered dementia care: a qualitative study. Int J Med Inform 2020 Jul;139:104132. [doi: 10.1016/j.ijmedinf.2020.104132] [Medline: 32416522]

47. Voruganti T, Grunfeld E, Jamieson T, Kurahashi AM, Lokuge B, Krzyzanowska MK, et al. My team of care study: a pilot randomized controlled trial of a web-based communication tool for collaborative care in patients with advanced cancer. J Med Internet Res 2017 Jul 18;19(7):e219 [FREE Full text] [doi: 10.2196/jmir.7421] [Medline: 28720558]
48. Barenfeld E, Ali L, Wallström S, Fors A, Ekman I. Becoming more of an insider: a grounded theory study on patients' experience of a person-centred e-health intervention. PLoS One 2020;15(11):e0241801 [FREE Full text] [doi: 10.1371/journal.pone.0241801] [Medline: 33226986]
49. Ford AR, Gibbons CM, Torres J, Kornmehl HA, Singh S, Young PM, et al. Access to dermatological care with an innovative online model for psoriasis management: results from a randomized controlled trial. Telemed J E Health 2019 Jul;25(7):619-627 [FREE Full text] [doi: 10.1089/tmj.2018.0160] [Medline: 30222518]
50. Jelin E, Granum V, Eide H. Experiences of a web-based nursing intervention--interviews with women with chronic musculoskeletal pain. Pain Manag Nurs 2012 Mar;13(1):2-10. [doi: 10.1016/j.pmn.2011.08.008] [Medline: 22341135]
51. Lie SS, Karlsen B, Graue M, Oftedal B. The influence of an eHealth intervention for adults with type 2 diabetes on the patient-nurse relationship: a qualitative study. Scand J Caring Sci 2019 Sep;33(3):741-749. [doi: 10.1111/scs.12671] [Medline: 30866066]
52. Petersen C, Weeks W, Norin O, Weinstein J. Development and implementation of a person-centered, technology-enhanced care model for managing chronic conditions: cohort study. JMIR Mhealth Uhealth 2019 Mar 20;7(3):e11082 [FREE Full text] [doi: 10.2196/11082] [Medline: 30892274]
53. Levine BA, Turner JW, Robinson JD, Angelus P, Hu TM. Communication plays a critical role in web-based monitoring. J Diabetes Sci Technol 2009 May 01;3(3):461-467 [FREE Full text] [doi: 10.1177/193229680900300309] [Medline: 20144283]
54. Haze KA, Lynaugh J. Building patient relationships: a smartphone application supporting communication between teenagers with asthma and the RN care coordinator. Comput Inform Nurs 2013 Jun;31(6):266-71; quiz 272. [doi: 10.1097/NXN.0b013e318295e5ba] [Medline: 23728445]
55. Haas K, Hayoz S, Maurer-Wiesner S. Effectiveness and feasibility of a remote lifestyle intervention by dietitians for overweight and obese adults: pilot study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e12289 [FREE Full text] [doi: 10.2196/12289] [Medline: 30973338]
56. Gammon D, Strand M, Eng LS, Børøsund E, Varsi C, Ruland C. Shifting practices toward recovery-oriented care through an e-recovery portal in community mental health care: a mixed-methods exploratory study. J Med Internet Res 2017 May 02;19(5):e145 [FREE Full text] [doi: 10.2196/jmir.7524] [Medline: 28465277]
57. Cho J, Lee H, Lim D, Kwon H, Yoon K. Mobile communication using a mobile phone with a glucometer for glucose control in Type 2 patients with diabetes: as effective as an internet-based glucose monitoring system. J Telemed Telecare 2009;15(2):77-82. [doi: 10.1258/jtt.2008.080412] [Medline: 19246607]
58. Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee J, et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemed J E Health 2011;17(1):5-13 [FREE Full text] [doi: 10.1089/tmj.2010.0133] [Medline: 21214498]
59. Greiner P, Sawka A, Imison E. Patient and physician perspectives on MSdialog, an electronic PRO diary in multiple sclerosis. Patient 2015 Dec;8(6):541-550 [FREE Full text] [doi: 10.1007/s40271-015-0140-1] [Medline: 26350792]
60. de Jong M, van der Meulen-de Jong A, Romberg-Camps M, Degens J, Becx M, Markus T, et al. Development and feasibility study of a telemedicine tool for all patients with IBD: MyIBDcoach. Inflamm Bowel Dis 2017 Apr;23(4):485-493. [doi: 10.1097/MIB.0000000000001034] [Medline: 28267047]
61. Triantafyllidis A, Velardo C, Chantler T, Shah SA, Paton C, Khorshidi R, SUPPORT-HF Investigators. A personalised mobile-based home monitoring system for heart failure: the SUPPORT-HF study. Int J Med Inform 2015 Oct;84(10):743-753. [doi: 10.1016/j.ijmedinf.2015.05.003] [Medline: 26037921]
62. Ruland CM, Andersen T, Jeneson A, Moore S, Grimsbø GH, Børøsund E, et al. Effects of an internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs 2013;36(1):6-17. [doi: 10.1097/NCC.0b013e31824d90d4] [Medline: 22495503]
63. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014 Dec 18;16(12):e295 [FREE Full text] [doi: 10.2196/jmir.3348] [Medline: 25525672]
64. Pinsker M, Schindler K, Morak J, Hayn D, Kastner P, Riedl M, et al. Experiences using mobile phones as patient-terminal for telemedical home care and therapy monitoring of patients suffering from chronic diseases. In: Proceedings of the 11th international conference on Computers Helping People with Special Needs. 2008 Presented at: ICCHP '08: 11th international conference on Computers Helping People with Special Needs; Jul 9 - 11, 2008; linz Austria. [doi: 10.1007/978-3-540-70540-6_195]
65. Zhang L, He X, Shen Y, Yu H, Pan J, Zhu W, et al. Effectiveness of smartphone app-based interactive management on glycemic control in Chinese patients with poorly controlled diabetes: randomized controlled trial. J Med Internet Res 2019 Dec 09;21(12):e15401 [FREE Full text] [doi: 10.2196/15401] [Medline: 31815677]
66. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]

67. Ignatowicz A, Slowther A, Elder P, Bryce C, Hamilton K, Huxley C, et al. Ethical implications of digital communication for the patient-clinician relationship: analysis of interviews with clinicians and young adults with long term conditions (the LYNC study). BMC Med Ethics 2018 Feb 23;19(1):11 [FREE Full text] [doi: 10.1186/s12910-018-0250-0] [Medline: 29475437]
68. Wannheden C, von Thiele Schwarz U, Östenson C, Pukk Härenstam K, Stenfors T. What's the name of the game? The impact of eHealth on productive interactions in chronic care management. Sustainability 2021 May 07;13(9):5221. [doi: 10.3390/su13095221]
69. Davidson AR, Kelly J, Ball L, Morgan M, Reidlinger DP. What do patients experience? Interprofessional collaborative practice for chronic conditions in primary care: an integrative review. BMC Prim Care 2022 Jan 14;23(1):8 [FREE Full text] [doi: 10.1186/s12875-021-01595-6] [Medline: 35172731]
70. Morimoto Y, Takahashi T, Sawa R, Saitoh M, Morisawa T, Kagiyama N, et al. Web portals for patients with chronic diseases: scoping review of the functional features and theoretical frameworks of telerehabilitation platforms. J Med Internet Res 2022 Jan 27;24(1):e27759 [FREE Full text] [doi: 10.2196/27759] [Medline: 35084355]
71. Eysenbach G. From intermediation to disintermediation and apomediation: new models for consumers to access and assess the credibility of health information in the age of Web2.0. Stud Health Technol Inform 2007;129(Pt 1):162-166. [Medline: 17911699]
72. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug 25;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
73. Emanuel EJ, Emanuel LL. Four models of the physician-patient relationship. JAMA 1992;267(16):2221-2226. [Medline: 1556799]
74. Boehmer KR, Gionfriddo MR, Rodriguez-Gutierrez R, Dabrh AM, Leppin AL, Hargraves I, et al. Patient capacity and constraints in the experience of chronic disease: a qualitative systematic review and thematic synthesis. BMC Fam Pract 2016 Sep 01;17:127 [FREE Full text] [doi: 10.1186/s12875-016-0525-9] [Medline: 27585439]
75. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LM. 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]
76. Papautsky EL, Patterson ES. Patients are knowledge workers in the clinical information space. Appl Clin Inform 2021 Jan;12(1):133-140 [FREE Full text] [doi: 10.1055/s-0041-1723022] [Medline: 33626585]
77. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

Abbreviations

CCM: chronic care model
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

---

Edited by T Leung; submitted 24.04.22; peer-reviewed by R Armstrong Junior, I Mircheva; comments to author 19.05.22; revised version received 02.06.22; accepted 16.06.22; published 01.08.22.

Please cite as:
Wannheden C, Åberg-Wennerholm M, Dahlberg M, Revenäs Å, Tolf S, Eftimovska E, Brommels M
Digital Health Technologies Enabling Partnerships in Chronic Care Management: Scoping Review
J Med Internet Res 2022;24(8):e38980
URL: https://www.jmir.org/2022/8/e38980
doi:10.2196/38980
PMID:35916720

Review

Social Media Use for Research Participant Recruitment: Integrative Literature Review

Elizabeth Mirekuwaa Darko[1*], BScN, MScN; Manal Kleib[1*], MScN, MBA, PhD; Joanne Olson[1*], PhD

College of Health Sciences, Faculty of Nursing, Edmonton Clinic Health Academy, University of Alberta, Edmonton, AB, Canada

[*] all authors contributed equally

Corresponding Author:
Elizabeth Mirekuwaa Darko, BScN, MScN
College of Health Sciences, Faculty of Nursing
Edmonton Clinic Health Academy
University of Alberta
11405-87 Avenue
Edmonton, AB, T6G 1C9
Canada
Phone: 1 7807826810
Email: darko@ualberta.ca

Abstract

Background: Social media tools have provided health researchers with the opportunity to engage with communities and groups in a nonconventional manner to recruit participants for health research. Using social media to advertise research opportunities and recruit participants facilitates accessibility to participants from broad geographical areas and diverse populations. However, little guidance is provided by ethics review boards for researchers to effectively use this recruitment method in their research.

Objective: This study sought to explore the literature on the use of social media for participant recruitment for research studies and identify the best practices for recruiting participants using this method.

Methods: An integrative review approach was used to synthesize the literature. A total of 5 health sciences databases, namely, EMBASE (Ovid), MEDLINE (Ovid and EBSCOhost), PsycINFO (Ovid), Scopus (Elsevier), and CINAHL Plus with Full Text (EBSCOhost), were searched using predefined keywords and inclusion and exclusion criteria. The initial search was conducted in October 2020 and was updated in February 2022. Descriptive and content analyses were applied to synthesize the results, and the findings are presented in a narrative and tabular format.

Results: A total of 96 records were included in this review, 83 (86%) from the initial search and 13 (14%) from the updated search. The publication year ranged between 2011 and 2022, with most publications (63/96, 66%) being from the United States. Regarding recruitment strategy, 45% (43/96) of the studies exclusively used social media, whereas 51% (49/96) used social media in conjunction with other strategies. The remaining 4% (4/96) provided guidelines and recommendations for social media recruitment. Notably, 38% (36/96) of these studies involved hard-to-reach populations. The findings also revealed that the use of social media is a cost-effective and efficient strategy for recruiting research participants. Despite the expanded use across different populations, there is limited participation of older adults in social media recruitment.

Conclusions: This review provides important insights into the current use of social media for health research participant recruitment. Ethics boards and research support services in academic institutions are encouraged to explicitly provide researchers with guidelines on the use of social media for health research participant recruitment. A preliminary guideline prepared based on the findings of this review is proposed to spark further development in this area.

(J Med Internet Res 2022;24(8):e38015) doi:10.2196/38015

KEYWORDS

advertisement; recruitment; research participants; social media; mobile phone

Introduction

Background

In this digital age, advancements in technology have created opportunities for researchers to use new techniques to recruit research participants. For health researchers, technological innovations present an opportunity to use digital platforms such as social media, the internet, web applications, multimedia, and smartphones to effectively and efficiently engage the community for research recruitment [1]. These digital platforms provide an additional source for participant recruitment for health research studies [2]. Within health sciences, social media is being quickly adopted because of its increased use as a method of communication with the public [3]. For many researchers, recruiting participants for trials can be a daunting task that can result in study delays or the termination of trials [4]. Less than one-third of trials reach their original target within a specified time frame, and approximately one-third required extension [5]. Hence, reaching targeted participants through social media platforms provides an important avenue for facilitating researchers' work.

Social media refers to a group of internet-based communication services through which users create and participate in web-based exchanges, contribute user-created content such as videos, or join web-based communities to share information and ideas [6]. The trends and patterns of social engagement worldwide help provide researchers, policy makers, and other stakeholders with an overview of the different social media applications that users are engaged with [7] and how these tools could potentially be used to leverage health research. With a global population of 7.8 billion inhabitants [8], internet users stand at 4.54 billion, representing a 59% penetration rate, and active social media users at 3.80 billion, representing 49% [9,10]. Active social media platforms users include Facebook (63%), YouTube (61%), WhatsApp (48%), Facebook Messenger (38%), Instagram (36%), Twitter (23%), and Snapchat (13%) [9,11]. Social media provides an appropriate medium for user connection and communication, information collection and dissemination, knowledge sharing, discussion, and collaboration with communities for professional networking and business purposes [12-14].

Despite the numerous benefits and opportunities associated with social media, its use in the recruitment of research participants is still evolving. Health researchers using digital platforms for research participant recruitment encounter challenges such as efficiency, cost, information reliability, informed consent, confidentiality, privacy-related concerns [15], internet accessibility, information overload, informed consent, and interaction quality [12,13]. In traditional recruitment methods, researchers often face costs associated with personnel and resources, administrative changes, time-consuming recruitment processes, recruitment bias, and population homogeneity [16-20]. Cost plays an essential part in the success of a research process as a higher fraction of the cost is allocated to participant recruitment [21]. The cost involved in research studies varies and is dependent on certain factors such as the targeted population, geographical location, and type of recruitment approach [18]. To overcome the challenges associated with the cost of participant recruitment, researchers need metrics to determine the cost of recruitment.

To access social media, users are required to create a profile that requires certain mandatory information such as first name and last name, email address, or mobile phone number [22]. Although interested social media users willingly provide these data, they are often unknowingly signing away their privacy, which increases the possibility of privacy breaches [23]. Although research ethics boards (REBs) require removing identifying information of research participants from data using unique identifiers, such guidelines are rendered ineffective in the context of social media data as participants' relational links are predictive of their attributes [24]. Nonetheless, Narayanan and Shmatikov [25] stated that such anonymization of participants' data might be insufficient to protect social media networks' privacy.

Researchers need guidance to navigate the ethical and logistical issues associated with using social platforms as a recruitment tool other than the "Terms and Conditions" stated by the application software providers [26]. Therefore, researchers often turn to ethics boards within their institutions for guidance on social media and internet recruitment; however, this information is not always readily available. To determine this, we reviewed the REBs of the top 10 higher education institutions in Canada to identify any standard ethical guidelines currently being used or recommended for using social media tools to recruit participants for research studies beyond adopting the Tri-Council Policy Statement on research. We used the QS World University Ranking criteria, which determine universities' rankings worldwide based on 6 metrics [27]. This strategy was deemed appropriate as these universities are known for their high-impact research productivity. The results revealed that only 3 universities had guidelines available on social media use in research studies, which further supported the need for this integrative review (Multimedia Appendix 1). A standard protocol that could be adopted by postsecondary institutions, research organizations, and researchers could help mitigate the pitfalls researchers encounter during participant recruitment for research via social media applications. Such protocols may facilitate the research process, expedite data collection, and ensure that digital research recruitment practices protect participants' data and rights.

Regarding the published literature, only 1 review [15] examined the evidence of cost, effectiveness, and the characteristics of participants recruited through Facebook compared with other web-based, social media, and traditional recruitment methods for adult health research. Little is known about the use of other social media platforms for participant recruitment in health research. Therefore, this study was warranted to address these gaps in the literature.

Objectives and Research Questions

This review sought to examine the evidence available on all the applications identified as social media tools and identify the best practices to facilitate participant recruitment through these tools. We addressed the following research questions:

1. What are the different social media tools commonly used by health science researchers for recruiting research participants and in what populations?
2. What is the proportion of nursing researchers who use social media platforms for recruitment?
3. What are the benefits and challenges of using social media to recruit research participants?
4. What are the best practices and ethical considerations for using social media tools to recruit research participants?

Methods

Overview

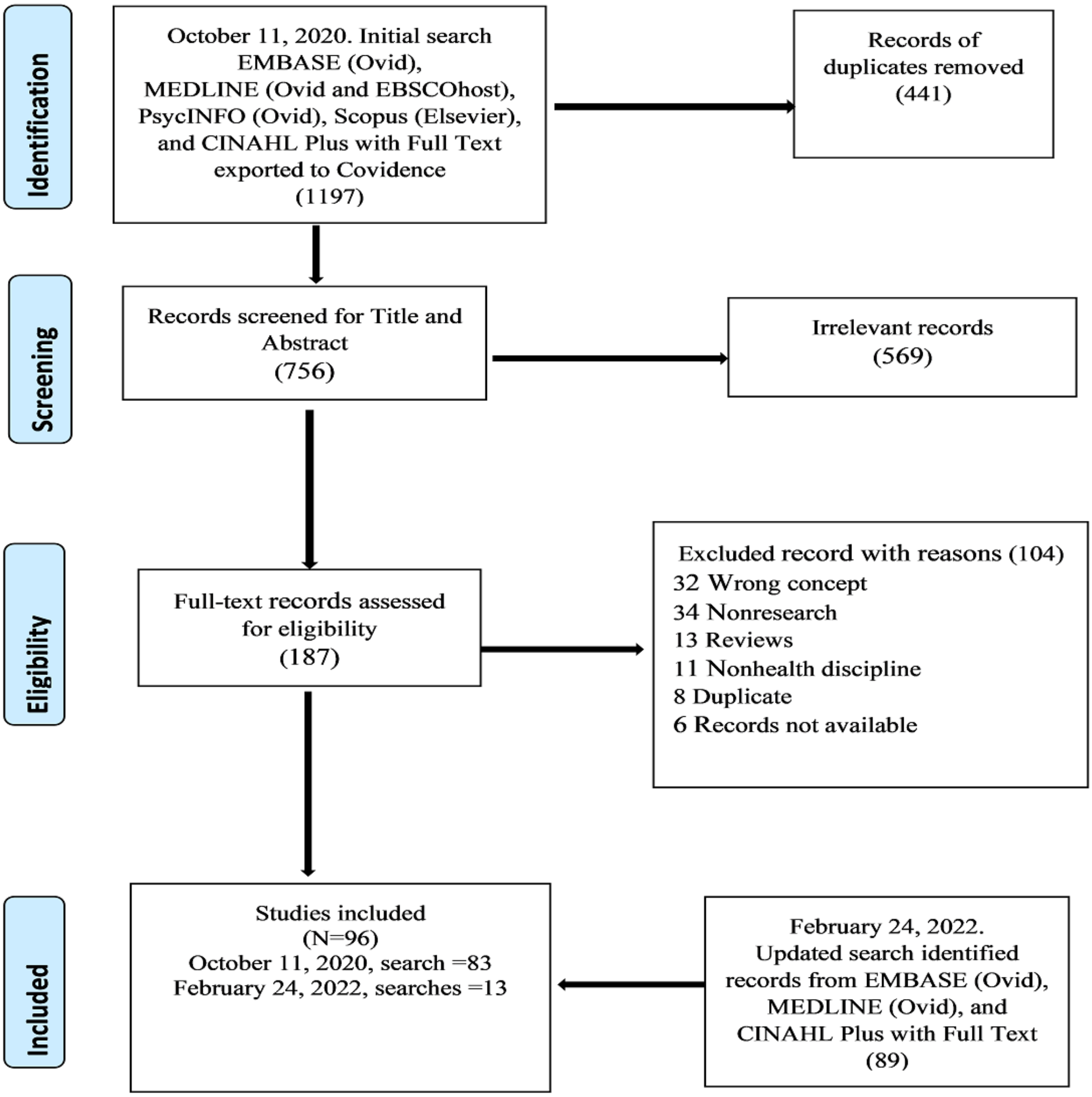
An integrative review guided by the Whittemore and Knafl [28] framework was conducted. This review type allows for the inclusion of diverse research methodologies and data sources to understand and generate new knowledge on the phenomenon of interest [28,29]. A comprehensive search strategy was formulated in consultation with a health science librarian. The initial search was conducted on October 11, 2020, and updated on February 24, 2022, in the EMBASE (Ovid), MEDLINE (Ovid and EBSCOhost), PsycINFO (Ovid), Scopus (Elsevier), and CINAHL Plus with Full Text (EBSCOhost) databases using a search strategy of keywords and subject headings through an iterative process (Multimedia Appendix 2). The criteria for eligibility were (1) all types of published research on primary and secondary studies, including qualitative, quantitative, and mixed methods; (2) discussion papers, white papers, reports, brief reports, specific guidelines, conference proceedings, dissertations, and published manuscripts that reported on social media use; (3) research reports published between January 2000 and February 2022; and (4) research reports that focused on research participant recruitment and advertisements on social media platforms, including all types of populations and health sciences disciplines, and (5) all geographical locations. The following articles were excluded: (1) non–English-language articles; (2) unpublished manuscripts and non–peer-reviewed publications such as descriptive papers, editorial papers, opinion papers, letters, book reviews, and article reviews; (3) review articles (scoping, integrative, narrative, and systematic) already published on the topic; and (4) all non–health sciences articles. The time frame for the published reports was chosen to capture the contemporary views that reflect the trends and popularity of digital platforms in participant recruitment.

Data Evaluation and Analysis

Records from the databases (initial search N=1197) were retrieved and imported into the Covidence Management Software for data screening and extraction. Overall, 2 reviewers (EMD and MK) independently conducted the screening process in Covidence, screening titles and abstracts, followed by full-text screening. All decisions made to either include or exclude records against the predetermined inclusion and exclusion criteria were documented. Where conflicts arose, the 2 reviewers consulted and resolved them through a voting process. We conducted another search on February 24, 2022, to update the results. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Protocols template in Covidence was used to map out the screening process, and the results as shown in Figure 1.

The following details were extracted from the included records: name or names of the author or authors, year of publication, country of publication, study design, study population, total number recruited, total number of participants enrolled or recruited through social media, social media platform used, other recruitment strategies, type of advertisement (paid or not paid), incentives provided, whether the study was funded, limitations of social media reported by the authors, and duration of advertisement (Multimedia Appendix 3 [2,18,20,30-115]). The extracted data from these records were analyzed by identifying codes and categories to characterize emerging themes, patterns, trends, and relationships to aid in synthesizing the findings logically and coherently. In addition, descriptive statistics were applied where appropriate to describe and summarize the data pertinent to the distribution of research and other characteristics. The Critical Appraisal Skills Program [116] was used per the research methodology to appraise and evaluate each of the included studies critically to ensure the quality of the available evidence included in this review (Multimedia Appendix 4 [18,20,31,50,53,54,55,58,60,61,76,77,81,106,109,114]). The studies were assessed and rated as “low quality” or “moderate quality” based on their theoretical or methodological rigor [28]. Ethics approval was not required as this study did not involve human participants.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 flow diagram.



Results

Overview

In total, 1197 records were retrieved from the initial database search; in Covidence, 441 (36.8%) duplicate records were removed from the imported references, and a total of 756 (63.2%) records were moved for screening. In the first stage of screening, the titles and abstracts of each record were screened for full-text review. In the second stage of screening, 187 full-text studies were reviewed entirely and assessed for eligibility for inclusion or exclusion. For records that were not available in the full text, the Health Science Library was contacted to obtain those records. A total of 6 articles that reported on conference proceedings were retrieved but did not have any substantial information, as reported in the abstracts. Nonetheless, the full texts of these articles were requested through library services but could not be retrieved; thus, a decision was made to exclude them from the results. It should be noted that although the Covidence software automatically removes duplicates, there were instances of errors where some records were missed; therefore, removing these duplicates manually was warranted. Of the initial search, a total of 83 records were included. The updated search returned a total of 89 records. Of the 89 articles, after 11 (12%) duplicates were removed, a total of 78 (88%) articles were screened. A total of 23 articles underwent full-text review, of which 10 (43%) were excluded as they did not meet the inclusion criteria, and 13 (57%) articles were retained and included in the review. Finally, 96 records were included in the review. In total, 114 records were excluded (Multimedia Appendix 5).

Characteristics of Included Studies

The range of publication years of the articles included in the review was between 2011 and 2022 (Multimedia Appendix 3). Most publications were from the United States (63/96, 66%) and Australia (20/96, 21%). Besides that, they were from Canada (5/96, 5%), the Netherlands (2/96, 2%), and the United Kingdom (2/96, 2%). There was only one publication from each of the following countries: Taiwan, Ecuador, India, and Brazil. Out of the 96 included studies, 92 (96%) were papers reporting primary research and 4 (4%) were reports on using social media. The methodological approaches used were cross-sectional studies (38/96, 40%), web-based surveys (15/96, 16%), secondary data analysis (14/96, 15%), randomized controlled trials (10/96, 10%), reports (4/96, 4%), mixed methods studies (4/96, 4%), qualitative studies (3/96, 3%), cohort studies (3/96, 3%), clinical trials (3/96, 3%), quasi-experimental studies (1/96, 1%), and longitudinal studies (1/96, 1%).

Social Media Use by Nursing Researchers and Other Health Researchers

We were interested in determining the proportion of nursing researchers using social media; however, this was not easy to identify as researchers have published studies that have used social media for the recruitment of research participants in a variety of interdisciplinary journals. On the basis of the journal names where these articles were published, these researchers could be from any health discipline, including nursing, medicine, psychology, rehabilitation, nutrition, pharmacy, or public health. Of the 96 included studies, 71 (74%) were published in general health science journals or interdisciplinary journals [18,20,26,30-92,117-121], and 25 (26%) were published in nursing-related journals [2,93-115,122].

Social Media Tools Commonly Used by Health Researchers

Researchers used a variety of social media platforms to recruit participants, as reported in the included studies. Researchers either exclusively used social media (43/96, 45%) or social media in conjunction with other recruitment methods (49/96, 51%) to recruit participants (Multimedia Appendix 3). For studies that exclusively used social media, ≥1 social media platform was used simultaneously. Social media platforms included Facebook, Twitter, Craigslist, Instagram, YouTube, LinkedIn, Reddit, Snapchat, and Tumblr. For social media in conjunction with other recruitment strategies, researchers used the identified social media platforms in addition to blogs, social media, Grindr, and WhatsApp Messenger. It was noted that, at times, researchers used the term “social media” but did not specify the type of social media used. In both approaches, most participants relied on the use of Facebook for research recruitment.

Age Groups of Research Participants Recruited via Social Media

Although researchers used social media platforms for advertisement and recruitment of participants for research, they sometimes did not target specific populations. In addition, the age group distribution of these populations varied, and the definition of the age group differed depending on the study aims. To address this, initially, the range of age groups was specified as follows: children (aged <9 years), adolescents (aged 10-18 years), young adults (aged 19-35 years), middle-aged adults (aged 36-55 years), older adults (aged 56-64 years), and older adults aged ≥65 years. The included studies were then scanned against this categorization to identify which age group was most targeted for social media recruitment (Table 1).

Table 1. Age groups of research participants recruited via social media (N=92).

Age category	Participants, n (%)	Research studies
≥18 years	28 (30)	[30,31,33-35,40,48,50,52,57,58,70,73-76,78,79,81-83,90,94,97,108,111,112,114,115]
Age group: not specified	22 (24)	[2,32,36,39,56,66,68,86,89,92,93,96,98-101,103,105-107,110,119]
Two age categories	18 (20)	[20,38,41,42,46,47,49,53,55,58,61,63,64,72,84,87,95,102]
Age groups between 3 age categories	9 (10)	[37,44,54,62,67,77,80,85,117]
Age groups between 4 age categories	3 (3)	[60,71,91]
Age groups between 5 age categories	3 (3)	[65,69,113]
Adolescents	2 (2)	[18,120]
Age groups of ≥21 years	3 (3)	[52,88,109]
≥30 years	1 (1)	[51]
≥33 years	1 (1)	[104]
≥45 years	1 (1)	[45]
≥60 years	1 (1)	[43]
Age group of ≤9 years	0 (0)	—[a]

[a] Not available.

Populations Targeted in Social Media Recruitment and Their Characteristics

Researchers have targeted different populations in their research studies. Largely, there were many studies (46/96, 48%) that included the general population [2,18,20,31,32, 36,37,40,41, 43,46,47,49, 51,54,55, 58,60,63,64,65, 67,69,70,71,72,73,74,77, 79,80,84, 85,91,92,93,94, 104,106,108, 112,114,115, 119,120]. In addition, a significant proportion (36/96, 38%) of the included studies focused on recruiting hard-to-reach populations (Table 2). Hard-to-reach populations are groups that are socially disadvantaged and present a challenge to access for researchers because of ethnicity, low income, or health literacy [123,124]. In this review, these populations had addiction problems, unique medical disease conditions, or lifestyle choices or belonged to an ethnic minority group. A few studies applied social media recruitment to target health care professionals as research participants (10/96, 10%) [96,98,99,100,101,103,105,106,107,113].

Table 2. Hard-to-reach populations targeted in social media recruitment.

Population	Records
Addiction: smoking and alcohol	[30,35,37,38,44,48,50,52,81]
Medical disease conditions: survivor of cancer, autism spectrum disorder, Lynch syndrome, people living with HIV, asthma, obstructive pulmonary disease, depression, and kidney transplant recipient	[34,39,42,53,56,58,61,66,68,89,97,107,111]
Lifestyle: men who have sex with men	[33,57,78,82,83,90]
Ethnic minorities: low-income and racial minority	[45,62,75,76,86,88,95,102]

Cost-effectiveness, Efficacy, and Feasibility of Social Media in Comparison With Other Recruitment Methods

The costs of recruitment reported in all the included studies are presented in tabular form (Tables 3 and 4) to help ascertain how money was dispensed, as well as the cost-effectiveness of each recruitment strategy. However, it was noted that although some researchers included personnel costs, advertising costs, and other recruitment costs in the total cost, other researchers did not include these costs. Hence, researchers should use only this information as a guide. Few studies specifically compared social media effectiveness to other platforms with the goal of establishing cost-effectiveness, efficacy, and feasibility. In this review, 7% (7/96) of the included studies aimed to determine the effectiveness of social media compared with other recruitment strategies [39,51,52,58,70,74,106], and 9% (9/96) of studies that did not compare social media platforms with any other strategy [30,38,41,46,63,72,73,93,111] found social media as an effective recruitment strategy in both instances.

Few studies did not conclusively find social media to be cost-effective or efficient. The findings reported by these researchers differ because of the different populations targeted, scale of recruitment, and whether the research was funded. In their funded study, Moreno et al [18] found that in-person strategies yielded more participants in a geographic area at a lower cost than social media, and the cost per enrollee by social media was higher than that of traditional methods. In addition, Frandsen et al [48] suggested that Facebook was cost-effective in obtaining eligible participants at the initial stage of the recruitment process. The mailing of letters was cost-effective compared with Facebook, according to Waltman et al [106].

Table 3. Cost of social media recruitment methods.

Study and social media	Cost per person	Total cost
Ahmed et al [39]		
Facebook	US $8.73	US $2950
Wilkerson et al [82]		
Social media advertisements, posts, and email blasts	US $40.44	US $3033.11
Social media posts and website banner advertisements	US $15.86	US $1380
Social media posts and film festival entrance wavier	US $2.78	US $6297.66
Gioia et al [52]		
Craigslist	US $1.46	US $275
Musiat et al [64]		
Facebook	Aus $105.77 (US $73.07)	—[a]
Twitter	Aus $422.03 (US $292.08)	—
YouTube	Aus $81.31 (US $56.27)	—
Frandsen et al [50]		
Facebook	Aus $42.34 (US $29.30)	Aus $5842.30 (US $4043.35)
Byaruhanga et al [35]		
Facebook	Aus $61.68 (US $42.69)	Aus $33,738.52 (US $23349.83)
Twitter	Aus $61.52 (US $42.58)	Aus $61.52 (US $42.58)
Harris et al [55]		
Facebook	—	Aus $28,571.54 (US $19,773.86)
Moreno et al [18]		
Social media (Facebook, Twitter, and blogs)	US $40.99	—
Wilkerson et al [83]		
Facebook and Twitter	—	Free
Guthrie et al [54]		
Facebook	—	US $14,825
Waltman et al [106]		
Facebook	US $119.38	US $5252.83
Watson et al [81]		
Facebook	US $40.51 (randomized)	US $49,791.49
Derrick et al [44]		
Facebook	US $498 per couple	US $10,966
Facebook	US $181 per couple	US $4145
Carter-Harris et al [117]		
Facebook	—	US $500
Jones et al [95]		
Facebook	US $66.46 (randomized)	—
Frandsen et al [48]		
Facebook	Aus $56.34 (US $38.99)	Aus $5183.33 (US $3587.29)
Juraschek et al [58]		
Facebook	US $794	US $5704
Facebook	US $1426	US $2383
Iott et al [57]		

Study and social media	Cost per person	Total cost
Grindr	US $87.35	US $1747.40
Scruff	US $69.30	US $207.90
Facebook	US $149.90	US $170.69
Facebook groups	—	US $10.40
van Gelder et al [77]		
Facebook	€10.88 (US $11.13)	€315.52 (US $322.89)
Facebook	€9.48 (US $9.70)	€284.48 (US $291.13)
Alley et al [70]		
Untargeted Facebook	Aus $68 (US $47.06)	Aus $1438 (US $995.21)
Targeted Facebook	Aus $42 (US $29.07)	Aus $7721(US $5343.57)
Gilligan et al [51]		
Facebook	Aus $5.94 (US $4.11)	Aus $1107 (US $766.14)
Barney et al[b] [84]		
Facebook and Instagram	US $42.21	US $21,867
Moseson et al[b] [85]		
Facebook	US $49.48	—
Reddit	US $182.78	—
Salvy et al[b] [20]		
Facebook	US $334	US $9020
Stuart and Moore[b] [96]		
Facebook	US $1.78	US $952.81
Cho et al[b] [97]		
Facebook	—	US $120,000
Facebook	—	US $215
Spahrkäs et al[b] [89]		
Facebook	€22.42 (US $22.94)	€17,000 (US $17,397.12)

[a] Not available.

[b] Records from the updated search.

Table 4. Cost of other recruitment methods.

Studies and other strategies	Cost per person	Total cost
Ahmed et al [39]		
Radio	—[a]	US $12,030
Wilkerson et al [82]		
Website banner advertisements	US $172.50	US $1380
Gioia et al [52]		
Print newspaper	US $116.88	US $33,311
Musiat et al [64]		
Recruitment agency	Aus $100 (US $69.21)	—
Google advertisements	Aus $195.83 (US $135.53)	—
Frandsen et al [50]		
Newspaper advertisements	Aus $21.52 (US $14.89)	Aus $2065.46 (US $1429.47)
Byaruhanga et al [35]		
Gumtree	Aus $7.29 (US $5.05)	Aus $36.43 (US $25.21)
Web promotions and internet searches	Aus $43.76 (US $30.29)	Aus $437.56 (US $302.83)
Emails	Aus $128.67 (US $89.05)	Aus $2315.98 (US $1602.85)
Newspaper	Aus $50.28 (US $34.80)	Aus $2363.38 (US $1635.65)
Radio (interviews)	Aus $102.78 (US $71.13)	Aus $205.55 (US $142.26)
Magazine	Aus $85.41 (US $59.11)	Aus $170.81 (US $118.21)
Posters	Aus $566.65 (US $392.17)	Aus $566.65 (US $392.17)
Flyers	Aus $2546.29 (US $1762.24)	Aus $2546.29 (US $1762.24)
Telephone	Aus $3990.84 (US $2761.99)	Aus $3990.84 (US $2761.99)
Harris et al [55]		
Access to organizational websites	—	Aus $5890 (US $4067.96)
Posters	—	Aus $195 (US $134.96)
Face-to-face events	—	Aus $43,000 (US $29,698.17)
Conference promotion	—	Aus $44,040 (US $30,416.45)
Moreno et al [18]		
In person	US $19.09	—
Wilkerson et al [83]		
Mobile banner advertisements	US $375	US $3000
Browser banner advertisements	US $187.50	US $1500
Guthrie et al [54]		
Mailings	US $356 per randomized participant	US $98,682
Waltman et al [106]		
Provider letter	US $29.36	US $1703
Postcards	US $926.96	US $43,567.49
Newspaper advertisements and television interviews	US $330.12	US $1650.63
Watson et al [81]		
Press releases	—	US $1995
Mailed letters	—	US $4054
Google advertisements	US $34.71 (randomized)	US $3506
Web-based survey company	—	US $7644

Studies and other strategies	Cost per person	Total cost
Derrick et al [44]		
Targeted mailing	US $303 per couple	US $3635
Carter-Harris et al [117]		
Newspaper advertisements	—	US $1224
Jones et al [95]		
On ground	US $149.62 (randomized)	—
Frandsen et al [48]		
Newspaper advertisements	Aus $52.33 (US $36.22)	Aus $4343.10 (US $3005.78)
Juraschek et al [58]		
Mailed brochure	US $799	US $51,950
Periodicals	US $437	US $10,906
Iott et al [57]		
Email groups	US $10.40	US $62.37
Personal networking	US $10.40	US $20.79
Unified staff	US $30.32	US $727.65
Bar outreach	US $1621.62	US $1621.62
Flyer per palm card	US $83.20	US $416
Publishing article in newsletter	—	US $20.79
van Gelder et al [77]		
Google AdWords	€54.28 (US $55.52)	€325.66 (US $333.12)
Care providers	—	—
Alley et al [70]		
Google AdWords	Aus $495 (US $342.58)	Aus $495 (US $342.58)
Posters	Aus $52 (US $35.99)	Aus $574 (US $397.26)
Health care leaflets	Aus $66 (US $45.68)	Aus $990 (US $685.16)
Letterbox drop	Aus $135 (US $93.43)	Aus $2425 (US $1678.30)
Newspaper advertisement	Aus $145 (US $100.35)	Aus $726 (US $502.45)
Community calendar	Aus $12 (US $8.30)	Aus $70 (US $48.45)
Newspaper article	Aus $3 (US $2.08)	Aus $53 (US $36.68)
Gilligan et al [51]		
Social networks, flyers, websites, posters, recruitment cards, email, and media coverage	Aus $58.70 (US $40.63)	Aus $4349 (US $3009.87
Barney et al[b] [84]		
Clinic-based and in person	US $865.93	US $102,180
Moseson et al[b] [85]		
Google advertisements	US $265.93	—
Salvy et al[b] [20]		
Targeted mailings	US $217	US $2387
In-person recruitment	US $290	US $11,328
Stuart and Moore[b] [96]		
Association journal	US $375.00	—
Cho et al[b] [97]		

Studies and other strategies	Cost per person	Total cost
Personal outreach	—	—
Public outreach	—	US $1686.04
Spahrkäs et al[b] [89]		
—	—	—

[a] Not available.

[b] Records from the updated search.

Best Practices and Strategies Used to Enhance Social Media Recruitment

Diverse advertisement strategies are adopted by researchers when recruiting research participants through social media platforms. Each social media platform advertisement differs in specification, advertisement content, word count, and design language [47]. In the included studies, researchers identified and used one or multiple paid, targeted advertisement campaigns with different themes to reach potential participants on various platforms within a specific advertisement duration. Some models of advertisement included the use of paid targeted advertisement [18,30,31, 37,39-44, 46-52,54, 58-60,63, 65,69-75, 77,78,80, 81,91,93, 95,102,109, 113,114,117,119,120]. In addition, some researchers used untargeted advertisements [38,62,111], untargeted but paid advertisements [45,74], "boosted" posts [94,106], posts [34,61,76,90,105,110,115], both advertisements and posts [55,57,64,68], tweets [79,104,108], targeted advertisements and posts [36], advertisements [56], blasts [33], paid and unpaid social media channels [32,35], and messengers [92] to strategically advertise and recruit their potential participants.

Another identified strategy was the use of cost-related model strategies to determine the cost of the advertisements. Researchers who are engaged with any social media platform to advertise and recruit participants are billed by cost per click, cost per thousand impressions, cost per view, or cost per action or per conversion [125]. With the cost per click model, researchers are billed when a potential participant clicks on the advertisement. This approach was one of the most preferred models for researchers in the included studies used for advertising. The cost per click model budget is set at a daily, weekly, or lifetime spending limit depending on the researcher's choice [18,30,33,38, 40,41,46, 47,48,51, 54,55,59, 60,64,69, 71,73,74, 75,77,80, 93,102, 111,126,129].

Researchers also noted considerations related to the display and design of an advertisement for a desktop application, which differed from that of a mobile app, and this affected how participants viewed and reacted to the recruitment advertisement. The displayed advertisements targeted either the user's browser or the newsfeed [91,117]. For Facebook, the advertisement is displayed on the user's web browser [58] or on the right-side panel of the Facebook newsfeed or placed directly in the newsfeed [47,55,65]. For Instagram, images are displayed in a linear format. Snapchat images are displayed using the story feature [47], and on Grindr, advertisements are displayed as pop-ups [57]. Therefore, cost is an influencing factor that determines the placement of the advertisement and the social media application of choice, thereby influencing the decisions that researchers make regarding recruitment.

Some researchers identified ethical challenges inherent to social media recruitment, such as privacy, confidentiality, and informed consent, and provided strategies to minimize the challenges for the researcher and the potential participant. The strategies offered and reported in the included studies included the use of a study-specific page for recruitment [2,51,76,78,82,91,94,107,112,113] and the use of secure landing sites or study webpages for data collection [35,43,48,50,58,75,106,118]. In addition to the strategies proposed by health researchers, there were 4 reports identified in the included studies that outlined guidelines and recommendations for social media recruitment (Table 5).

Table 5. Recommendations for best practices on social media recruitment.

Study	Study purpose or aim	Key findings	Recommendation
Curtis [118] (the United States)	To outline ethical challenges associated with social media recruitment	Social media platforms are challenged with issues of confidentiality, informed consent, and privacy issues.	Recruiting participants through secure landing sites; researchers regularly reviewing social media websites for regular updates; verifying participants' age through cross-checking with other information may provide solutions to the challenges; setting web-based quizzes to test participant competency; and providing a summary of the research study via email
Kamp et al [122] (the United States)	To examine and describe the challenges of the Facebook recruitment method and provide recommendations	The Facebook platform presents an inherent challenge with privacy, data security, and recruiting participants.	Researchers can implement a multifactor authentication process to access research data and regularly review the privacy settings and policies on social media sites
Gelinas et al [26] (United States)	To examine the norms governing social media recruitment and analyze the ethics of recruiting, and the implication of web-based communication	The foundational norms in research ethics include respect for persons, beneficence, and justice; however, in social media, the key norms governing social media recruitment include respect for privacy and researcher's transparency. The lack of regulatory guidance on ethics in social media recruitment poses a risk for ethical issues.	The authors proposed a checklist that researchers can use in social media recruitment
Bender et al [121] (Canada)	To develop a framework on ethics and privacy for social media and internet recruitment	The Privacy by Design framework evaluates the privacy strengths, thereby providing privacy protection in web-based recruitment.	Adhering to the Privacy by Design framework, which provides privacy-enhancing measures such as developing privacy notices, disabling comment features, or monitoring comments and removing identifiable information before it becomes public

Discussion

Principal Findings

Although we intentionally excluded 13 reviews from this study, 7 (54%) of them are discussed here against the findings of our review. The results from our review show an increased interest in using social media for research recruitment by researchers from different health disciplines in which social media strategies have fulfilled researchers' recruitment needs. Considering the wide range of publications, the scope of this literature review, and the social media applications examined in this review, it can be concluded that the use of social media is on the rise, as evidenced by the increase in the number of publications in the past few years. The different research methods identified in the included studies suggest increased use of social media for a variety of research methods. Notably, a few of the included studies recruited participants for clinical studies, with most recruiting participants for cross-sectional studies. Despite social media's reach within a broad geographic location, health researchers are still challenged with participant recruitment for clinical trials. This suggests that social media may be best suited for recruiting participants for noninterventional studies. Researchers recruiting for clinical trials may have to diversify their recruitment strategies to reach their recruitment goals until a comprehensive strategy to navigate social media platforms is established. This finding is similar to that of the review by Topolovec-Vranic and Natarajan [127], which found that only a few studies used social media to recruit participants for interventional studies as opposed to observational studies. Although the Topolovec-Vranic and Natarajan review [127] used a smaller number of studies to draw this conclusion, their findings are still significant, considering this review.

Researchers from different health disciplines, including nursing, medicine, public health, mental health, and pharmacy, have used social media for recruitment and have published their findings in a variety of journals. Some of these journals are discipline specific or interdisciplinary. This suggests different avenues for health researchers to publish their work. Within the nursing discipline, nurses are increasingly using social media for the recruitment of research studies, as published in multiple nursing and nonnursing journals. However, there are opportunities to continue promoting the use of social media among nurses for research and educational purposes.

On the basis of this review, researchers used different social media applications to advertise and recruit potential research participants. The preferred social media applications were Facebook, Instagram, Snapchat, LinkedIn, Twitter, Grindr, Reddit, Tumblr, WhatsApp, Craigslist, YouTube, and blogs to be used either solely to recruit or in conjunction with other recruitment strategies to achieve recruitment and study goals. Owing to its popularity among users and global penetration, Facebook was the most widely used application among researchers. Different social media applications enabled researchers to recruit participants with different demographics and characteristics. For instance, Facebook was used to recruit younger participants [77,113] and older individuals [67], whereas other researchers recruited young people through Tumblr [61]. This finding is similar to that of Arigo et al [128], who identified web-based platforms such as social networking sites (Facebook, Twitter, Instagram, Tumblr, and LinkedIn) as some of the common platforms that health researchers use to recruit research participants, including a diverse population for their research studies. In addition, researchers used multiple approaches and strategies to recruit participants. An approximately equal number of participants were recruited through social media alone or social media in conjunction with other strategies. This finding agrees with the general literature on the increasing acceptance of digital platforms for recruitment and with some health researchers using social media and traditional methods [129] for recruitment. In addition to the findings in this review, reviews conducted by some researchers [15,127,130-134], although focusing on only one social media application or using the term “social networking sites” broadly in their research, exclusively and comprehensively reported on a wide range of different social media applications used in research recruitment.

The different social media platforms used to target the different groups of populations such as the general population, hard-to-reach populations, and specialized populations, depended on the research aim. The hard-to-reach populations included people with addictions, sensitive health issues, ethnic groups, and poor and stigmatized populations [135]. Social media was found to be effective in reaching and recruiting hard-to-reach potential participants who were otherwise unreceptive to traditional recruitment methods because of their conditions and representations within their communities and society [50]. Researchers must weigh all available options to determine the best approach to proceed when recruiting from these populations.

The age group distribution of the research participants included in this review spanned different age categories. As shown in Table 1, the most targeted population from an age perspective was young adults. According to Kemp [10], the engagement of social media platforms among youth stands at 58% between the ages of 16 and 24 years. This is not surprising because of the acceptability of social media among youth who are considered technology savvy and their tendency to use social media regularly. As such, targeting such an age group for research studies can lead to increased participation. An observation of interest in this review is the low involvement of children and adolescents aged <18 years and older adults. Only 2% (2/96) of studies [18,120] involved adolescents between the ages of 13 and 14 years. As researchers require parental consent among the children and the adolescent group, research studies involving these groups are relatively limited. This finding is similar to that of Amon et al [130], who suggested that instead of focusing on adolescents who require parental consent, targeting parents or guardians of the intended group could help waive parental consent.

For older adults, the usability of social media platforms presents a challenge, such as platform design and content, as these platforms are tailored to the interests of the younger population [136]. Owing to the complex design, nature, and privacy-related concerns associated with social media platforms, older adults are more comfortable and familiar with traditional forms of recruitment than social-mediated platforms [137]. Other barriers encountered by older adults include intrapersonal, interpersonal, functional, and structural elements that hinder the use of social media platforms [138]. Although social media presents a challenge for recruitment in the older population, researchers can continue to explore traditional methods in such populations to offer an equal chance of participation in research studies. The trade-off between using traditional methods and social media for recruiting research participants is a complicated issue, requiring health researchers to weigh options and the benefits and risks to the participant and the research study, as well as more creative ways of engaging low participating groups.

There is a debate on the cost-effectiveness of social media in the literature. Some studies found the social media method to be cost-effective, whereas other studies disagree with this assertion [18,58]. In this review, the cost of other recruitment strategies compared with social media recruitment strategies was presented as part of this review to assist researchers in making an informed decision (Tables 3 and 4). In addition, in this review, the factors that influenced the cost associated with recruitment varied from one study to the next. Some researchers reported advertising, recruitment, and other administrative costs as the total cost, whereas others reported only aspects of social media advertisement and recruitment as the total cost. Owing to the inconsistency in cost reporting, having a standardized cost reporting system to maintain consistency would help to effectively determine whether social media recruitment is cost-effective. On the basis of the analysis of the cost-effectiveness of both social media and other recruitment strategies, this review found that social media was viewed by researchers as a cost-effective strategy. Although 28% (27/96) of studies in this review reported on the cost of social media compared with other recruitment methods, not all researchers found social media as a cost-effective method. Nonetheless, given that a large proportion of these studies found social media to be cost-effective, this review supports this conclusion. Compared with the previously published reviews by Reagan et al [15] and Topolovec-Vranic and Natarajan [127], this review provides additional insights and includes a broader range of studies. This review captured additional literature not included in the review by Reagan et al [15], which relied only on 18 articles, of which only 10 articles reported on cost. In the review by Topolovec-Vranic and Natarajan [127], the authors included 30 studies, of which 5 reported on cost-effectiveness, and 7 did not find social media to be a cost-effective method. The findings also revealed that the cost of recruitment for hard-to-reach populations differs from that for the general population. Juraschek et al [58] found that recruiting through Facebook advertisements for the African American population was costly; however, advertisements were effective in directing eligible participants to the website. Hence, researchers hoping to recruit research participants through social media must consider these factors to make decisive choices on the most suitable method for recruitment.

Best Practices for the Use of Social Media in Recruitment of Research Participants

Researchers are increasingly tapping into the available opportunities to use social media platforms for their research studies. However, there is a need for best practices to guide this process. To adequately explore and navigate social media platforms successfully for recruitment, adhering to best practices, including those of ethical considerations (informed consent, privacy, confidentiality, and transparency) that protect the researcher and participants, is of utmost importance [26,118,122]. In the review by Amon et al [130], the authors found that participants recruited on web-based platforms were subjected to the same ethical standards as though they were responding to a traditional recruitment method. In that regard, Gelinas et al [26] were of the view that REBs should standardize social media techniques by clarifying their similarity to traditional recruitment. Furthermore, the findings from the review also establish the need to take additional steps to make available informed consent through other means, where the potential participants are well informed with detailed information about the research study before participation. To curb and curtail the complexities and complicated nature of informed consent, the findings from this review support the recommendations suggested by Herbell and Zauszniewski [94] and Stokes et al [105] in their studies to make an information sheet in a downloadable version available for participants and send web-based consent forms to potential participants after meeting the eligibility criteria. To maintain the confidentiality of both researchers and participants, Shaver et al [71] suggested using anonymous surveys and directing interested participants through a survey link to a landing page for study information. Researchers are discouraged from directly recruiting participants on social media platforms but instead using the social media platform to advertise, as the confidentiality and privacy of participants' data cannot be guaranteed. To further ensure the provision of privacy, Bender et al [121] used privacy-enhancing measures aligned with the principles of Privacy by Design by disabling the comment feature, developing privacy notices for social media campaigns, sending disclaimers about the privacy risks of social media pages, and building privacy protection into the recruitment strategy. Although the tenets of the foundational principles were incorporated to avoid privacy-related issues, Bender et al [121] were of the view that the principles of transparency and user-centric options of Privacy by Design provide inadequate guidance on how to design privacy notices using these key principles.

In addition, some factors were identified to influence advertisements, such as advertisement targets, crafting of multiple advertisement campaigns with different wordings and themes, rotating and alternating advertisements, payment model, duration of the advertisement, and location of the advertisement on the social media platform. To favorably achieve the results of recruiting an increasing number of participants for research, researchers advertising on social media must strategically reach out to their participants. On the basis of the findings of this review, using an appealing image and simple and consistent language through both the text caption and image [69] influences and attracts participants to the study. Some social media platforms' advertising policies provide details on advertisement content, including the choice of words and counts and the duration of an advertisement on their platforms. The advertisement policies differ from platform to platform. Researchers must research any platform they wish to engage in, understand the policies, and adhere to them. In addition, working with REBs on social recruitment messages and strategies helps avoid ineffective strategies and enhance ethical conduct. Incorporating prescreening questions before allowing participants to enter details for study participation reduces the rate of ineligible participants and maximizes the reach and sample representativeness. Researchers can use these applications simultaneously because of the feasibility of incorporating social media platforms such as Instagram, Snapchat, and Facebook into a study without difficulty [47].

Implications

The findings from the review show the increasing accessibility and multifunctionality of social media platforms that could be leveraged to further support health science research. In fact, one of the benefits of social media for conducting research recruitment has been amplified during the COVID-19 pandemic because of the limitations to in-person recruitment, thus sustaining the continuity of research.

Generally, social media platforms provide avenues for a practical approach to reaching diverse, extensive, and targeted audiences [139] or populations, particularly those that are hard to reach. Further research may be needed to understand the barriers to and facilitators of older adults' engagement with social media platform recruitment.

Although different approaches to recruitment, advertisement, cost determination, and efficiency reporting can challenge novice researchers planning to use social media, there are ways of mitigating some of these challenges. For example, with the availability of funds and resources, researchers can benefit from hiring specialized companies or third-party service organizations to assist with the marketing and development of social media recruitment strategies and other innovative recruitment approaches targeting potential research participants. It is also recommended that these strategies be discussed and coordinated with the researcher's academic institution's REB to ensure no risks to participants.

The lack of explicit regulations by REBs to guide researchers continues to prevent the full exploration of social media platforms to support health science research. As such, stakeholders and collaborative efforts from research-based organizations, academia, researchers, think tanks, and student groups must partner to develop guidelines that reflect the use of social media in research studies. The different guidelines developed and published by researchers and academic institutions can provide a context for what is available. Therefore, based on our review, we propose a tentative description or guideline to guide researchers based on what we have synthesized from the literature included in this review (Multimedia Appendix 6). Ultimately, this guide could serve as a starting point to inform stakeholders in the development of a standardized protocol to guide health science researchers in the use of various social media platforms for research participant recruitment.

Finally, there are opportunities to advance health science education regarding social media use in general and its use for the recruitment of research participants. As students become technologically savvy, incorporating social media into their learning process will allow them to effectively engage with the platform. Schools can also provide guidelines on social media platforms on their websites to enhance learning about their applications in research processes. In addition, teaching students about best practices that support professional social media use and including social media applications as part of ethics training programs are also recommended.

Strengths and Limitations

The findings of this review offer a broad perspective on the use of social media platforms for participant recruitment by health researchers. A large number of studies were included for analysis in this review. The timelines for the included studies span >20 years and provide sufficient time to capture all studies published during the popularity of social media. This study comprehensively synthesized available literature from all health science disciplines. However, the review was limited to studies reported only in English.

Conclusions

The purpose of this integrative review was to explore the literature on recruiting participants for research studies through social media application tools and identify best practices to assist researchers in conducting research participant recruitment via social media tools. This integrative review expanded on the review by Reagan et al [15], which focused primarily on Facebook, by including other social media applications used by health researchers to recruit research participants, such as Facebook, Craigslist, Instagram, LinkedIn, Reddit, Tumblr, Twitter, and YouTube. Overall, the findings showed that social media is a suitable, viable, and cost-effective channel for recruiting research participants, despite some challenges associated with its use. Health researchers are increasingly embracing various social media platforms in their research to recruit participants from various age groups and diverse backgrounds; however, there is less use of social media to recruit older adults. Adhering to best practices when targeting various populations through social media advertisements is vitally important to protect participants' and researchers' rights and increase participation. REBs must proactively provide protocols and best practice guidelines that researchers can apply during the advertisement and recruitment of research participants.

Acknowledgments

The authors acknowledge the assistance of Janice Kung, Laura Hamonic, and Allison Sivak of the University of Alberta Health Sciences Library for their support during the literature search.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Top 10 Universities in Canada QS World University Rankings 2021.
[DOCX File , 17 KB - jmir_v24i8e38015_app1.docx]

Multimedia Appendix 2
Search strategy for keywords and subject headings of the search terms used in each database.
[DOCX File , 31 KB - jmir_v24i8e38015_app2.docx]

Multimedia Appendix 3
Characteristics of included studies.
[DOCX File , 72 KB - jmir_v24i8e38015_app3.docx]

Multimedia Appendix 4
Critical Appraisal Skills Program critical appraisal tools used to critically appraise and evaluate each of the included studies.
[DOCX File , 26 KB - jmir_v24i8e38015_app4.docx]

Multimedia Appendix 5
List of excluded studies.
[DOCX File , 37 KB - jmir_v24i8e38015_app5.docx]

Multimedia Appendix 6
Tentative guidelines to assist health sciences researchers in social media recruitment.
[DOCX File , 21 KB - jmir_v24i8e38015_app6.docx]

References

1. Raju J. Knowledge and skills for the digital era academic library. J Acad Librarianship 2014 Mar;40(2):163-170. [doi: 10.1016/j.acalib.2014.02.007]

2. Johnson LA, Dias N, Clarkson G, Schreier AM. Social media as a recruitment method to reach a diverse sample of bereaved parents. Appl Nurs Res 2019 Dec;50:151201. [doi: 10.1016/j.apnr.2019.151201] [Medline: 31677928]
3. Collins K, Shiffman D, Rock J. How are scientists using social media in the workplace? PLoS One 2016;11(10):e0162680 [FREE Full text] [doi: 10.1371/journal.pone.0162680] [Medline: 27732598]
4. Darmawan I, Bakker C, Brockman TA, Patten CA, Eder M. The role of social media in enhancing clinical trial recruitment: scoping review. J Med Internet Res 2020 Oct 26;22(10):e22810 [FREE Full text] [doi: 10.2196/22810] [Medline: 33104015]
5. Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, et al. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. Health Technol Assess 2007 Nov;11(48):iii, ix-iii,105 [FREE Full text] [Medline: 17999843]
6. Dewing M. Social media. 1, An introduction. In: Background paper (Canada. Parliamentary Information and Research Service). Ottawa: Library of Parliament; 2010.
7. Gruzd A, Jacobson J, Mai P, Dubois E. The state of social media in Canada 2017. SSRN J 2017. [doi: 10.2139/ssrn.3158771]
8. Chamie J. World Population: 2020 overview. YaleGlobal Online. 2020. URL: https://archive-yaleglobal.yale.edu/content/world-population-2020-overview# [accessed 2020-02-11]
9. Social media: statistics and facts. Statista. URL: https://www.statista.com/topics/1164/social [accessed 2020-05-18]
10. Digital 2020: global digital overview. We are Social. URL: https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media [accessed 2020-08-10]
11. Digital 2020. Hootsuite. URL: https://hootsuite.com/pages/digital-2020#c-204634 [accessed 2020-09-10]
12. Dhar VK, Kim Y, Graff JT, Jung AD, Garrett J, Dick LE, et al. Benefit of social media on patient engagement and satisfaction: results of a 9-month, qualitative pilot study using Facebook. Surgery 2018 Mar;163(3):565-570. [doi: 10.1016/j.surg.2017.09.056] [Medline: 29195735]
13. Kashif Imran M, Fatima T, Aslam U, Muhammad Javed Iqbal S. Exploring the benefits of social media towards knowledge sharing among doctors. Pakistan J Psychol Res 2019 Jul 23;34(2):331-352. [doi: 10.33824/pjpr.2019.34.2.18]
14. Vuori M. Exploring uses of social media in a global corporation. J Syst Inf Technol 2012;14(2):155-170. [doi: 10.1108/13287261211232171]
15. Reagan L, Nowlin SY, Birdsall SB, Gabbay J, Vorderstrasse A, Johnson C, et al. Integrative review of recruitment of research participants through Facebook. Nurs Res 2019;68(6):423-432. [doi: 10.1097/NNR.0000000000000385] [Medline: 31693547]
16. Kapse A, Patil V, Patil NV. E-recruitment. Int J Eng Adv Technol 2012;1(4):82-86 [FREE Full text]
17. Arnobit CI, Loo K, Pagano I, Uchiyama M, Fukui J, Braun-Inglis C, et al. Recruiting cancer survivors to a mobile mindfulness intervention in the United States: exploring online and face-to-face recruitment strategies. Int J Environ Res Public Health 2021 Sep 27;18(19):10136 [FREE Full text] [doi: 10.3390/ijerph181910136] [Medline: 34639439]
18. Moreno MA, Waite A, Pumper M, Colburn T, Holm M, Mendoza J. Recruiting adolescent research participants: in-person compared to social media approaches. Cyberpsychol Behav Soc Netw 2017 Jan;20(1):64-67. [doi: 10.1089/cyber.2016.0319] [Medline: 27976951]
19. Robinson JL, Fuerch JH, Winiewicz DD, Salvy SJ, Roemmich JN, Epstein LH. Cost effectiveness of recruitment methods in an obesity prevention trial for young children. Prev Med 2007 Jun;44(6):499-503 [FREE Full text] [doi: 10.1016/j.ypmed.2007.03.004] [Medline: 17475318]
20. Salvy S, Carandang K, Vigen CL, Concha-Chavez A, Sequeira PA, Blanchard J, et al. Effectiveness of social media (Facebook), targeted mailing, and in-person solicitation for the recruitment of young adult in a diabetes self-management clinical trial. Clin Trials 2020 Dec;17(6):664-674 [FREE Full text] [doi: 10.1177/1740774520933362] [Medline: 32627589]
21. Huynh L, Johns B, Liu S, Vedula SS, Li T, Puhan MA. Cost-effectiveness of health research study participant recruitment strategies: a systematic review. Clin Trials 2014 Oct;11(5):576-583. [doi: 10.1177/1740774514540371] [Medline: 24980278]
22. Shozi N, Mtsweni J. Big data privacy in social media sites. In: Proceedings of the 2017 IST-Africa Week Conference (IST-Africa). 2017 Presented at: 2017 IST-Africa Week Conference (IST-Africa); May 30-Jun 02, 2017; Windhoek, Namibia. [doi: 10.23919/istafrica.2017.8102311]
23. Custers B, van der Hof S, Schermer B, Appleby-Arnold S, Brockdorff N. Informed consent in social media use – the gap between user expectations and EU personal data protection law. scrip 2013 Dec 17;10(4):435-457. [doi: 10.2966/scrip.100413.435]
24. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/AJPH.2017.304249] [Medline: 29346005]
25. Narayanan A, Shmatikov V. De-anonymizing social networks. In: Proceedings of the 2009 30th IEEE Symposium on Security and Privacy. 2009 Presented at: 2009 30th IEEE Symposium on Security and Privacy; May 17-20, 2009; Oakland, CA, USA. [doi: 10.1109/sp.2009.22]
26. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
27. QS World University Rankings: methodology. QS Top Universities. URL: https://www.topuniversities.com/subject-rankings/methodology [accessed 2020-03-04]

28. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
29. Badu E, O'Brien AP, Mitchell R. An integrative review on methodological considerations in mental health research - design, sampling, data collection procedure and quality assurance. Arch Public Health 2019 Oct 10;77(1):37 [FREE Full text] [doi: 10.1186/s13690-019-0363-z] [Medline: 31624592]
30. Bold KW, Hanrahan TH, O'Malley SS, Fucito LM. Exploring the utility of web-based social media advertising to recruit adult heavy-drinking smokers for treatment. J Med Internet Res 2016 May 18;18(5):e107 [FREE Full text] [doi: 10.2196/jmir.5360] [Medline: 27194456]
31. Burgess JD, Kimble RM, Watt K, Cameron CM. The adoption of social media to recruit participants for the cool runnings randomized controlled trial in Australia. JMIR Res Protoc 2017 Oct 24;6(10):e200 [FREE Full text] [doi: 10.2196/resprot.8189] [Medline: 29066427]
32. Burke-Garcia A, Mathew S. Leveraging social and digital media for participant recruitment: a review of methods from the Bayley Short Form Formative Study. J Clin Transl Sci 2017 Jun 29;1(3):205-207 [FREE Full text] [doi: 10.1017/cts.2017.9] [Medline: 29093968]
33. Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, et al. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. AIDS Behav 2012 Oct;16(7):1816-1820 [FREE Full text] [doi: 10.1007/s10461-012-0277-z] [Medline: 22851153]
34. Burton-Chase AM, Parker WM, Hennig K, Sisson F, Bruzzone LL. The use of social media to recruit participants with rare conditions: lynch syndrome as an example. JMIR Res Protoc 2017 Jan 23;6(1):e12 [FREE Full text] [doi: 10.2196/resprot.6066] [Medline: 28115298]
35. Byaruhanga J, Tzelepis F, Paul C, Wiggers J, Byrnes E, Lecathelinais C. Cost per participant recruited from rural and remote areas into a smoking cessation trial via online or traditional strategies: observational study. J Med Internet Res 2019 Nov 12;21(11):e14911 [FREE Full text] [doi: 10.2196/14911] [Medline: 31714253]
36. Cahill TJ, Wertz B, Zhong Q, Parlato A, Donegan J, Forman R, et al. The search for consumers of web-based raw DNA interpretation services: using social media to target hard-to-reach populations. J Med Internet Res 2019 Jul 30;21(7):e12980 [FREE Full text] [doi: 10.2196/12980] [Medline: 31364607]
37. Carter-Harris L. Facebook targeted advertisement for research recruitment: a primer for nurse researchers. Appl Nurs Res 2016 Nov;32:144-147. [doi: 10.1016/j.apnr.2016.07.006] [Medline: 27969019]
38. Cavallo D, Lim R, Ishler K, Pagano M, Perovsek R, Albert E, et al. Effectiveness of social media approaches to recruiting young adult cigarillo smokers: cross-sectional study. J Med Internet Res 2020 Jul 22;22(7):e12619 [FREE Full text] [doi: 10.2196/12619] [Medline: 32459629]
39. Ahmed KL, Simon AR, Dempsey JR, Samaco RC, Goin-Kochel RP. Evaluating two common strategies for research participant recruitment into autism studies: observational study. J Med Internet Res 2020 Sep 24;22(9):e16752 [FREE Full text] [doi: 10.2196/16752] [Medline: 32969826]
40. Choi SH, Milne DN, Glozier N, Peters D, Harvey SB, Calvo RA. Using different Facebook advertisements to recruit men for an online mental health study: engagement and selection bias. Internet Interv 2017 Jun;8(1):27-34 [FREE Full text] [doi: 10.1016/j.invent.2017.02.002] [Medline: 30135825]
41. Chu JL, Snider CE. Use of a social networking web site for recruiting Canadian youth for medical research. J Adolesc Health 2013 Jun;52(6):792-794. [doi: 10.1016/j.jadohealth.2012.12.002] [Medline: 23352727]
42. Close S, Smaldone A, Fennoy I, Reame N, Grey M. Using information technology and social networking for recruitment of research participants: experience from an exploratory study of pediatric Klinefelter syndrome. J Med Internet Res 2013 Mar 19;15(3):e48 [FREE Full text] [doi: 10.2196/jmir.2286] [Medline: 23512442]
43. Cowie JM, Gurney ME. The use of Facebook advertising to recruit healthy elderly people for a clinical trial: baseline metrics. JMIR Res Protoc 2018 Jan 24;7(1):e20 [FREE Full text] [doi: 10.2196/resprot.7918] [Medline: 29367186]
44. Derrick JL, Eliseo-Arras RK, Hanny C, Britton M, Haddad S. Comparison of internet and mailing methods to recruit couples into research on unaided smoking cessation. Addict Behav 2017 Dec;75:12-16 [FREE Full text] [doi: 10.1016/j.addbeh.2017.06.012] [Medline: 28662435]
45. Dunn PH, Woo BK. Facebook recruitment of Chinese-speaking participants for hypertension education. J Am Soc Hypertens 2018 Sep;12(9):690-692. [doi: 10.1016/j.jash.2018.06.017] [Medline: 30054194]
46. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012 Feb 01;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
47. Ford KL, Albritton T, Dunn TA, Crawford K, Neuwirth J, Bull S. Youth study recruitment using paid advertising on Instagram, Snapchat, and Facebook: cross-sectional survey study. JMIR Public Health Surveill 2019 Oct 09;5(4):e14080 [FREE Full text] [doi: 10.2196/14080] [Medline: 31599739]
48. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016 Aug 10;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]

49. Ahmed N, Jayasinghe Y, Wark JD, Fenner Y, Moore EE, Tabrizi SN, et al. Attitudes to Chlamydia screening elicited using the social networking site Facebook for subject recruitment. Sex Health 2013 Jul;10(3):224-228. [doi: 10.1071/SH12198] [Medline: 23702200]
50. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]
51. Gilligan C, Kypri K, Bourke J. Social networking versus facebook advertising to recruit survey respondents: a quasi-experimental study. JMIR Res Protoc 2014 Sep 17;3(3):e48 [FREE Full text] [doi: 10.2196/resprot.3317] [Medline: 25230740]
52. Gioia CJ, Sobell LC, Sobell MB, Agrawal S. Craigslist versus print newspaper advertising for recruiting research participants for alcohol studies: cost and participant characteristics. Addict Behav 2016 Mar;54:24-32 [FREE Full text] [doi: 10.1016/j.addbeh.2015.11.008] [Medline: 26675247]
53. Gorman JR, Roberts SC, Dominick SA, Malcarne VL, Dietz AC, Su HI. A diversified recruitment approach incorporating social media leads to research participation among young adult-aged female cancer survivors. J Adolesc Young Adult Oncol 2014 Jun 01;3(2):59-65 [FREE Full text] [doi: 10.1089/jayao.2013.0031] [Medline: 24940529]
54. Guthrie KA, Caan B, Diem S, Ensrud KE, Greaves SR, Larson JC, et al. Facebook advertising for recruitment of midlife women with bothersome vaginal symptoms: a pilot study. Clin Trials 2019 Oct;16(5):476-480 [FREE Full text] [doi: 10.1177/1740774519846862] [Medline: 31055949]
55. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young Australian women. Am J Epidemiol 2015 May 15;181(10):737-746. [doi: 10.1093/aje/kwv006] [Medline: 25883155]
56. Hsiao Y, Lee I, Lin T, Liao C, Lee C, Ku W, et al. Demographic differences in people living with HIV according to recruitment sources: comparison between health-care systems and social media networks. AIDS Care 2020 Jul 18;32(7):901-906. [doi: 10.1080/09540121.2019.1668524] [Medline: 31533457]
57. Iott BE, Veinot TC, Loveluck J, Kahle E, Golson L, Benton A. Comparative analysis of recruitment strategies in a study of men who have sex with men (MSM) in metropolitan Detroit. AIDS Behav 2018 Jul 22;22(7):2296-2311. [doi: 10.1007/s10461-018-2071-z] [Medline: 29470808]
58. Juraschek SP, Plante TB, Charleston J, Miller ER, Yeh H, Appel LJ, et al. Use of online recruitment strategies in a randomized trial of cancer survivors. Clin Trials 2018 Apr;15(2):130-138 [FREE Full text] [doi: 10.1177/1740774517745829] [Medline: 29361843]
59. Kapp JM, Peters C, Oliver DP. Research recruitment using Facebook advertising: big potential, big challenges. J Cancer Educ 2013 Mar;28(1):134-137. [doi: 10.1007/s13187-012-0443-z] [Medline: 23292877]
60. Akers L, Gordon JS. Using Facebook for large-scale online randomized clinical trial recruitment: effective advertising strategies. J Med Internet Res 2018 Nov 08;20(11):e290 [FREE Full text] [doi: 10.2196/jmir.9372] [Medline: 30409765]
61. Kelleher E, Moreno M, Wilt MP. Recruitment of participants and delivery of online mental health resources for depressed individuals using Tumblr: pilot randomized control trial. JMIR Res Protoc 2018 Apr 12;7(4):e95 [FREE Full text] [doi: 10.2196/resprot.9421] [Medline: 29650507]
62. Lohse B. Facebook is an effective strategy to recruit low-income women to online nutrition education. J Nutr Educ Behav 2013;45(1):69-76. [doi: 10.1016/j.jneb.2012.06.006] [Medline: 23305805]
63. McCarthy E, Mazza D. Cost and effectiveness of using Facebook advertising to recruit young women for research: PREFER (contraceptive preferences study) experience. J Med Internet Res 2019 Nov 29;21(11):e15869 [FREE Full text] [doi: 10.2196/15869] [Medline: 31782738]
64. Musiat P, Winsall M, Orlowski S, Antezana G, Schrader G, Battersby M, et al. Paid and unpaid online recruitment for health interventions in young adults. J Adolesc Health 2016 Dec;59(6):662-667. [doi: 10.1016/j.jadohealth.2016.07.020] [Medline: 27663927]
65. Nash EL, Gilroy D, Srikusalanukul W, Abhayaratna WP, Stanton T, Mitchell G, et al. Facebook advertising for participant recruitment into a blood pressure clinical trial. J Hypertens 2017 Dec;35(12):2527-2531. [doi: 10.1097/HJH.0000000000001477] [Medline: 28704263]
66. Paige SR, Krieger JL. Recruitment in online research for COPD: leveraging social media and research registries. ERJ Open Res 2019 Apr 08;5(2):00167-2018 [FREE Full text] [doi: 10.1183/23120541.00167-2018] [Medline: 30972351]
67. Peñaherrera CA, Palacios M, Duarte MC, Santibáñez R, Tamariz L, Palacio A. Use of media for recruiting clinical research volunteers in Ecuador. Medwave 2015 Dec 10;15(11):e6334. [doi: 10.5867/medwave.2015.11.6334] [Medline: 26731413]
68. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020 Oct 07;22(10):e23021 [FREE Full text] [doi: 10.2196/23021] [Medline: 33026360]
69. Ryan J, Lopian L, Le B, Edney S, Van Kessel G, Plotnikoff R, et al. It's not raining men: a mixed-methods study investigating methods of improving male recruitment to health behaviour research. BMC Public Health 2019 Jun 24;19(1):814 [FREE Full text] [doi: 10.1186/s12889-019-7087-4] [Medline: 31234825]

70. Alley S, Jennings C, Plotnikoff RC, Vandelanotte C. An evaluation of web- and print-based methods to attract people to a physical activity intervention. JMIR Res Protoc 2016 May 27;5(2):e94 [FREE Full text] [doi: 10.2196/resprot.4826] [Medline: 27235075]
71. Shaver LG, Khawer A, Yi Y, Aubrey-Bassler K, Etchegary H, Roebothan B, et al. Using Facebook advertising to recruit representative samples: feasibility assessment of a cross-sectional survey. J Med Internet Res 2019 Aug 19;21(8):e14021 [FREE Full text] [doi: 10.2196/14021] [Medline: 31429409]
72. Subasinghe AK, Nguyen M, Wark JD, Tabrizi SN, Garland SM. Targeted Facebook advertising is a novel and effective method of recruiting participants into a human papillomavirus vaccine effectiveness study. JMIR Res Protoc 2016 Jul 22;5(3):e154 [FREE Full text] [doi: 10.2196/resprot.5679] [Medline: 27450586]
73. Teo AR, Liebow SB, Chan B, Dobscha SK, Graham AL. Reaching those at risk for psychiatric disorders and suicidal ideation: Facebook advertisements to recruit military veterans. JMIR Ment Health 2018 Jul 05;5(3):e10078 [FREE Full text] [doi: 10.2196/10078] [Medline: 29980498]
74. Thornton LK, Harris K, Baker A, Johnson M, Kay-Lambkin FJ. Recruiting for addiction research via Facebook. Drug Alcohol Rev 2016 Jul;35(4):494-502. [doi: 10.1111/dar.12305] [Medline: 26179224]
75. Tsai W, Zavala D, Gomez S. Using the Facebook advertisement platform to Recruit Chinese, Korean, and Latinx Cancer survivors for psychosocial research: web-based survey study. J Med Internet Res 2019 Jan 10;21(1):e11571 [FREE Full text] [doi: 10.2196/11571] [Medline: 30632966]
76. Valdez RS, Guterbock TM, Thompson MJ, Reilly JD, Menefee HK, Bennici MS, et al. Beyond traditional advertisements: leveraging Facebook's social structures for research recruitment. J Med Internet Res 2014;16(10):e243 [FREE Full text] [doi: 10.2196/jmir.3786] [Medline: 25348050]
77. van Gelder MM, van de Belt TH, Engelen LJ, Hooijer R, Bredie SJ, Roeleveld N. Google AdWords and Facebook ads for recruitment of pregnant women into a prospective cohort study with long-term follow-up. Matern Child Health J 2019 Oct;23(10):1285-1291 [FREE Full text] [doi: 10.1007/s10995-019-02797-2] [Medline: 31222599]
78. Vial AC, Starks TJ, Parsons JT. Finding and recruiting the highest risk HIV-negative men who have sex with men. AIDS Educ Prev 2014 Feb;26(1):56-67 [FREE Full text] [doi: 10.1521/aeap.2014.26.1.56] [Medline: 24450278]
79. Wasilewski MB, Stinson JN, Webster F, Cameron JI. Using Twitter to recruit participants for health research: an example from a caregiving study. Health Informatics J 2019 Dec 30;25(4):1485-1497 [FREE Full text] [doi: 10.1177/1460458218775158] [Medline: 29843545]
80. Arcia A. Facebook advertisements for inexpensive participant recruitment among women in early pregnancy. Health Educ Behav 2014 Jun;41(3):237-241 [FREE Full text] [doi: 10.1177/1090198113504414] [Medline: 24082026]
81. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant recruitment and retention in remote eHealth intervention trials: methods and lessons learned from a large randomized controlled trial of two web-based smoking interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
82. Wilkerson JM, Patankar P, Rawat SM, Simon Rosser BR, Shukla KM, Rhoton J, et al. Recruitment strategies of Indian men who have sex with men in the state of Maharashtra into an online survey. Int J Sex Health 2016;28(3):221-227 [FREE Full text] [doi: 10.1080/19317611.2016.1193079] [Medline: 27668029]
83. Wilkerson JM, Shenk JE, Grey JA, Simon Rosser BR, Noor SW. Recruitment strategies of methamphetamine-using men who have sex with men into an online survey. J Subst Use 2015;20(1):33-37 [FREE Full text] [doi: 10.3109/14659891.2013.868936] [Medline: 25642143]
84. Barney A, Rodriguez F, Schwarz EB, Reed R, Tancredi D, Brindis CD, et al. Adapting to changes in teen pregnancy prevention research: social media as an expedited recruitment strategy. J Adolesc Health 2021 Aug;69(2):349-353. [doi: 10.1016/j.jadohealth.2020.12.140] [Medline: 33632643]
85. Moseson H, Wollum A, Seymour JW, Zuniga C, Thompson T, Gerdts C. Comparison of Facebook, Google Ads, and Reddit for the recruitment of people who considered but did not obtain abortion care in the United States: cross-sectional survey. JMIR Form Res 2021 Feb 24;5(2):e22854 [FREE Full text] [doi: 10.2196/22854] [Medline: 33625368]
86. Turner-McGrievy GM, Wilson MJ, Bailey S, Bernhart JA, Wilcox S, Frongillo EA, et al. Effective recruitment strategies for African-American men and women: the Nutritious Eating with Soul study. Health Educ Res 2021 Apr 12;36(2):206-211 [FREE Full text] [doi: 10.1093/her/cyab003] [Medline: 33582788]
87. Harfield S, Elliott S, Ramsey L, Housen T, Ward J. Using social networking sites to recruit participants: methods of an online survey of sexual health, knowledge and behaviour of young South Australians. Aust N Z J Public Health 2021 Aug 07;45(4):348-354. [doi: 10.1111/1753-6405.13117] [Medline: 34097339]
88. Lindsay AC, Valdez MJ, Pineda JA, Muñoz MA. Lessons learned from recruiting Latino fathers in child health research. Health Promot Pract 2021 Jul 19;22(4):462-468. [doi: 10.1177/1524839920963704] [Medline: 33073618]
89. Spahrkäs SS, Looijmans A, Sanderman R, Hagedoorn M. Recruiting participants for an international mHealth study via Facebook Ads: experiences from the Untire App RCT. Internet Interv 2021 Mar;23:100362 [FREE Full text] [doi: 10.1016/j.invent.2021.100362] [Medline: 33489782]
90. Grov C, Ventuneac A, Rendina HJ, Jimenez RH, Parsons JT. Recruiting men who have sex with men on Craigslist.org for face-to-face assessments: implications for research. AIDS Behav 2013 Feb 17;17(2):773-778 [FREE Full text] [doi: 10.1007/s10461-012-0345-4] [Medline: 23073646]

91. Bennetts SK, Hokke S, Crawford S, Hackworth NJ, Leach LS, Nguyen C, et al. Using paid and free Facebook methods to recruit Australian parents to an online survey: an evaluation. J Med Internet Res 2019 Mar 06;21(3):e11206 [FREE Full text] [doi: 10.2196/11206] [Medline: 30839282]
92. Blumenberg C, Menezes AM, Gonçalves H, Assunção MC, Wehrmeister FC, Barros AJ. How different online recruitment methods impact on recruitment rates for the web-based coortesnaweb project: a randomised trial. BMC Med Res Methodol 2019 Jun 19;19(1):127 [FREE Full text] [doi: 10.1186/s12874-019-0767-z] [Medline: 31217008]
93. Akard TF, Wray S, Gilmer MJ. Facebook advertisements recruit parents of children with cancer for an online survey of web-based research preferences. Cancer Nurs 2015;38(2):155-161 [FREE Full text] [doi: 10.1097/NCC.0000000000000146] [Medline: 24945264]
94. Herbell K, Zauszniewski JA. Facebook or Twitter?: effective recruitment strategies for family caregivers. Appl Nurs Res 2018 Jun;41:1-4. [doi: 10.1016/j.apnr.2018.02.004] [Medline: 29853206]
95. Jones R, Lacroix LJ, Porcher E. Facebook advertising to recruit young, urban women into an HIV prevention clinical trial. AIDS Behav 2017 Nov;21(11):3141-3153 [FREE Full text] [doi: 10.1007/s10461-017-1797-3] [Medline: 28528463]
96. Stuart WP, Moore B. Evidence-based Facebook recruitment of study participants. Comput Inform Nurs 2021 Mar 29;39(7):355-361. [doi: 10.1097/CIN.0000000000000709] [Medline: 34224415]
97. Cho E, Gilmer MJ, Friedman DL, Hendricks-Ferguson VL, Hinds PS, Akard TF. Facebook recruitment for children with advanced cancer and their parents: lessons from a web-based pediatric palliative intervention study. Prog Palliat Care 2021 Mar 29;29(5):264-271 [FREE Full text] [doi: 10.1080/09699260.2021.1898077] [Medline: 34737490]
98. Avery-Desmarais SL, McCurry MK, Sethares KA, Batchelder A, Stover C. Internet recruitment of a diverse population of lesbian, gay, and bisexual nurses in a study of substance use and minority stress. J Transcult Nurs 2022 Jan 27;33(1):118-125. [doi: 10.1177/10436596211042071] [Medline: 34448433]
99. Bethel C, Rainbow JG, Dudding KM. Recruiting nurses via social media for survey studies. Nurs Res 2021;70(3):231-235. [doi: 10.1097/NNR.0000000000000482] [Medline: 33060416]
100. Leighton K, Kardong-Edgren S, Schneidereith T, Foisy-Doll C. Using social media and snowball sampling as an alternative recruitment strategy for research. Clin Simulation Nursing 2021 Jun;55:37-42. [doi: 10.1016/j.ecns.2021.03.006]
101. Smith J, Brown K, Hutchings M, Merwin E. Strategies for recruiting and retaining nurses for rural hospital closure research. Online J Rural Nurs Heal Care 2021 Dec;21(2):14-43. [doi: 10.14574/ojrnhc.v21i2.682]
102. Jones R, Lacroix LJ, Nolte K. "Is your man stepping out?" an online pilot study to evaluate acceptability of a guide-enhanced HIV prevention soap opera video series and feasibility of recruitment by Facebook advertising. J Assoc Nurses AIDS Care 2015;26(4):368-386 [FREE Full text] [doi: 10.1016/j.jana.2015.01.004] [Medline: 26066692]
103. Amerson R. Facebook: a tool for nursing education research. J Nurs Educ 2011 Jul;50(7):414-416. [doi: 10.3928/01484834-20110331-01] [Medline: 21449530]
104. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. J Adv Nurs 2014 Mar;70(3):599-609. [doi: 10.1111/jan.12222] [Medline: 23909740]
105. Stokes Y, Vandyk A, Squires J, Jacob J, Gifford W. Using Facebook and Linkedin to recruit nurses for an online survey. West J Nurs Res 2019 Jan 07;41(1):96-110. [doi: 10.1177/0193945917740706] [Medline: 29113542]
106. Waltman NL, Smith KM, Kupzyk KA, Lappe JM, Mack LR, Bilek LD. Approaches to recruitment of postmenopausal women for a community-based study. Nurs Res 2019;68(4):307-316 [FREE Full text] [doi: 10.1097/NNR.0000000000000356] [Medline: 30829836]
107. Williamson GR, O'Connor A, Chamberlain C, Halpin D. mHealth resources for asthma and pregnancy care: methodological issues and social media recruitment. A discussion paper. J Adv Nurs 2018 Oct 17;74(10):2442-2449. [doi: 10.1111/jan.13773] [Medline: 29943472]
108. Wilson RL, Usher K. Social media as a recruitment strategy: using Twitter to explore young people's mental health. Nurse Res 2017 Dec 18;25(3):36-41. [doi: 10.7748/nr.2017.e1478] [Medline: 29251447]
109. Chambers M, Bliss K, Rambur B. Recruiting research participants via traditional snowball vs Facebook advertisements and a website. West J Nurs Res 2020 Oct 02;42(10):846-851. [doi: 10.1177/0193945920904445] [Medline: 32009549]
110. Child RJ, Mentes JC, Pavlish C, Phillips LR. Using Facebook and participant information clips to recruit emergency nurses for research. Nurse Res 2014 Jul;21(6):16-21. [doi: 10.7748/nr.21.6.16.e1246] [Medline: 25059083]
111. Chung S, Hacker ED, Rawl S, Ellis R, Bakas T, Jones J, et al. Using Facebook in recruiting kidney transplant recipients for a REDCap study. West J Nurs Res 2019 Dec;41(12):1790-1812. [doi: 10.1177/0193945919832600] [Medline: 30836840]
112. Corey KL, McCurry MK, Sethares KA, Bourbonniere M, Hirschman KB, Meghani SH. Utilizing internet-based recruitment and data collection to access different age groups of former family caregivers. Appl Nurs Res 2018 Dec;44:82-87 [FREE Full text] [doi: 10.1016/j.apnr.2018.10.005] [Medline: 30389065]
113. Desroches ML. Facebook recruitment of nurses as research participants: methodological considerations. Appl Nurs Res 2020 Aug;54:151282. [doi: 10.1016/j.apnr.2020.151282] [Medline: 32650889]
114. Hansen D, Sheehan DK, Stephenson P. Family caregiver recruitment via social media: challenges, opportunities and lessons. Int J Palliat Nurs 2017 Feb 02;23(2):84-87. [doi: 10.12968/ijpn.2017.23.2.84] [Medline: 28245165]
115. Herbell K. Using Facebook to recruit pregnant women for research. Nurs Res 2019;68(3):242-245. [doi: 10.1097/NNR.0000000000000353] [Medline: 31034458]

116. CASP - Critical Appraisal Skills Programme. Casp Checklists. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2022-05-04]
117. Carter-Harris L, Bartlett Ellis R, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging Facebook-targeted advertisement to recruit long-term smokers for research. J Med Internet Res 2016 Jun 15;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]
118. Curtis BL. Social networking and online recruiting for HIV research: ethical challenges. J Empir Res Hum Res Ethics 2014 Feb;9(1):58-70 [FREE Full text] [doi: 10.1525/jer.2014.9.1.58] [Medline: 24572084]
119. Leach LS, Bennetts SK, Giallo R, Cooklin AR. Recruiting fathers for parenting research using online advertising campaigns: evidence from an Australian study. Child Care Health Dev 2019 Nov 22;45(6):871-876. [doi: 10.1111/cch.12698] [Medline: 31216596]
120. Schwinn T, Hopkins J, Schinke SP, Liu X. Using Facebook ads with traditional paper mailings to recruit adolescent girls for a clinical trial. Addict Behav 2017 Feb;65:207-213 [FREE Full text] [doi: 10.1016/j.addbeh.2016.10.011] [Medline: 27835860]
121. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
122. Kamp K, Herbell K, Magginis WH, Berry D, Given B. Facebook recruitment and the protection of human subjects. West J Nurs Res 2019 Sep;41(9):1270-1281 [FREE Full text] [doi: 10.1177/0193945919828108] [Medline: 30729866]
123. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14(1):42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
124. Sydor A. Conducting research into hidden or hard-to-reach populations. Nurse Res 2013 Jan;20(3):33-37. [doi: 10.7748/nr2013.01.20.3.33.c9495] [Medline: 23346777]
125. How much does social media advertising cost? Strike Social. URL: https://strikesocial.com/blog/social-media-advertising-costs/ [accessed 2020-09-18]
126. Stern MJ, Fordyce E, Hansen C, Heim Viox M, Michaels S, Schlissel A, et al. Social media recruitment for a web survey of sexual and gender minority youth: an evaluation of methods used and resulting sample diversity. LGBT Health 2020;7(8):448-456. [doi: 10.1089/lgbt.2019.0311] [Medline: 33147121]
127. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
128. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
129. Im E, Chee W. An online forum as a qualitative research method: practical issues. Nurs Res 2006;55(4):267-273 [FREE Full text] [doi: 10.1097/00006199-200607000-00007] [Medline: 16849979]
130. Amon KL, Campbell AJ, Hawke C, Steinbeck K. Facebook as a recruitment tool for adolescent health research: a systematic review. Acad Pediatr 2014;14(5):439-47.e4. [doi: 10.1016/j.acap.2014.05.049] [Medline: 25169155]
131. Frampton GK, Shepherd J, Pickett K, Griffiths G, Wyatt JC. Digital tools for the recruitment and retention of participants in randomised controlled trials: a systematic map. Trials 2020 Jun 05;21(1):478 [FREE Full text] [doi: 10.1186/s13063-020-04358-3] [Medline: 32498690]
132. Park BK, Calamaro C. A systematic review of social networking sites: innovative platforms for health research targeting adolescents and young adults. J Nurs Scholarsh 2013 Sep;45(3):256-264. [doi: 10.1111/jnu.12032] [Medline: 23676115]
133. Reaves AC, Bianchi DW. The role of social networking sites in medical genetics research. Am J Med Genet A 2013 May;161A(5):951-957. [doi: 10.1002/ajmg.a.35903] [Medline: 23554131]
134. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
135. Jones J, Salazar LF. A review of HIV prevention studies that use social networking sites: implications for recruitment, health promotion campaigns, and efficacy trials. AIDS Behav 2016 Nov 22;20(11):2772-2781. [doi: 10.1007/s10461-016-1342-9] [Medline: 26902295]
136. Kane L. Usability for seniors: challenges and changes. Nielsen Norman Group. URL: https://www.nngroup.com/articles/usability-for-senior-citizens/ [accessed 2022-05-15]
137. Hutto C, Bell C, Farmer S, Fausset C, Harley L, Nguyen J, et al. Social media gerontology: understanding social media usage among older adults. Web Intell Agent Syst 2015 Apr 29;13(1):69-87. [doi: 10.3233/web-150310]
138. Lee B, Chen Y, Hewitt L. Age differences in constraints encountered by seniors in their use of computers and the internet. Comput Human Behav 2011 May;27(3):1231-1237. [doi: 10.1016/j.chb.2011.01.003]
139. Surdam J, Daly B, Fulton S, Margevicius S, Schluchter M, Flocke S, et al. Recruitment strategies for nurse enrollment in an online study. Nurs Res 2020;69(1):69-73 [FREE Full text] [doi: 10.1097/NNR.0000000000000393] [Medline: 31804433]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
REB: research ethics board

Edited by R Kukafka; submitted 15.03.22; peer-reviewed by N Frisch, T Risling, L Southwick, A McLean; comments to author 06.05.22; accepted 15.06.22; published 04.08.22.

Please cite as:
Darko EM, Kleib M, Olson J
Social Media Use for Research Participant Recruitment: Integrative Literature Review
J Med Internet Res 2022;24(8):e38015
URL: https://www.jmir.org/2022/8/e38015
doi:10.2196/38015
PMID:35925655

Review

Digital Patient Experience: Umbrella Systematic Review

Tingting Wang[1], MA; Guido Giunti[1,2], PhD; Marijke Melles[1], PhD; Richard Goossens[1], PhD

[1]Industrial Design Engineering, Delft University of Technology, Delft, Netherlands
[2]Digital Health Design and Development, University of Oulu, Oulu, Finland

Corresponding Author:
Tingting Wang, MA
Industrial Design Engineering
Delft University of Technology
Gebouw 32
Landbergstraat 15
Delft, 2628 CE
Netherlands
Phone: 31 623018218
Email: t.wang-8@tudelft.nl

Abstract

Background: The adoption and use of technology have significantly changed health care delivery. Patient experience has become a significant factor in the entire spectrum of patient-centered health care delivery. Digital health facilitates further improvement and empowerment of patient experiences. Therefore, the design of digital health is served by insights into the barriers to and facilitators of digital patient experience (PEx).

Objective: This study aimed to systematically review the influencing factors and design considerations of PEx in digital health from the literature and generate design guidelines for further improvement of PEx in digital health.

Methods: We performed an umbrella systematic review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology. We searched Scopus, PubMed, and Web of Science databases. Two rounds of small random sampling (20%) were independently reviewed by 2 reviewers who evaluated the eligibility of the articles against the selection criteria. Two-round interrater reliability was assessed using the Fleiss-Cohen coefficient ($k1=0.88$ and $k2=0.80$). Thematic analysis was applied to analyze the extracted data based on a small set of a priori categories.

Results: The search yielded 173 records, of which 45 (26%) were selected for data analysis. Findings and conclusions showed a great diversity; most studies presented a set of themes (19/45, 42%) or descriptive information only (16/45, 36%). The digital PEx–related influencing factors were classified into 9 categories: patient capability, patient opportunity, patient motivation, intervention technology, intervention functionality, intervention interaction design, organizational environment, physical environment, and social environment. These can have three types of impacts: positive, negative, or double edged. We captured 4 design constructs (personalization, information, navigation, and visualization) and 3 design methods (human-centered or user-centered design, co-design or participatory design, and inclusive design) as design considerations.

Conclusions: We propose the following definition for digital PEx: “Digital patient experience is the sum of all interactions affected by a patient’s behavioral determinants, framed by digital technologies, and shaped by organizational culture, that influence patient perceptions across the continuum of care channeling digital health.” In this study, we constructed a design and evaluation framework that contains 4 phases—define design, define evaluation, design ideation, and design evaluation—and 9 design guidelines to help digital health designers and developers address digital PEx throughout the entire design process. Finally, our review suggests 6 directions for future digital PEx–related research.

(J Med Internet Res 2022;24(8):e37952) doi:10.2196/37952

KEYWORDS

digital health; eHealth; telemedicine; telehealth; mobile health; mHealth; patient experience; user experience; influencing factors; user-centered design; human-computer interaction

Introduction

Background

Recently, there has been a significant increase in the use of digital health technologies. In addition, many countries currently use digital health technologies to support health care service delivery to overcome the disruptions caused by the COVID-19 pandemic. These include web-based patient consultations and requesting pharmacy and medication refills [1]. Digital health offers care without the risk of exposure to the virus, especially for vulnerable patients such as older adults and patients with chronic diseases [2]. Before the COVID-19 pandemic, there was increasing recognition of the potential of digital health to improve the accessibility of health care in different clinical settings (eg, ambulatory care, acute care, and inpatient care) [3]. Digital health provides an opportunity to both reduce the costs of care and improve patient affordability [4,5], and previous research suggests that digital health has the potential to provide health prevention, consultation, treatment, and management [5-10]. With digital health solutions continuing to grow in both number and functionality, patient interest in digital health has rapidly increased, leading to an expanding reliance on digital health technologies [11].

As digital health has become a more familiar term, it has generated many definitions, and the concept has been expanded to encompass a much broader set of scientific concepts and technologies [12]. These include digital health applications, ecosystems and platforms [13], patient portals [14], mobile health apps [15], eHealth records, and appointment scheduling applications [16]. For the purposes of this study, we will use eHealth, mobile health, telemedicine, telehealth, virtual health, remote health, electronic consultations, and health information systems (HISs) as interchangeable terms for digital health.

Patient Experience in Digital Health

Digital health has the potential to improve patients' overall health care experience [17-19]. However, there is currently no common concept for describing patient experience (PEx) in digital health. Neither the general PEx nor user experience (UX) adequately reflects the experience of a patient using a digital service. For example, in a hospital setting, the environment's cleanliness, background noise, and even food provision could affect PEx [20]; however, these factors would not be expected to influence the experience of a patient using a digital service. Similarly, the fact that the system passes usability heuristics does not necessarily mean that the overall experience of a patient using digital health services is positive [21]. Therefore, it is vital to understand the experiences of individuals using digital health and how the design of new technologies can affect them [17,22,23].

The concept of (nondigital) PEx has many definitions in general health care practice and research. The Beryl Institute defines PEx as "the sum of all interactions, shaped by an organization's culture, that influence patient perceptions, across the continuum of care" [24]. Other definitions and studies note that the core elements of optimized PEx include access to appropriate care, patients' active participation in care, a good patient-physician relationship, reliable evidence-based care, comprehensible information, physical comfort, emotional support, involvement of family and friends, individualized approaches, responsiveness of services, and continuity of care [19,25-27]. These core elements of PEx help to recognize patients' priorities when receiving care and in providing patient-centered care. However, patients' priorities may differ for digital health, in which traditional face-to-face interaction is replaced by human to digital interface interaction. Therefore, to address patient priorities in digital health, it is essential to consider UX in the design of digital health [28]. In this study, we define UX as a person's perceptions and responses that result from the use or anticipated use of a product, system, or service [18,29]. Usable, useful, findable, accessible, credible, valuable, and desirable products are more likely to succeed in delivering a positive UX [30]. However, the full impact of digital health technologies on PEx or UX still remains unclear [31]; some products even result in negative effects such as increased patient anxiety [32]. Therefore, more insights into the barriers to and facilitators of individuals' experiences with digital health are required [33].

Objectives

The objectives of this paper were to systematically review (1) the factors that influence PEx in digital health and (2) the design considerations of PEx that are in digital health. The overall aim was to generate a design framework and guidelines for further improving PEx in digital health.

Methods

We performed an umbrella systematic review compiling evidence from multiple systematic reviews [34] on PEx and UX in digital health. This review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology, which is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses [35].

Digital PEx Working Definition

Throughout this study, we use the term digital PEx as a working definition to describe people's experiences in various digital health contexts. As the study progressed, the definition underwent several revisions, which resulted in a more inclusive final definition.

Search Strategy

We searched Scopus, PubMed, and Web of Science for studies published between January 1, 2000, and December 16, 2020. The search time window was limited to 2000 as the term digital health was first introduced by Frank [36] in 2000. To be inclusive, we used broad interchangeable search terms with varying combinations of digital health, PEx, and UX:

1. Category 1: "patient experience" OR "health experience" OR "user experience" OR "customer experience" OR "client experience"
2. Category 2: "ehealth" OR "e-health" OR "mhealth" OR "m-health" OR "telehealth" OR "tele-health" OR "digital health" OR "virtual health" OR "remote health" OR "telemedicine" OR "telemonitoring" OR "teleconsultation"

3. Category 3: “patient digital experience” OR “patient experience in digital health” OR “e-patient experience” OR “epatient experience” OR “online patient experience”

After combining categories 1, 2, and 3, limits were set to restrict studies to English-language literature reviews published in journals after 2000. The final search strategy was ([category 1 AND category 2] OR category 3) AND (DOCTYPE [review]) AND (PUBYEAR>2000) AND (LIMIT-TO [SRCTYPE, “journal”]) AND (LIMIT-TO [LANGUAGE, “English”]). Google Scholar was used as an additional database to manually search for additional related references based on the snowballing method during the review process.

Selection Criteria

Eligibility criteria were developed for title and abstract screening and refined for full-text screening. The following inclusion criteria were proposed by TW and GG and adjusted by MM and RG:

1. No duplicated articles
2. Full text available
3. English language
4. Only completed peer-reviewed journal articles
5. Only review articles
6. Related to digital health (ie, use of information and communication technology in health) and PEx, UX, or health care experience

Screening Process

The collected articles were included in the final analysis if they met all the inclusion criteria after a 2-stage screening process: first, a title and abstract review, followed by a full-text review. In the screening process, 2-round, small random samples (20%) were independently reviewed by 2 reviewers (TW and GG) who evaluated the eligibility of the articles against the selection criteria. The interrater reliability and clarity of the selection criteria were assessed using the Fleiss-Cohen coefficient until it reached the required strength (≥0.60). Uncertainties around paper inclusion and exclusion were resolved by discussions with the research team (TW, GG, MM, and RG) when necessary.

Data Extraction and Thematic Analysis

Articles meeting the eligibility criteria were imported into ATLAS.ti (Scientific Software Development GmbH; version 9.0.7; 1857) for data extraction. Data were extracted for the following aspects: (1) study characteristics, including authors, year of publication, research aims, review methods, target users, and digital health intervention (DHI) characteristics; (2) the overall impression of digital PEx (eg, the foci or types of findings regarding digital PEx); (3) influencing factors of digital PEx; and (4) design considerations for improving digital PEx.

We used the Braun and Clarke 6-phase thematic analysis method [37] to analyze the extracted data; these include (1) familiarization with the data, (2) generation of initial codes, (3) searching for themes among codes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the final report (analytical themes). A total of 4 researchers participated in the review process. After data familiarization, a set of a priori categories was defined by TW and refined by all the coauthors (Table 1). The coding was based on the Performance of Routine Information System Management (PRISM) framework [38], which states that routine HIS performance is affected by the system’s inputs (ie, technical, behavioral [39], and organizational determinants) and progress. Please note that other elements of the framework (outputs, outcomes, and impact) are discussed in another study addressing the evaluation of digital PEx (work in progress).

Group discussions among the authors were used to reach an agreement on the produced a priori categories. TW quoted the relevant data across the included reviews, generated initial codes based on a priori categories, and then searched for themes among codes. Frequently used terms in the included reviews were used as inspiration to generate subsequent codes and themes. The latter process was independently and randomly validated by GG, MM, and RG.

Table 1. A priori categories of influencing factors of digital patient experience based on the Performance of Routine Information System Management framework [38].

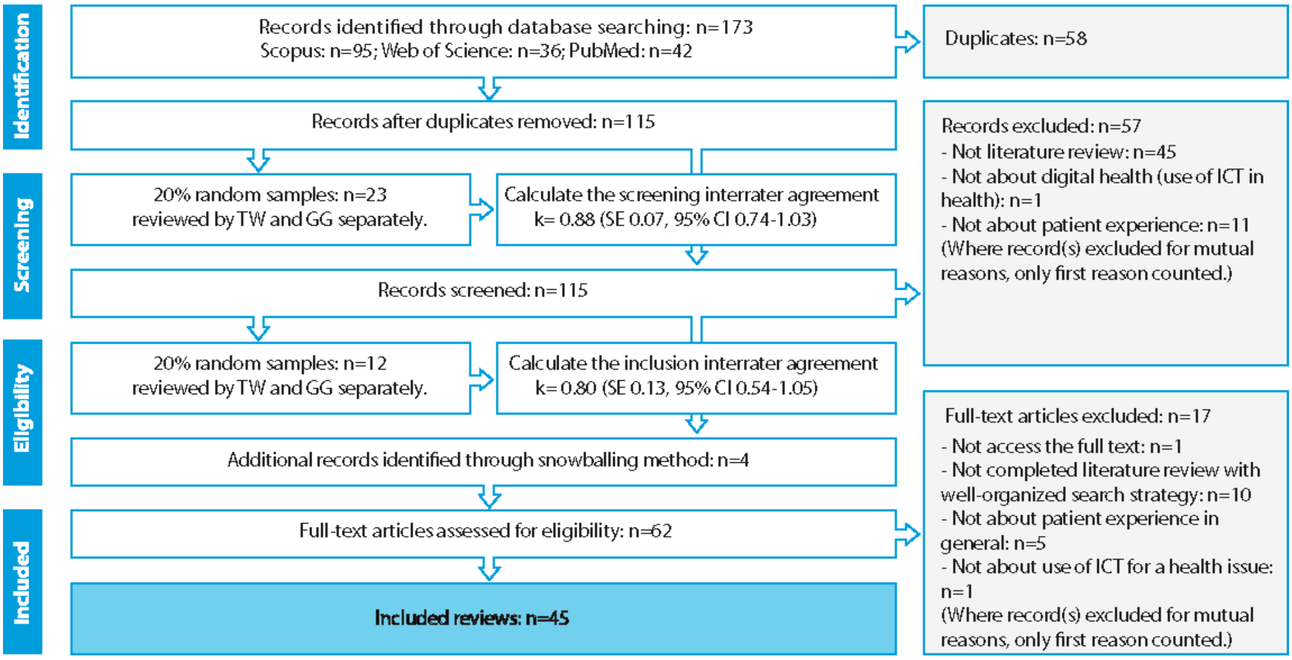
Determinants and a priori categories	Description
Behavioral determinants	
Patient capability	The individual’s psychological and physical capacity to engage in the concerned digital health activity
Patient opportunity	The individual’s internal conditions that enable or disrupt patients to engage in digital health
Patient motivation	The reflective and automatic brain processes that energize and direct patients’ goal setting and decision-making and their behaviors regarding using digital health
Technical determinants	
Intervention technology	The integration of telecommunications and computers, as well as necessary enterprise software, middleware, and storage and audiovisual software, which enables users to access, store, transmit, understand, and manipulate health information
Intervention functionality	The ability of digital health to work as expected to help users meet their health goals and needs
Intervention interaction design	The process of moving digital health from its existing state to a preferred state to optimize interactions between patients and digital health interventions
Organizational determinants	
Organizational environment	The management of the health service system, as affected by the rules, values, and practices of the involved people or community
Physical environment	The tangible surroundings (such as space, light, or sound) around patients, which affects their interactions with digital health
Social environment	The cultural environment (such as policy, business, or customs) that affect patients’ interactions with digital health

Results

Overview

Figure 1 shows the flow diagram of the systematic search. A total of 173 records were generated after the computer search; 58 (33.5%) duplicates were removed, and the titles and abstracts of 115 (66.5%) articles were reviewed. Subsequently, 53.9% (62/115) of full-text articles (including 4 additional records collected through snowballing) were reviewed for inclusion. Ultimately, 45 studies were included in the review for data extraction.

Figure 1. Study flow diagram. ICT: information and communications technology.



Study Characteristics

Embase, MEDLINE, PubMed, PsycINFO, CINAHL, and the Cochrane Library were the most common databases for the included reviews. Of these, 62% (28/45) were systematic review articles. The remainder included scoping reviews (6/45, 13%), literature reviews (3/45, 7%), integrative reviews (3/45, 7%), narrative reviews (2/45, 4%), comprehensive overviews (1/45, 2%), review of systematic reviews (1/45, 2%), and umbrella reviews (1/45, 2%). More than half of the included reviews (24/45, 53%) conducted quality assessments. The reviews included >1400 studies, which mainly or partially reported qualitative and quantitative analyses of PEx in digital health. The data analysis methods varied and included thematic analysis (8/45, 18%), meta-synthesis (5/45, 11%), meta-ethnography synthesis (2/45, 4%), taxonomy (1/45, 2%), hermeneutic synthesis (1/45, 2%), qualitative evidence synthesis (1/45, 2%), and state-of-the-art survey analysis (1/45, 2%).

Among the included reviews, some focused on specific populations, such as children (3/45, 7%), college students (1/45, 2%), younger people (1/45, 2%), adults (7/45, 16%), or older adults (4/45, 9%). Others either focused on the general population or did not mention the target population. The most common health issues across the included articles were chronic diseases (17/45, 38%), including chronic obstructive pulmonary disease, heart failure, cardiovascular disease, cancer, diabetes, and hypertension. Mental health problems (7/45, 16%), including depression, anxiety, psychological well-being, psychotic disorders, and schizophrenia, were the second most common health issues. The remainder either focused on other issues (8/45, 18%), including audiology, asthma, reproductive health, maternal health, newborn health, child health, adolescent health, surgery, postpartum, somatic diseases, or palliative care, or did not mention any specific health issues (14/45, 31%). Some papers (8/45, 18%) also provided multistakeholder perspectives, including health care professionals, providers, surgeons, clinicians, staff and organizations, implementers (such as health policy makers, clinicians, and researchers), and the participation of information technology.

The degree of detail provided about the interventions varied greatly across the studies. Phone-based apps, websites, handheld sensing devices, and ambient assisted living health care systems were common digital health deliveries. Interaction techniques included synchronous, asynchronous, and hybrid models. Diverse intervention platforms, systems, or functions were used to deliver various health care services, including supporting disease management (14/45, 31%); patient-to-physician communication or consultation (9/45, 20%); symptom monitoring (9/45, 20%); information transmission (4/45, 9%); health promotion activities (3/45, 7%); screening, diagnosis, or self-assessment (2/45, 4%); behavior changes (2/45, 4%); self-education (1/45, 2%); and decision-making (1/45, 2%). Multimedia Appendix 1 [28,40-83] provides detailed information regarding the characteristics of the included studies.

Overall Impression of Digital PEx

Our study revealed great diversity in the perspectives and definitions describing patients' experiences and characteristics when using digital health, presenting a variety of influencing factors and design considerations for digital PEx. The included studies showed different foci regarding digital PEx, including influencing factors (21/45, 47%) [28,40-59], digital health performance (19/45, 42%) [40-43,46,48,49,56,57,59-68], patient perceptions (9/45, 20%) [28,45,47,49,69-73], evaluation methods of digital health or digital PEx (8/45, 18%) [43,64,74-79], and design considerations (9/45, 20%) [48-50,53,54,59,80-82]. The findings and conclusions of the 45 reviews showed a great diversity. Most studies presented a set of themes (19/45, 42%) [28,44,45,48,49,51,54-57,59,62,69-73,76,79] or descriptions only (16/45, 36%) [40-42,46,47,58,60,61,63-67,74,75,83]. Other studies concluded with a theory-based description (5/45, 11%) [52,68,77,78,80], framework (4/45, 9%) [28,49,50,82], model (2/45, 4%) [53,69], method (2/45, 4%) [43,81], or checklist (1/45, 2%) [59]. Only a few studies transformed findings into design considerations (9/45, 20%) or visualized or structured their results into frameworks, models, checklists, or methods (9/45, 20%). Limited information was found on participant dropout reasons during the interventions [28,41,43,51,53,63,69,71]. The overall impression of the researchers on the DHIs was positive. In 51% (23/45) of reviews [41-44,48,49,52-54,57,59,61-63,65-69,72,73,77,80], the DHIs either showed promising results or at least results comparable with face-to-face health care services. Only 4% (2/45) of reviews [47,60] reported concrete evidence of the negative impact of current DHIs on digital PEx. In general, digital PEx was addressed because of the interactions between the DHIs and the patients involved and how the service was organized and carried out.

Influencing Factors of Digital PEx

An influencing factor is an aspect of the existing situation that influences other aspects of the situation, and it is formulated as an attribute of an element that is considered relevant and can be observed, measured, or assessed [84]. In this study, influencing factors refer to specific factors that lead to a positive or negative experience (digital PEx). Some factors have either positive or negative consistent and concrete impacts, whereas others have double-edged impacts; that is, impacts that are different per individual or change over time. Among the included papers, a common understanding of the potential influencing factors was captured from 3 aspects—behavioral, technical, and organizational determinants—following the categorization of the PRISM framework. These determinants were each classified into 3 categories, resulting in nine categories: patient capability, patient opportunity, patient motivation, intervention technology, intervention functionality, intervention interaction design, organizational environment, physical environment, and social environment. Multimedia Appendix 2 [28,40-83] presents an overview of the themes identified for each category, the influencing factors per theme (positive, negative, and double-edged), and references. Most factors appear to be related to technical determinants, followed by behavioral and organizational determinants. For *technical determinants*, we summarized 3 categories with 13 themes, containing 58 positive, 35 negative, and 13 double-edged factors. For example, DHIs with multiple behavioral change techniques appeared to be more effective [42,56,80] and reported higher patient satisfaction [54,57]. *Behavioral determinants* included 3 categories with 9 themes containing 11 positive, 21 negative, and 5 double-edged factors. For instance, some studies mentioned a lack of confidence in patients' own ability to use the technology [43,45,47,48,71,81], leading to a negative digital PEx. *Organizational determinants* were classified into 3 categories with 5 themes, including 13 positive and 23 negative factors. For example, unrealistic financial reimbursement and higher costs related to the internet or equipment were practical challenges of using digital health [47,48,51,55,56]. For the behavioral and organizational determinants, we collected more negative factors than positive factors. This is in contrast to the technical determinants, in which more positive factors were identified. Double-edged factors were less than both positive and negative factors for all the 3 determinants. Multimedia Appendix 3 [28,40-61,63-67,69-73,75-78,80-83] provides detailed information and examples.

Design Considerations of Digital PEx

Table 2 provides an overview of the identified themes for each design construct or method, related considerations, and references. To address the abovementioned influencing factors, several the included articles referred to *design constructs* (personalization, information, navigation, and visualization) [48,49,53,54,59,80] and *design methods* (ie, human-centered design [HCD] or user-centered design [UCD], co-design or participatory design, and inclusive design) [48-50,54,80-82], either as recommendations or implications for improving digital PEx from a design perspective. Notably, there was an overlap between design considerations and influencing factors. The former focuses on concluding possible design suggestions, recommendations, and implications proposed by the reviewed articles. The latter involves mapping the impacts of interaction design on digital PEx in different contexts; therefore, they refer to different themes and references. Generally, the *personalization construct* identifies patient profiles and tailors digital health according to patients' needs and preferences. The *information construct* addresses the source, language, presentation, content, and architecture of delivered health information. The *navigation construct* considers the interactive, delivered, and instructional elements of digital health to guide users to different areas of content within digital health. The *visualization construct* focuses on the aesthetics, attractiveness, visibility, and consistency of digital health appearance and interface. Furthermore, co-design and UCD or HCD were recommended as the most common methods for designing digital health, which involve multi-stakeholders and multi-disciplinaries in the design process to facilitate the designers' work, as designers need to understand end user needs and be aware of potential barriers to engaging in DHIs. Finally, *inclusive design* provides flexible design and is usable for a broader population. Notably, the design considerations identified in the included papers are not meant to be applied to every project; the implementation depends on the project's focus. Designers always need to balance project requirements (such as profits), user needs (such as privacy concerns), and policy regulations (such as data security). For example, peer-to-peer patient communication may not be appropriate for more sensitive health issues.

Table 2. Design considerations of digital patient experience.

Themes	Considerations	References
Design constructs		
Personalization		
Profiling	• Careful patient selection for digital health use • Assess specific metrics (eg, sociodemographic characteristics, basic health status, individual preferences, and habits) • Create an accurate patient profile	[53,59]
Tailoring	• Provide personalized information, tailored intervention content, and customized feedback (eg, predicted possible causes and consequences of a health problem and advice on the behavior under investigation) • Tailor the content to the user's needs and preferences • Tailor images, colors, text quantity, and font size and color to what users find appealing • Tailor multiple variables rather than a single variable • Align with end users' habitual routines	[49,54,59,80]
Autonomy	• Choose desirable and accessible forms of delivery • Choose when and how to receive reminders • Select or change personalized goals for future use throughout the time span of intervention • Select preferred styles (eg, color and font)	[59,80]
Information		
Content	• Provide comprehensive health information (eg, medical history, test results, and medication information) • Provide appropriate education and training on a health condition • Provide concise information (not overwhelm) • Provide evidence-based information from a credible source (eg, no advertisements and validated advice) • Appropriate encryption and digital health security (eg, password setting and privacy policy)	[48,53,54,59]
Communication	• Provide peer-to-peer communication through web-based forums and communities using instant messages • Access to professionals directly via email, SMS text message, or live chat • Share duties between health care staff	[53,54,59,80]
Functionality	• Rewards (eg, material incentives, intangible rewards, and messages of congratulations when a task is completed) • Reminders (eg, email messages, SMS text messages, words of the day, and pop-ups) for task completion • Reflective feedback, persuasive features, and gaming features (eg, knowledge quizzes and games) • Functional characteristics enable accurate and continuous self-management (eg, activity planning, activity tracking, self-monitoring, and diaries), person-centered care, and sustained behavior change • Appropriate dose of treatment	[48,49,53,54,59,80]
Navigation		
Forms of delivery	• Readily accessible and downloadable • Improve DHI[a] delivery (eg, change from a website to a mobile phone app) • Ability to print and email the information	[49,54]
User flows	• Minimum input (eg, voice commands) • Efficient access to information • Clarify what to do next • Provide search bar and menu bar • Tools and aids to help understand health information and complete health tasks	[48,59]

Themes	Considerations	References
Instruction and tutorials	• Guide users to a greater extent if the design is not self-explanatory • Provide more concrete, explicit, and context-sensitive instructions (eg, a virtual guided tour and extra internal links) • Adopt features from common (ie, familiarized) user interfaces (eg, the iPhone interface) • Provide appropriate education and training on digital health use	[54,80]
Visualization		
Message presentation	• Visualize continuous monitoring data (eg, present data as graphs and tables) • Provide a coherent presentation in terms of colors, pictures, and themes • Simple nontechnical language • Straightforward and concise text • Comprehensive descriptions of actionable message • Provide positive, nonauthoritarian, friendly, and nonjudgmental tone of voice • Multimedia messages (eg, text combined with relevant pictures or video) • Highlight information using various font styles, sizes, and colors	[59]
Interface aesthetic	• Show graphics (ie, visual aids) rather than (too much) text • Provide a pleasing color scheme (eg, bright colors) • Simple interface	[54,59,80]
Design methods		
Co-design and participatory design approaches		
Multistakeholder	• Involve end users and other stakeholders • Include the user at the beginning of the design process	[48,49,54]
Interdisciplinary	• An interdisciplinary approach to the development and implementation	[48,80,81]
User-centered design and human-centered design approaches		
Needs assessment	• Know the needs, capabilities, and environment of users through focus groups, surveys, interviews, and personas • Composing, preparing, and organizing content	[48,81,82]
Usability testing	• Gain early feedback from users through prototypes; benchmark testing, user testing, heuristic analysis, failure modes and effects analysis, and observations in other health care settings	[48,81,82]
Implementation	• Fit the technology to the person, not the person to the technology; pilot testing, task analysis, and reporting mechanism	[81,82]
Monitor and sustain	• Understanding work as imagined often differs from work as done; pre- and posttesting, contextual inquiry, and safety and hazard reporting	[82]
Inclusive design approaches		
Inclusive	• Provide a flexible design that is usable by people with no limitations, as well as by people with functional limitations related to disabilities or old age	[48,50]

[a] DHI: digital health intervention.

Design Implications

On the basis of our findings regarding influencing factors and design considerations for digital PEx, in this section, we define digital PEx and present design guidelines for the implementation of improving PEx in digital services.

Definition of Digital PEx

Our review reveals the absence of a commonly used concept for PEx in digital health. An increasing number of studies have been conducted on surveying PEx, satisfaction with, and expectations in varied digital health. With the growing academic interest in this topic and increasing efforts to address PEx in digital health design practice, a common concept with a concise definition will strengthen and align efforts overall. After reviewing the alignment of widely accepted concepts of PEx, UX, and DHIs with our generated influencing factors, we observed that many of our findings are included in the PEx definition offered by The Beryl Institute. Therefore, by including the sum of all interactions shaped by an organization’s culture, which influence patient perceptions across the continuum of care [33] along with the constructs of UX (people’s perceptions and responses [18,29]), DHIs (digital health technologies [13]), and the determinants (ie, technical, behavioral, and organizational determinants) identified in this review, we propose a concise, practical definition of digital PEx to guide the future design of digital health: "Digital patient experience is the sum of all interactions, affected by a patient's behavioral determinants, framed by digital technologies, and shaped by organizational culture, that influence patient perceptions across the continuum of care channeling digital health." Compared with the original definition of general PEx, this new definition underlines the digital part of health care delivery and includes 2 new determinants (technical and behavioral) that go beyond the organization's culture to clarify what can influence patient perceptions while traveling along a digital care pathway.

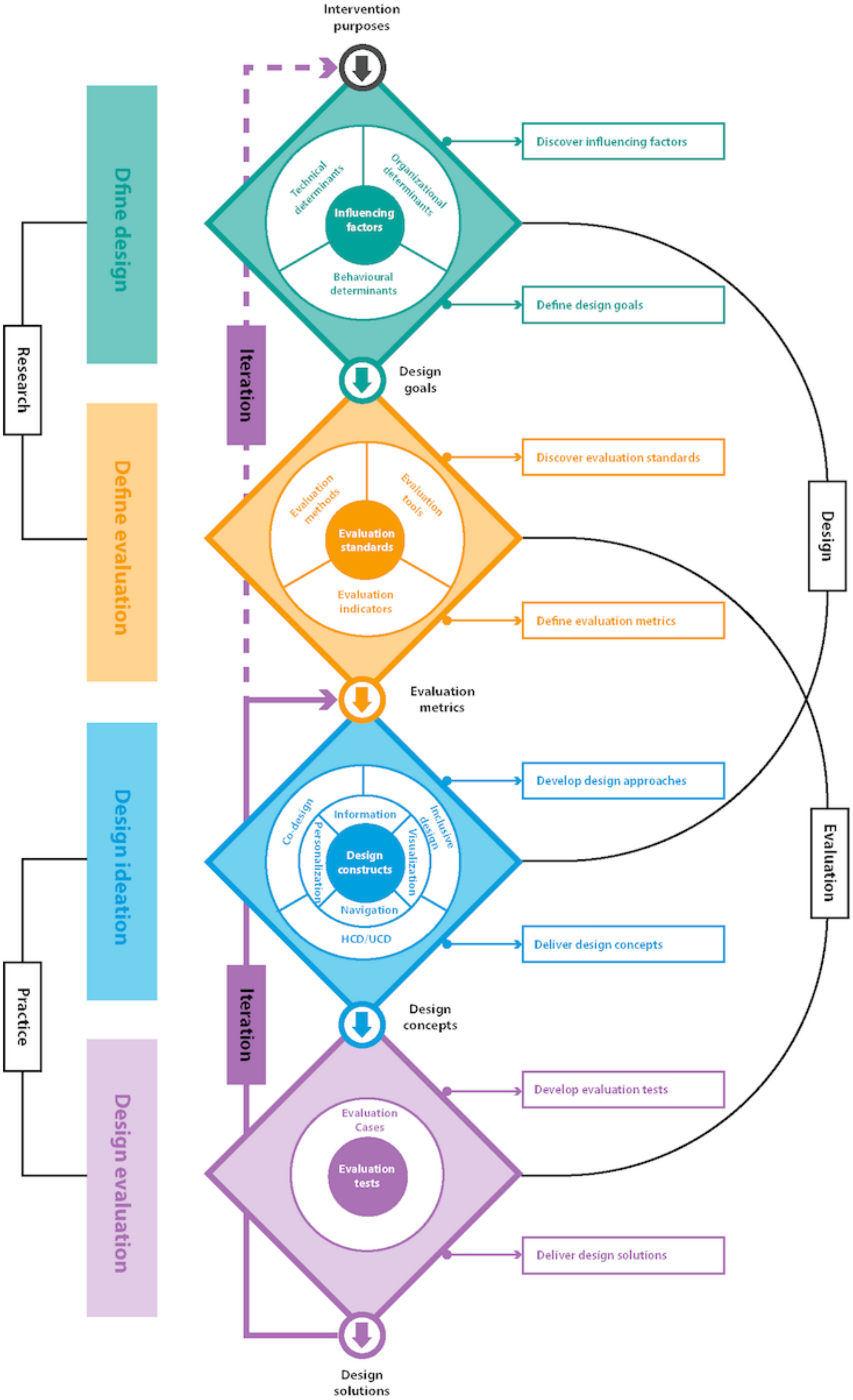
Design Guidelines for Improving Digital PEx

We developed a design and evaluation framework to help digital health designers or developers improve digital PEx in the design process (Figure 2). This framework was based on the findings of this umbrella review and was inspired by the double diamond model [85,86]. Our framework shows four phases: *define design, define evaluation, design ideation*, and *designevaluation*. The first and third phases focus on the design itself, and the second and fourth phases focus on design evaluation. In this study, we focus on explaining the first and third phases. In the first phase, designers must define the design goals by considering the factors that affect digital PEx. In this phase, we provided 3 determinants referring to 9 categories of influencing factors that have 3 types of impact on digital PEx (positive, negative, and double-edged) for designers to discover and explore. Designers can frame their design goals based on the intervention purposes and the selection of influencing factors. For example, if the purpose of the intervention is to improve patient eHealth literacy, designers need to pay more attention to patient capability and frame a design goal to develop suitable intervention functionality for improving patient capability. After defining the design goals, designers can move to the second phase, which is the *define evaluation* phase. In this phase, designers need to consider evaluation indicators (patient emotional, behavioral, and health outcomes) and evaluation methods (surveys and interviews) that are used to assess digital PEx. Detailed information regarding this phase will be discussed in a parallel study. Following this, we provide 4 design constructs (personalization, information, navigation, and visualization) and 3 design methods (ie, HCD or UCD, co-design, and inclusive design) for the design ideation phase. *Personalization* [41,54,56,57,59,69,81,87] refers to ascertaining user needs with design goals. It encompasses the design of intervention technology and functionality needs that meet the patients' ability, opportunity, and motivation to trigger behavior changes and promote health outcomes. *UCD/HCD* and *inclusive design* are valuable at this stage for the inclusion of patient perspectives. Driven by user needs and intervention goals, *information* includes content, communication, and functionality [54,59,81], and *navigation* comprises forms of delivery, user flows, instructions, and tutorials [54,59,80,81]. This relates to how relevant content presented in multimedia with a clear information architecture can attract patient attention and help them understand and complete tasks efficiently [88]. *Co-design* and *participatory design* are multidisciplinary collaborations that are necessary at these 2 stages. Finally, designers need to consider *visualization* [54,57,59,80,81], which determines the product look. The digital health interface can affect patients' first impressions when using DHIs. An attention-grabbing, simple, and consistent interface [59], layout (colors and images) [80], and message presentation [59] can all lead to positive UX. The design guidelines (Textbox 1) can be used at this stage to produce design concepts. In addition, this phase contains the digital health design workflow, challenges, and tips from a design practice perspective (which will be presented in an ongoing interview study). Finally, we ended up with this framework by introducing the design evaluation phase, in which designers need to develop tests (based on evaluation metrics) to evaluate design concepts. If the evaluation outcomes do not meet the evaluation standards, designers can return to the design ideation phase to adjust the design concepts or return to the first phase to reconsider the design goals.

Compared with the original double diamond model, our framework separates the evaluation part from the design part. This aligns with the design research methodology framework [84], which suggests generating success criteria after clarifying design research goals and before producing design support, formulating criteria for success is essential to be able to determine whether the results help achieve this aim. Therefore, we paid equal attention to design and evaluation. In addition, our framework provides detailed reference materials (such as 3 determinants) for each phase to provide designers with more practical support. Notably, in our framework, we retain some typical features of the double diamond model: the first 2 phases are research related, the last 2 phases are practice related, and each phase starts from divergence and ends at convergence.

On the basis of our findings on influencing factors and design considerations, we mapped the combinations of design constructs and design methods into 9 design guidelines to address different influencing factors (Textbox 1), which can be used to guide the design ideation process. Some of the design guidelines uncovered in this study have already been implemented, resulting in a positive digital PEx, such as the digital platform PatientsLikeMe, which aims to empower patients to navigate their health journeys together through peer support, personalized health insights, tailored digital health services, and patient-friendly clinical education [89]. One of the studies pointed out that patients can greatly benefit from using this platform as it improves patient health literacy, and its condition-specific customization may still further improve PEx [90], which aligns with our design guidelines on improving "patient capability" and providing "personalized information."

Figure 2. Digital patient experience design and evaluation framework. HCD: human-centered design; UCD: user-centered design.



Textbox 1. Design guidelines for improving digital patient experience.

Categories and design guidelines

- Patient capability
  - Identify patients' knowledge and skill levels by understanding their technology, language, and health literacy; consider their previous experience and current confidence level in using digital health; improve their actual literacy and correct their perceived inability; tailor design to their ability
- Patient opportunity
  - Profile patients' identity (eg, age, gender, economic status, and daily routines) and health status (eg, illness complexity, severity, and stability); consider patients' accessibility and affordance to digital health; tailor design to their individual opportunity
- Patient motivation
  - Recognize patients' mindset and perceived advantages and disadvantages; inform them of the potential benefits of using digital health; address their concerns and worries; understand their expectations and needs; tailor design to their preferences to trigger their motivation
- Intervention technology
  - Increase technical usability; ensure ease of use, ready to use, and timely feedback on digital health; select technical features (eg, data accessibility) and delivery media or devices (eg, device ownership) to meet patients' preferences and needs
- Intervention functionality
  - Strengthen theory-based interventions (eg, behavior change techniques and evidence-based interventions); improve intervention quality, considering privacy, security, and accuracy issues; provide regular and continuous social support combining both remote communication and real human contact; tailor health promotion and intervention structure to patients' needs and preferences
- Intervention interaction design
  - Provide personalized and consistent information, clear tutorials or technical support, and visualized data; allow patients to choose personalized interactive elements; follow human or user-centered design, co-design, and inclusive design methods; involve multi-stakeholders and multi-disciplines in the design process
- Organizational environment
  - Reduce equipment or service cost and time; improve health care providers' professional ability, communication skills, and service attitudes across the use of digital health; increase workflow transparency and clarify accountability; improve system integration and compatibility
- Physical environment
  - Provide a familiar, warm, and comfortable environment rather than cold and unfamiliar settings; reduce environmental distractions (eg, background noise or lighting)
- Social environment
  - Provide adequate support policies and legislation; develop plausible business cases

Discussion

Principal Findings

We systematically reviewed review articles on factors that influence digital PEx and considerations regarding how best to design digital PEx. The reviews varied greatly in type, including studies and data analysis methods, as well as in HIS, health issues, target patient groups, intervention content, and structure. Of the selected reviews, 62% (28/45) were systematic reviews, the rest were *other* types. These included qualitative, quantitative, and mixed methods studies. Thematic analysis and meta-analysis were the most common data analysis methods used in the reviews. We note that the studies described in the selected reviews were extremely heterogeneous, and information about interventions and digital PEx were often mixed and complex, making comparison difficult.

Our results are in line with the findings reported by previous authors [25,30,38] on the factors that affect PEx, UX, or the implementation of digital health. On the basis of the identified influencing factors and design considerations, we developed 9 design guidelines for improving digital PEx. Our findings reveal that among the selected reviews, only a few formulated design strategies or guidelines. This lack of design knowledge transformation makes it difficult for designers or developers to apply the findings directly. This aligns with the studies by Sakaguchi-Tang et al [48] and Søgaard Neilsen and Wilson [80]; the former indicated that the absence of specific design recommendations impairs the design of digital health, with the latter suggesting that there was a lack of understanding of the most beneficial design aspects for some specific digital health and how design principles can best be applied. Moreover, the use of UCD has been recommended in many studies to address

UX-relevant issues in digital health [3,80,91], which also supports our findings.

Digital PEx Versus General PEx and UX

We found a lack of a common term to describe PEx in digital health; UX (25/45, 56%) and PEx (17/45, 38%) were the most commonly used terms. Patient UX, patient perceptions, client experiences, patient empowerment, and user engagement were also used to describe similar concepts. Many reviews indicated that there was limited information about UX or PEx in varied digital health and underlined the need for a more holistic view of patient needs and priorities to better shape digital health design strategies and provide tailored digital health [28,40,42,45,60,74].

Influencing Factors Are More Complex Than Facilitators and Barriers

The information provided about digital PEx–influencing factors was complex and heterogeneous. Digital health is often treated as a whole, whereas digital PEx is affected by the additive effect of varying digital health factors. A single change in a factor may affect everything else. We found that without a concrete interaction context, factors could be regarded concurrently with facilitators or barriers. For example, regular contact with health care providers (HCPs) could be perceived to increase a sense of reassurance or perceived as a burden to patients' daily lives [28]; some patients experienced digital health as time consuming or an additional burden, whereas others experienced it as time saving or convenient [69]. Some influencing factors may have a soft or indirect influence on digital PEx [44,76]. For instance, users who are completely unaware of privacy or security risks may have excellent experience with digital health that fails to meet privacy or security requirements [76]. A lack of concrete solutions to address these barriers was mentioned [48]. It is likely that digital health cannot serve all populations equally well [71], which aligns with the results of a scoping review that investigated the inequities caused by the adoption of digital health technologies [92]. Some researchers indicated that older adults can also experience benefits by using digital health [53], whereas others suggested that telehealth is, at best, a partial solution for younger and fitter subpopulations [47,71]. Again, although some mentioned that patients preferred using personal devices [49,55,61,64], others noted the opposite [64].

Unclear Benefit From the Different Elements in Digital Health

It is likely that some patient groups benefit more than others from specific DHIs. For example, one of the reviews suggested that in telemedicine treatment for type 2 diabetes, behavioral change and continuous management were the keys to success [43]. However, it was unclear precisely which elements of digital health resulted in patients' satisfaction or dissatisfaction and how they could be addressed [83]. Moreover, we found limited data and even contradictory results on which factors affect digital PEx the most, which elements should be considered first when developing DHIs, and who benefits more from them. The latter is commonly mentioned [48,53,70], with some authors suggesting that patients with unstable chronic diseases might benefit the most [47,93]. However, another review indicated that even if patients are provided with the latest state-of-the-art technology at home, the intervention will not be beneficial if it remains unused [43]. Patients who are less activated are likely to have less positive experiences than those who are highly engaged [74].

Lack of Multiple Perspectives During the Design of Digital PEx

Clear communication between experts, designers, and patients regarding their understanding of digital PEx is required. Some reviews acknowledged the need for a multistakeholder perspective on digital PEx [55,69]. However, we found circumstances in which this was not possible. For example, in some cases, UCD for DHIs was conducted on nonpatient users either because of ethical reasons or relevant regulations [43], and in others, apps that are not specifically designed for patients with cancer were being used for this patient group [64]. HCPs are often isolated from the decision-making process to incorporate digital health into their current service provision [28]. Moreover, a lack of clinician perceptions of digital health use was also reported [40,54]. Furthermore, no studies focused on exploring designers' views, opinions, experiences, or values in addressing PEx or UX in the design of digital health. There was little information on whether experienced designers had worked with patients in their design process.

Over- or Underestimated Results

Some studies suggested that a lack of interest was the main reason for patients' refusal of digital health and that reasons for patient withdrawal were patients not wanting to use equipment, deteriorating health, and technical problems [94,95]. We need to gain better insights into the reasons for patients choosing not to engage in or withdraw from digital health, as these will significantly inform future DHI development and design [43,53,69]. However, it is likely that most studies only included patients who had already agreed to or were using digital health technologies; those who refused to use, withdrew from, or had no accessibility were excluded [28,51,63,69]. One of the reviews suggested that this would result in over- or underestimated results of DHIs' effects on digital PEx, as participants who completed the intervention may differ from those who did not [41]. Another review found that patients only reported positive themes associated with remote monitoring, which may indicate a selection bias [71].

Conflicts Between Benefits and Cost for Developing DHIs

The provision of digital health can reduce the treatment burden and better integrate care into patients' daily routines [69], which is consistent with our findings; we found that most reviews had a positive perspective of DHIs. However, in one of the reviews, it was suggested that although there was agreement among most professionals that health information technology can have a positive impact on PEx, when weighing the benefits against the potential cost, demonstrating this will be challenging [44,47]. Moreover, unnecessary high-frequency monitoring could result in a waste of health resources and an increased workload for HCPs [52]. Compared with existing health care services, the application of new technology needs to demonstrate clinical evidence of improved health conditions [43]. However, there were discordant findings in terms of the benefits of using DHIs. For example, there was no concrete evidence that telemedicine consultations were quicker than face-to-face consultations [40,57,68,83]. In another case, the impact of DHIs on health care use was not examined [57]. In conclusion, only user-friendly and quality-certified DHIs should be provided to patients [64]; health care organizations should not shift their focus from the basic and inexpensive strategies that affect patient care. Care is needed: new technology should not overwhelm the patient or ignore patient needs [44].

Limitations

First, when undertaking a review of review articles, some important details included in the original studies may have been lost, which increases the possibility of reporting bias. We also noted differences in the interpretation of terms and methods between the reviews. There is a lack of consistency in the terminology used to describe the functions of DHIs, HISs, or digital PEx itself. For example, in some cases, "eHealth" and "mHealth" were used as interchangeable terms [75], "persuasive technology" and "behavior change techniques" were presented as having a similar meaning [43], and "patient engagement" and "patient activation" were also regarded as being the same [74]. This inconsistent use of terms may impede knowledge translation and dissemination [57]. To counter this, we summarized the varied factors with unified descriptions to build a common understanding of the digital PEx–influencing factors.

Second, the intervention types and patient groups varied widely among the reviews, limiting meaningful comparisons between different studies. In addition, the digital health landscape is rapidly evolving, and the technology infrastructure is constantly shifting [41], as are the continuous updates of the UX design area. It is important to keep the influencing factors updated or adapted as the technology develops. Possibly, relevant original studies may have been excluded because of our focus on review papers. However, our approach to conducting an overarching review provides readers with a quick overview of the relevant digital PEx studies and a basis for further research.

Third, our umbrella review did not account for the multimodal relationships between subthemes or the potential overlap between subthemes within different domains. For example, different subthemes, such as "personalized design" in "interventions' interaction design" also interconnect with "interventions' technology" and "interventions' functionality." Moreover, our review process did not aim to address the question of whether some influencing factors are more important than others or how different aspects of DHIs influence them. This warrants further investigation as we suspect that differences may exist between the influencing factors, as some elements in digital health are more likely to increase or inhibit a positive digital PEx.

Finally, as we used qualitative thematic analysis to synthesize the findings and generate themes, the generated themes could have been influenced by the authors' previous research experiences and personal understanding. By asking other researchers to repeat the coding process, the resulting themes are likely to be different. However, to minimize the potential coding bias, the generation of categories was based on the PRISM framework; 4 researchers with different backgrounds, including design, medical, and human factors, were involved in the iterative coding process, group discussion, and independent and random validation, and existing theories were used.

Further Research

The goals of this umbrella review were to systematically review the influencing factors that affect digital PEx and the design considerations for improving digital PEx that are summarized in the existing literature. We must conclude that, currently, much remains unknown, and the topic of digital PEx is relatively new. We propose 6 directions that require further research. The first direction is to develop frameworks or models that translate digital PEx–related research findings into design practices or implications. For example, in this study, we used design guidelines and a design framework to summarize the findings. The second direction is to identify those who will benefit more from which elements in DHIs and which influencing factors could be addressed by combining design constructs and design methods. The third direction is to further examine how designers understand and address digital PEx in the digital health design process. To address this, we conducted a qualitative study on how designers address digital PEx in design practice. The fourth direction is to standardize evaluation indicators, methods, or tools for assessing digital PEx; we are currently evaluating digital PEx in a parallel study. The fifth direction is to quantify the balance between the benefits and costs of developing user-friendly and validated DHIs. The sixth direction is to identify participants' reasons for dropping out and their impact on the reported digital PEx–related results.

Conclusions

To the best of our knowledge, this is the first study to propose the term "digital patient experience" as a common phrase to describe PEx in digital health and define digital PEx by synthesizing the reported PEx or UX of varied DHIs from multiple reviews. Multimedia Appendix 4 shows more details about the structure of this study. In this review, information on influencing factors was identified and summarized into 9 categories (ie, patient capability, opportunity, motivation, intervention technology, functionality, interaction design, organizational, physical environment, and social environment). These categories were classified into positive, negative, and double-edged factors based on their positive, negative, and diverse impacts on digital PEx. Our review uncovered 4 design constructs (personalized, information, navigation, and visual design) and 3 common design methods (UCD or HCD, co-design, and inclusive design) as design considerations for addressing digital PEx. Finally, we proposed a design and evaluation framework and design guidelines to help digital health designers and developers address digital PEx throughout the entire design process.

Acknowledgments

This work was supported by the China Scholarship Council under grant 201906790084.

Authors' Contributions

TW conceived, designed, and led the overall study, conducted data collection, led the data analysis and interpretation, and drafted the manuscript. TW and GG performed the review selection and data interpretation. GG, MM, and RG participated in the overall study; contributed to the analysis and interpretation of the study data; and conceptualized, reviewed, and suggested modifications regarding the presentation of results. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study characteristics and digital health intervention characteristics of included reviews.
[DOCX File , 48 KB - jmir_v24i8e37952_app1.docx]

Multimedia Appendix 2
Influencing factors on digital patient experience (double-edged factors imply diverse impact, positive factors imply positive impact, and negative factors imply negative impact).
[DOCX File , 29 KB - jmir_v24i8e37952_app2.docx]

Multimedia Appendix 3
Detailed information on themes of influencing factors of the digital patient experience.
[DOCX File , 36 KB - jmir_v24i8e37952_app3.docx]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - jmir_v24i8e37952_app4.docx]

References

1. Second round of the national pulse survey on continuity of essential health services during the COVID-19 pandemic. World Health Organization. 2021 Apr 23. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-EHS-continuity-survey-2021.1 [accessed 2021-12-06]
2. Mehrotra A, Ray K, Brockmeyer DM, Barnett ML, Bender JA. Rapidly converting to “Virtual Practices”: outpatient care in the era of Covid-19. NEJM Catalyst Innovations in Care Delivery. 2020 Apr 1. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0091 [accessed 2021-12-06]
3. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. Pediatr Res 2016 Jan;79(1-2):169-176. [doi: 10.1038/pr.2015.192] [Medline: 26466080]
4. Atlas of eHealth country profiles: the use of eHealth in support of universal health coverage. World Health Organization. 2016 Jan 1. URL: https://www.who.int/publications/i/item/9789241565219 [accessed 2021-12-06]
5. Davis SW, Oakley-Girvan I. mHealth education applications along the cancer continuum. J Cancer Educ 2015 Jun;30(2):388-394. [doi: 10.1007/s13187-014-0761-4] [Medline: 25482319]
6. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
7. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014 May 20;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
8. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital health interventions for adults with type 2 diabetes: qualitative study of patient perspectives on diabetes self-management education and support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
9. McLean G, Band R, Saunderson K, Hanlon P, Murray E, Little P, DIPSS co-investigators. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. J Hypertens 2016 Apr;34(4):600-612 [FREE Full text] [doi: 10.1097/HJH.0000000000000859] [Medline: 26845284]
10. Escriva Boulley G, Leroy T, Bernetière C, Paquienseguy F, Desfriches-Doria O, Préau M. Digital health interventions to help living with cancer: a systematic review of participants' engagement and psychosocial effects. Psychooncology 2018 Dec;27(12):2677-2686. [doi: 10.1002/pon.4867] [Medline: 30152074]

11. Augenstein J. Opportunities to expand telehealth use amid the coronavirus pandemic. Health Affairs. 2020 Mar 16. URL: https://www.healthaffairs.org/do/10.1377/forefront.20200315.319008/full/ [accessed 2021-04-26]
12. Tecco H. 2016 Year end funding report: a reality check for digital health. Rock Health. 2017 Jan 3. URL: https://rockhealth.com/insights/2016-year-end-funding-report-a-reality-check-for-digital-health/ [accessed 2022-05-17]
13. Digital implementation investment guide (DIIG): integrating digital interventions into health programmes. World Health Organization. 2020 Sep 15. URL: https://www.who.int/publications/i/item/9789240010567 [accessed 2021-05-13]
14. 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]
15. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
16. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
17. Alkire (née Nasr) L, O'Connor GE, Myrden S, Köcher S. Patient experience in the digital age: an investigation into the effect of generational cohorts. J Retail Consum Serv 2020 Nov;57:102221. [doi: 10.1016/j.jretconser.2020.102221]
18. Bolton RN, McColl-Kennedy JR, Cheung L, Gallan A, Orsingher C, Witell L, et al. Customer experience challenges: bringing together digital, physical and social realms. J Serv Manag 2018 Sep 07;29(5):776-808. [doi: 10.1108/josm-04-2018-0113]
19. Kneeland PP. Patient experience. In: Anderson ME, Glasheen JJ, editors. Hospital Medicine Clinics. Amsterdam, The Netherlands: Elsevier Health Sciences; Jan 2016:137-151.
20. Reeves R, Coulter A, Jenkinson CJ, Cartwright J, Bruster S, Richards N. Development and pilot testing of questionnaires for use in the acute NHS trust inpatient survey programme. Europe: Picker Institute. 2002. URL: https://www.researchgate.net/publication/230687421_Development_and_pilot_testing_of_questionnaire_for_use_in_the_acute_NHS_Trust_inpatient_survey_programme [accessed 2021-05-13]
21. Richardson B, Campbell-Yeo M, Smit M. Mobile application user experience checklist: a tool to assess attention to core UX principles. Int J Hum Comput Interact 2021 Feb 01;37(13):1283-1290. [doi: 10.1080/10447318.2021.1876361]
22. Kellermann AL, Jones SS. What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Aff (Millwood) 2013 Jan;32(1):63-68. [doi: 10.1377/hlthaff.2012.0693] [Medline: 23297272]
23. Larivière B, Bowen D, Andreassen TW, Kunz W, Sirianni NJ, Voss C, et al. “Service Encounter 2.0”: an investigation into the roles of technology, employees and customers. J Bus Res 2017 Oct;79:238-246. [doi: 10.1016/j.jbusres.2017.03.008]
24. Defining Patient Experience. The Beryl Institute. URL: https://www.theberylinstitute.org/page/DefiningPX [accessed 2021-12-06]
25. Staniszewska S, Boardman F, Gunn L, Roberts J, Clay D, Seers K, et al. The Warwick Patient Experiences Framework: patient-based evidence in clinical guidelines. Int J Qual Health Care 2014 Apr;26(2):151-157. [doi: 10.1093/intqhc/mzu003] [Medline: 24556816]
26. NHS Patient Experience Framework. Department of Health, National Health Service. 2011. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215159/dh_132788.pdf [accessed 2021-03-07]
27. Shandley LM, Hipp HS, Anderson-Bialis J, Anderson-Bialis D, Boulet SL, McKenzie LJ, et al. Patient-centered care: factors associated with reporting a positive experience at United States fertility clinics. Fertil Steril 2020 Apr;113(4):797-810. [doi: 10.1016/j.fertnstert.2019.12.040] [Medline: 32147181]
28. Brunton L, Bower P, Sanders C. The contradictions of telehealth user experience in chronic obstructive pulmonary disease (COPD): a qualitative meta-synthesis. PLoS One 2015 Oct 14;10(10):e0139561 [FREE Full text] [doi: 10.1371/journal.pone.0139561] [Medline: 26465333]
29. Jokela T, Iivari N, Matero J, Karukka M. The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In: Proceedings of the Latin American conference on Human-computer interaction. 2003 Presented at: CLIHC '03; August 17-20, 2003; Rio de Janeiro, Brazil p. 53-60. [doi: 10.1145/944519.944525]
30. Morville P. Experience design unplugged. In: ACM SIGGRAPH 2005 Web program. 2005 Presented at: SIGGRAPH '05; July 31-August 4, 2005; Los Angeles, CA, USA p. 10-es. [doi: 10.1145/1187335.1187347]
31. Mobasheri MH, Johnston M, King D, Leff D, Thiruchelvam P, Darzi A. Smartphone breast applications - what's the evidence? Breast 2014 Oct;23(5):683-689 [FREE Full text] [doi: 10.1016/j.breast.2014.07.006] [Medline: 25153432]
32. Foley NM, O'Connell EP, Lehane EA, Livingstone V, Maher B, Kaimkhani S, et al. PATI: patient accessed tailored information: a pilot study to evaluate the effect on preoperative breast cancer patients of information delivered via a mobile application. Breast 2016 Dec;30:54-58. [doi: 10.1016/j.breast.2016.08.012] [Medline: 27611236]
33. Wolf JA, Niederhauser V, Marshburn D, LaVela SL. Defining patient experience. Patient Exp J 2014;1(1):7-19.
34. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]

35. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
36. Frank SR. Digital health care--the convergence of health care and the Internet. J Ambul Care Manage 2000 Apr;23(2):8-17. [doi: 10.1097/00004479-200004000-00003] [Medline: 10848396]
37. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Aqil A, Lippeveld T, Hozumi D. PRISM framework: a paradigm shift for designing, strengthening and evaluating routine health information systems. Health Policy Plan 2009 May;24(3):217-228 [FREE Full text] [doi: 10.1093/heapol/czp010] [Medline: 19304786]
39. 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]
40. Swanepoel DW, Hall 3rd JW. A systematic review of telehealth applications in audiology. Telemed J E Health 2010 Mar;16(2):181-200. [doi: 10.1089/tmj.2009.0111] [Medline: 20187743]
41. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of Web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013 Feb 20;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
42. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014 Feb 18;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]
43. Jalil S, Myers T, Atkinson I. A meta-synthesis of behavioral outcomes from telemedicine clinical trials for type 2 diabetes and the Clinical User-Experience Evaluation (CUE). J Med Syst 2015 Mar;39(3):28. [doi: 10.1007/s10916-015-0191-9] [Medline: 25677954]
44. Werder M. Health information technology: a key ingredient of the patient experience. Patient Exp J 2015 Apr 29;2(1):143-147. [doi: 10.35680/2372-0247.1071]
45. Jones L, Grech C. The patient experience of remote telemonitoring for heart failure in the rural setting: a literature review. Contemp Nurse 2016;52(2-3):230-243. [doi: 10.1080/10376178.2015.1111154] [Medline: 26525077]
46. Stokke R. The personal emergency response system as a technology innovation in primary health care services: an integrative review. J Med Internet Res 2016 Jul 14;18(7):e187 [FREE Full text] [doi: 10.2196/jmir.5727] [Medline: 27417422]
47. Greenhalgh T, A'Court C, Shaw S. Understanding heart failure; explaining telehealth - a hermeneutic systematic review. BMC Cardiovasc Disord 2017 Jun 14;17(1):156 [FREE Full text] [doi: 10.1186/s12872-017-0594-2] [Medline: 28615004]
48. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient portal use and experience among older adults: systematic review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
49. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
50. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
51. Ames HM, Glenton C, Lewin S, Tamrat T, Akama E, Leon N. Clients' perceptions and experiences of targeted digital communication accessible via mobile devices for reproductive, maternal, newborn, child, and adolescent health: a qualitative evidence synthesis. Cochrane Database Syst Rev 2019 Oct 14;10(10):CD013447 [FREE Full text] [doi: 10.1002/14651858.CD013447] [Medline: 31608981]
52. Cheung KL, Durusu D, Sui X, de Vries H. How recommender systems could support and enhance computer-tailored digital health programs: a scoping review. Digit Health 2019 Jan 24;5:2055207618824727 [FREE Full text] [doi: 10.1177/2055207618824727] [Medline: 30800414]
53. De La Cruz Monroy MF, Mosahebi A. The use of smartphone applications (apps) for enhancing communication with surgical patients: a systematic review of the literature. Surg Innov 2019 Apr;26(2):244-259. [doi: 10.1177/1553350618819517] [Medline: 30602332]
54. Lim S, Tan A, Madden S, Hill B. Health professionals' and postpartum women's perspectives on digital health interventions for lifestyle management in the postpartum period: a systematic review of qualitative studies. Front Endocrinol (Lausanne) 2019 Nov 8;10:767 [FREE Full text] [doi: 10.3389/fendo.2019.00767] [Medline: 31787932]
55. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider- and patient-related barriers to and facilitators of digital health technology adoption for hypertension management: scoping review. JMIR Cardio 2019 Mar 26;3(1):e11951 [FREE Full text] [doi: 10.2196/11951] [Medline: 31758771]
56. Brigden A, Anderson E, Linney C, Morris R, Parslow R, Serafimova T, et al. Digital behavior change interventions for younger children with chronic health conditions: systematic review. J Med Internet Res 2020 Jul 31;22(7):e16924 [FREE Full text] [doi: 10.2196/16924] [Medline: 32735227]

57. Eze ND, Mateus C, Cravo Oliveira Hashiguchi T. Telemedicine in the OECD: an umbrella review of clinical and cost-effectiveness, patient experience and implementation. PLoS One 2020 Aug 13;15(8):e0237585 [FREE Full text] [doi: 10.1371/journal.pone.0237585] [Medline: 32790752]
58. Ingemann C, Hansen NF, Hansen NL, Jensen K, Larsen CV, Chatwood S. Patient experience studies in the circumpolar region: a scoping review. BMJ Open 2020 Oct 05;10(10):e042973 [FREE Full text] [doi: 10.1136/bmjopen-2020-042973] [Medline: 33020108]
59. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
60. Memon M, Wagner SR, Pedersen CF, Beevi FH, Hansen FO. Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes. Sensors (Basel) 2014 Mar 04;14(3):4312-4341 [FREE Full text] [doi: 10.3390/s140304312] [Medline: 24599192]
61. Firth J, Torous J. Smartphone apps for schizophrenia: a systematic review. JMIR Mhealth Uhealth 2015 Nov 06;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
62. Liddy C, Drosinis P, Keely E. Electronic consultation systems: worldwide prevalence and their impact on patient care-a systematic review. Fam Pract 2016 Jun;33(3):274-285. [doi: 10.1093/fampra/cmw024] [Medline: 27075028]
63. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
64. Rincon E, Monteiro-Guerra F, Rivera-Romero O, Dorronzoro-Zubiete E, Sanchez-Bocanegra CL, Gabarron E. Mobile phone apps for quality of life and well-being assessment in breast and prostate cancer patients: systematic review. JMIR Mhealth Uhealth 2017 Dec 04;5(12):e187 [FREE Full text] [doi: 10.2196/mhealth.8741] [Medline: 29203459]
65. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
66. Wesselman LM, Hooghiemstra AM, Schoonmade LJ, de Wit MC, van der Flier WM, Sikkes SA. Web-based multidomain lifestyle programs for brain health: comprehensive overview and meta-analysis. JMIR Ment Health 2019 Apr 09;6(4):e12104 [FREE Full text] [doi: 10.2196/12104] [Medline: 30964438]
67. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. J Am Med Inform Assoc 2020 Jun 01;27(6):939-945 [FREE Full text] [doi: 10.1093/jamia/ocaa029] [Medline: 32357368]
68. O'Keefe M, White K, Jennings JC. Asynchronous telepsychiatry: a systematic review. J Telemed Telecare 2021 Apr;27(3):137-145. [doi: 10.1177/1357633X19867189] [Medline: 31357908]
69. Cox A, Lucas G, Marcu A, Piano M, Grosvenor W, Mold F, et al. Cancer survivors' experience with telehealth: a systematic review and thematic synthesis. J Med Internet Res 2017 Jan 09;19(1):e11 [FREE Full text] [doi: 10.2196/jmir.6575] [Medline: 28069561]
70. Barken TL, Söderhamn U, Thygesen E. A sense of belonging: a meta-ethnography of the experience of patients with chronic obstructive pulmonary disease receiving care through telemedicine. J Adv Nurs 2019 Dec;75(12):3219-3230. [doi: 10.1111/jan.14117] [Medline: 31225664]
71. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.013] [Medline: 30784430]
72. Leonardsen AL, Hardeland C, Helgesen AK, Grøndahl VA. Patient experiences with technology enabled care across healthcare settings- a systematic review. BMC Health Serv Res 2020 Aug 24;20(1):779 [FREE Full text] [doi: 10.1186/s12913-020-05633-4] [Medline: 32838784]
73. Steindal SA, Nes AA, Godskesen TE, Dihle A, Lind S, Winger A, et al. Patients' experiences of telehealth in palliative home care: scoping review. J Med Internet Res 2020 May 05;22(5):e16218 [FREE Full text] [doi: 10.2196/16218] [Medline: 32369037]
74. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 8;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
75. Feather JS, Howson M, Ritchie L, Carter PD, Parry DT, Koziol-McLain J. Evaluation methods for assessing users' psychological experiences of Web-based psychosocial interventions: a systematic review. J Med Internet Res 2016 Jun 30;18(6):e181 [FREE Full text] [doi: 10.2196/jmir.5455] [Medline: 27363519]
76. Baumel A, Birnbaum ML, Sucala M. A systematic review and taxonomy of published quality criteria related to the evaluation of user-facing eHealth programs. J Med Syst 2017 Aug;41(8):128. [doi: 10.1007/s10916-017-0776-6] [Medline: 28735372]
77. Rising KL, Ward MM, Goldwater JC, Bhagianadh D, Hollander JE. Framework to advance oncology-related telehealth. JCO Clin Cancer Inform 2018 Dec;2:1-11 [FREE Full text] [doi: 10.1200/CCI.17.00156] [Medline: 30652582]

78. Bashi N, Fatehi F, Mosadeghi-Nik M, Askari MS, Karunanithi M. Digital health interventions for chronic diseases: a scoping review of evaluation frameworks. BMJ Health Care Inform 2020 Mar;27(1):e100066 [FREE Full text] [doi: 10.1136/bmjhci-2019-100066] [Medline: 32156751]
79. Lemon C, Huckvale K, Carswell K, Torous J. A narrative review of methods for applying user experience in the design and assessment of mental health smartphone interventions. Int J Technol Assess Health Care 2020;36(1):64-70. [doi: 10.1017/S0266462319003507] [Medline: 31973787]
80. Søgaard Neilsen A, Wilson RL. Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: an integrative literature review. Int J Ment Health Nurs 2019 Feb;28(1):22-39. [doi: 10.1111/inm.12527] [Medline: 30133096]
81. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
82. Fouquet SD, Miranda AT. Asking the right questions-human factors considerations for telemedicine design. Curr Allergy Asthma Rep 2020 Aug 29;20(11):66 [FREE Full text] [doi: 10.1007/s11882-020-00965-x] [Medline: 32862299]
83. Chaudhry H, Nadeem S, Mundi R. How satisfied are patients and surgeons with telemedicine in orthopaedic care during the COVID-19 pandemic? A systematic review and meta-analysis. Clin Orthop Relat Res 2021 Jan 01;479(1):47-56 [FREE Full text] [doi: 10.1097/CORR.0000000000001494] [Medline: 33009231]
84. Blessing LT, Chakrabarti A. DRM, a Design Research Methodology. London, UK: Springer; 2009.
85. Design methods for developing services. Design Council. 2015 Mar 17. URL: https://www.designcouncil.org.uk/our-work/skills-learning/resources/design-methods-developing-services/ [accessed 2021-12-10]
86. Ball J. The Double Diamond Process Model. Design Council. 2019 Oct 1. URL: https://www.designcouncil.org.uk/our-work/news-opinion/double-diamond-universally-accepted-depiction-design-process/ [accessed 2022-05-17]
87. Slater H, Briggs A, Stinson J, Campbell JM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: a qualitative systematic review protocol. JBI Database System Rev Implement Rep 2017 Aug;15(8):2047-2054. [doi: 10.11124/JBISRIR-2016-003299] [Medline: 28800054]
88. Dekkers T. Data-driven Patient Profiles: definition, validation, and implementation for tailored orthopaedic healthcare services. Delft University of Technology. 2020 Sep 3. URL: https://research.tudelft.nl/en/publications/data-driven-patient-profiles-definition-validation-and-implementa [accessed 2021-05-13]
89. PatientsLikeMe. URL: https://www.patientslikeme.com/about [accessed 2022-05-17]
90. Wicks P, Mack Thorley E, Simacek K, Curran C, Emmas C. Scaling PatientsLikeMe via a "generalized platform" for members with chronic illness: Web-based survey study of benefits arising. J Med Internet Res 2018 May 07;20(5):e175 [FREE Full text] [doi: 10.2196/jmir.9909] [Medline: 29735472]
91. Vagal A, Wahab S, Lecky S, Washburn E, Schwartz R, Vogel C, et al. Optimizing patient experience using human-centered design. J Am Coll Radiol 2020 May;17(5):668-672. [doi: 10.1016/j.jacr.2019.11.020] [Medline: 31891673]
92. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]
93. Zanaboni P, Ngangue P, Mbemba GI, Schopf TR, Bergmo TS, Gagnon MP. Methods to evaluate the effects of Internet-based digital health interventions for citizens: systematic review of reviews. J Med Internet Res 2018 Jun 07;20(6):e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]
94. Gorst SL, Armitage C, Hawley M, Coates E. Exploring patient beliefs and perceptions about sustained use of telehealth. Int J Integr Care 2013 Nov 20;13(7). [doi: 10.5334/ijic.1393]
95. Sanders C, Rogers A, Bowen R, Bower P, Hirani S, Cartwright M, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. BMC Health Serv Res 2012 Jul 26;12:220 [FREE Full text] [doi: 10.1186/1472-6963-12-220] [Medline: 22834978]

Abbreviations

DHI: digital health intervention
HCD: human-centered design
HCP: health care provider
HIS: health information system
PEx: patient experience
PRISM: Performance of Routine Information System Management
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UCD: user-centered design
UX: user experience

Edited by T Leung; submitted 17.03.22; peer-reviewed by H Mehdizadeh, H Veldandi, FM Calisto; comments to author 15.05.22; revised version received 26.05.22; accepted 31.05.22; published 04.08.22.

Please cite as:
Wang T, Giunti G, Melles M, Goossens R
Digital Patient Experience: Umbrella Systematic Review
J Med Internet Res 2022;24(8):e37952
URL: https://www.jmir.org/2022/8/e37952
doi:10.2196/37952
PMID:35925651

Review

Self-management Interventions for People With Parkinson Disease: Scoping Review

Madison Milne-Ives[1], BAS, MSc; Camille Carroll[2], BSc, BM, BCh, MRCP, PhD; Edward Meinert[1,3,4], MA, MSc, MBA, MPA, PhD

[1]Centre for Health Technology, University of Plymouth, Plymouth, United Kingdom
[2]Peninsula Medical School, Faculty of Health, University of Plymouth, Plymouth, United Kingdom
[3]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom
[4]Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States

Corresponding Author:
Edward Meinert, MA, MSc, MBA, MPA, PhD
Centre for Health Technology
University of Plymouth
6 Kirkby Place, Room 2
Plymouth, PL4 6DN
United Kingdom
Phone: 44 01752600600
Email: edward.meinert@plymouth.ac.uk

Abstract

Background: Parkinson disease can impose substantial distress and costs on patients, their families and caregivers, and health care systems. To address these burdens for families and health care systems, there is a need to better support patient self-management. To achieve this, an overview of the current state of the literature on self-management is needed to identify what is being done, how well it is working, and what might be missing.

Objective: The aim of this scoping review was to provide an overview of the current body of research on self-management interventions for people with Parkinson disease and identify any knowledge gaps.

Methods: The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) and Population, Intervention, Comparator, Outcome, and Study type frameworks were used to structure the methodology of the review. Due to time and resource constraints, 1 reviewer systematically searched 4 databases (PubMed, Ovid, Scopus, and Web of Science) for the evaluations of self-management interventions for Parkinson disease published in English. The references were screened using the EndNote X9 citation management software, titles and abstracts were manually reviewed, and studies were selected for inclusion based on the eligibility criteria. Data were extracted into a pre-established form and synthesized in a descriptive analysis.

Results: There was variation among the studies on study design, sample size, intervention type, and outcomes measured. The randomized controlled trials had the strongest evidence of effectiveness: 5 out of 8 randomized controlled trials found a significant difference between groups favoring the intervention on their primary outcome, and the remaining 3 had significant effects on at least some of the secondary outcomes. The 2 interventions included in the review that targeted mental health outcomes both found significant changes over time, and the 3 algorithms evaluated performed well. The remaining studies examined patient perceptions, acceptability, and cost-effectiveness and found generally positive results.

Conclusions: This scoping review identified a wide variety of interventions designed to support various aspects of self-management for people with Parkinson disease. The studies all generally reported positive results, and although the strength of the evidence varied, it suggests that self-management interventions are promising for improving the care and outcomes of people with Parkinson disease. However, the research tended to focus on the motor aspects of Parkinson disease, with few nonmotor or holistic interventions, and there was a lack of evaluation of cost-effectiveness. This research will be important to providing self-management interventions that meet the varied and diverse needs of people with Parkinson disease and determining which interventions are worth promoting for widespread adoption.

(J Med Internet Res 2022;24(8):e40181) doi:10.2196/40181

KEYWORDS

Parkinson disease; self-management; self-care; home nursing; self-efficacy; quality of life; signs and symptoms; health behaviour

Introduction

Background

Parkinson disease has a substantial impact on patients, their caregivers and families, and health care systems globally [1,2]. The United Kingdom's aging population is expected to nearly double the prevalence of Parkinson disease by 2065 [3]. The National Health Service (NHS) Long Term Plan has emphasized the need for supported self-management to improve patient outcomes and reduce the strain of an aging population on the health care system [4]. Self-management support interventions for chronic illnesses have been demonstrated to decrease health care use without negatively affecting patient health outcomes [5], but research is still needed on how they are used by and affect all users (including patients, caregivers, and health care professionals) [6]. The NHS has estimated that 25% to 40% of patients have low self-management knowledge, skills, and confidence (patient activation) [7]. A recent study focusing on people with Parkinson disease found that more than half of the patients rated themselves high on patient activation, whereas perceived self-management support was rated much lower [8].

Although Parkinson disease itself is not fatal, its complications and motor and nonmotor symptoms can have serious negative effects on the quality of life for both patients and care partners (CPs). The nonmotor symptoms of Parkinson disease are often undeclared in routine appointments [9] but can have severe negative effects on symptom burden and the quality of life [10,11]. For instance, it has been estimated that around half of the people with Parkinson disease have a mental health comorbidity [12]. Parkinson disease has a substantial impact on patients, their CPs and families, and health care systems [1,13,14]. Successful self-management is associated with improvements in chronic conditions and achieved by supported self-efficacy [15]. For all Parkinson disease symptoms, there are pharmacological and nonpharmacological approaches to management. Self-management interventions focus on the nonpharmacological approaches to symptoms by providing people with Parkinson disease and CPs with support to identify and monitor their symptoms and behavioral approaches to manage their symptoms [16,17].

Preliminary Literature Review

Previous systematic reviews have examined various aspects of support for people with Parkinson disease, particularly interventions that support a shift toward more home-based care, but none were identified that provided a comprehensive overview of self-management interventions. There are 2 recent systematic reviews that examined the use of digital technologies and wearables to monitor or support the care of people with Parkinson disease and provide a comprehensive and recent overview of the available technologies, what they are being used for, and how they are being evaluated [18,19]. A recent preprint review provided an overview of the trends in research in the use of mobile and wearable technology for Parkinson disease over the past decade and identified 4 main applications: assisting with diagnosis, monitoring and prognosis, predicting the outcomes of treatments, and therapy [20]. A scoping review conducted in 2018 summarized the literature about home-based rehabilitation interventions [21].

There were 2 reviews that focused specifically on self-management for people with Parkinson disease [22,23]. The first was a systematic review of the qualitative experience of self-management components by people with Parkinson disease and their carers [22]. This review identified 7 key aspects of self-management interventions for people with Parkinson disease: "(1) medication management, (2) physical exercise, (3) self-monitoring techniques, (4) psychological strategies, (5) maintaining independence, (6) encouraging social engagement, and (7) providing knowledge and information" [22]. However, it did not provide an overview or evaluation of the impact of the self-management interventions on health, behavioral, or other outcomes. The other review was an integrative literature review, which provided an overview of the characteristics of self-management support programs for people with Parkinson disease and their effectiveness [23]. It identified a wide variety of interventions, most of which were specific to Parkinson disease, but found limited evidence of their effectiveness. The review provided a good summary of the state of the field but was conducted in 2016 and did not examine the integration of digital technologies in self-management interventions.

Rationale

A search of the international prospective register of systematic reviews (PROSPERO) also did not find any relevant reviews on self-management and Parkinson disease in progress. A search for "parkinson AND (digital OR technolog*) AND (self-management OR home based care)" only retrieved 4 registrations: 1 focused on diabetes, and the others included a range of neurological conditions. A broader search for "parkinson AND self-management" identified 1 relevant registration—a systematic review and meta-analysis of self-management interventions in Parkinson disease. However, the registration is 2 years old (published on PROSPERO on April 15, 2019), has not been updated, and was not identified in a search for a published final article [24].

Given the rapid evolution of digital technology [25] and its growing role in health care [4], the state of the literature on self-management interventions has likely changed since the 2016 review was conducted, necessitating an updated overview that intentionally includes digital interventions, which are becoming a desired support for Parkinson disease care [26]. The variety of self-management aspects and applications identified in previous reviews indicates that an overview of the different types of self-management interventions and their potential impact is needed. The needs emphasized by the NHS Long Term Plan [4] for self-management and technology-enabled, personalized care demonstrate the potential for digital technology to help people with Parkinson disease and CPs improve their identification and management of Parkinson disease symptoms. Understanding the types of self-management interventions currently being developed and implemented will help inform the development of future, digitally enabled self-management interventions.

Objectives and Research Questions

The aim of this scoping review was to provide an overview of the current state of the field and the evidence of the effectiveness of self-management interventions for Parkinson disease and to identify any gaps. Specifically, the review asked, "What types of self-management interventions are available to support people with Parkinson disease, what outcomes do they target, and what evidence is there in the literature of their effectiveness?"

Methods

Search Strategy

The PRSIMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews; Multimedia Appendix 1 [27]) and Population, Intervention, Comparator, Outcome, and Studies (PICOS) frameworks were used to structure the review and develop the search strategy (see Table 1). Based on the PICOS, relevant Medical Subject Headings (MeSH) terms were identified from a preliminary search, and the search string was created using the following structure: Population (MeSH terms) AND Interventions (MeSH terms) AND Outcomes (MeSH terms). There was no limit on the publication date. The search was performed in 4 databases—PubMed, Ovid, Scopus, and Web of Science—using the University of Plymouth's search tool Primo. PubMed was chosen because it provides a good synthesis of biomedical literature, and the other search engines were selected because they capture a broad, multidisciplinary set of databases to ensure that no relevant literature was missed. Multimedia Appendix 2 provides a complete record of the specific search strings (modified slightly to fit the specific structure and requirements of each database) and the number of references retrieved. The database searches were performed on April 8, 2021.

Table 1. PICOS framework.

PICOS[a]	Detail	MeSH[b] terms used in search
Population	People with Parkinson disease and their carers	Parkinson Disease
Intervention	Self-management interventions for people with Parkinson disease	Self-Management OR Self-Care OR Home Nursing OR Delivery of Health Care, Integrated OR Telemedicine OR Mobile Applications OR Internet-based Interventions OR Internet of Things
Comparator	None or standard care	—[c]
Outcomes	• Primary outcome: self-management (with measures including, but not limited to, health outcomes, behaviors, perceived self-efficacy, quality of life, and use of health care services, etc) • Secondary outcomes: factors that could affect self-management (eg, demographics and disease factors, etc)	Self Efficacy OR Quality of Life OR Signs and Symptoms OR Health Behaviour OR Patient Admission OR Patient Readmission
Study types	Case-control studies, cohort studies, and RCTs[d]	—

[a] PICOS: Population, Intervention, Comparator, Outcome, Study type.

[b] MeSH: Medical Subject Headings.

[c] Not applicable.

[d] RCT: randomized controlled trial.

Inclusion Criteria

Studies were eligible for inclusion in the review if they evaluated a self-management intervention for people with Parkinson disease or their CPs. A broad definition of self-management interventions was used, so that an overview of the different types of intervention could be collected. Any intervention type (remote or in person) was eligible for inclusion if it aimed to help improve any elements of the patient self-management of Parkinson disease. Randomized controlled trials (RCTs), cohort studies, and case-control studies were eligible for inclusion. Studies published at any date were eligible for inclusion.

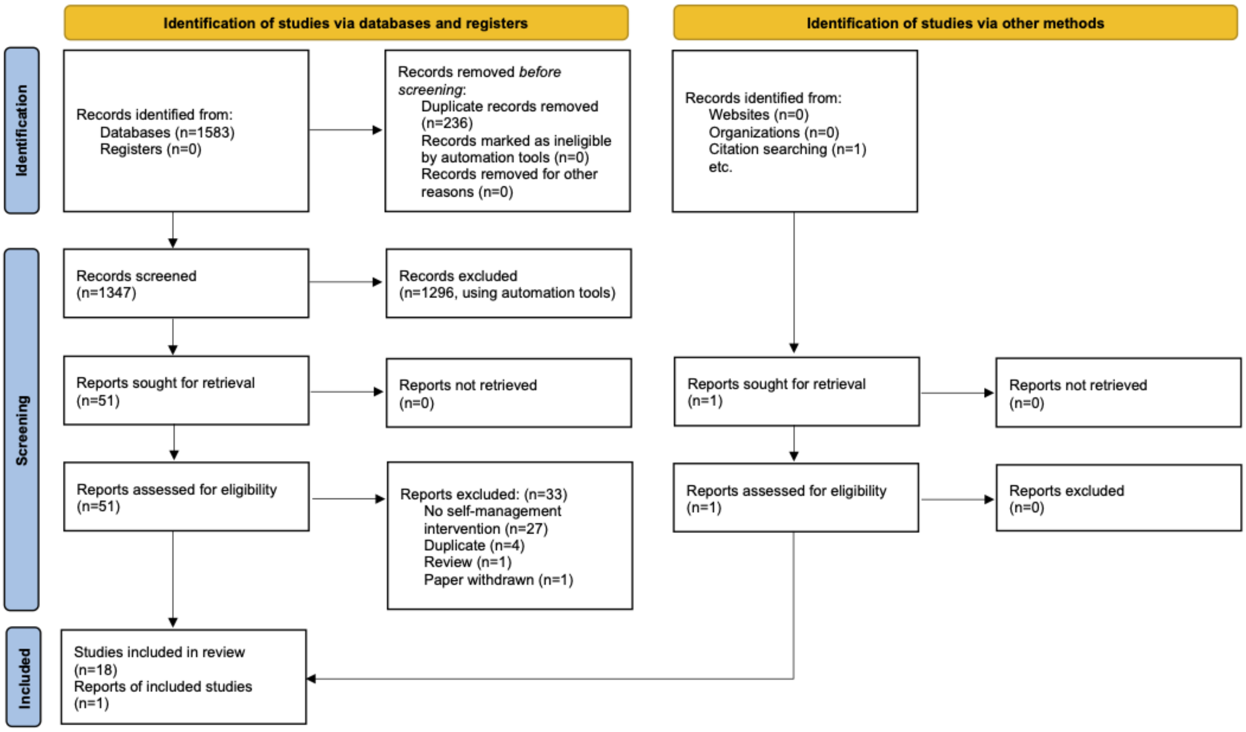
Exclusion Criteria

Studies were excluded if they did not include a self-management intervention for Parkinson disease or if they described an intervention without evaluating it. Protocols and reviews were also excluded. Studies that were published in languages other than English were also excluded, as the review team did not have the necessary resources to assess them.

Screening and Article Selection

The EndNote X9 citation management software (Clarivate) was used to store references, remove duplicates, and conduct the initial screening. The screening was done in several stages, using keywords based on the PICOS (see Multimedia Appendix 3). Next, 1 reviewer screened the remaining titles and abstracts (excluding articles with reasons) and conducted a full-text review to determine final eligibility (see Figure 1). The review team did not have the resources—in terms of time and budget—to have a second reviewer screen and extract data.

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



Data Extraction

The full texts of the included studies were reviewed, and the data were extracted by 1 reviewer based on a predetermined form (see Textbox 1). Given the anticipated variety of study types and aims, the specific outcomes to extract were not prespecified but included as data to be extracted.

Textbox 1. Article information and data extraction.

General study information

- Title
- Year of publication
- Sample size
- Population
- Method

Intervention

- Type of self-management intervention
- Description of self-management intervention

Evaluation

- Primary outcome
- Secondary outcomes
- Summary of reported results
- Evidence of effectiveness at achieving stated outcomes

Data Analysis and Synthesis

A descriptive analysis was used to summarize the data extracted from the studies and provide an overview of the state of the literature on self-management interventions for Parkinson disease. The implications of the findings are examined in the discussion.

Results

Included Studies

The database search retrieved 1583 references (see Multimedia Appendix 2). The EndNote X9 software was used to remove 236 duplicates, and the keyword search tool was used to screen out 1296 references (see Multimedia Appendix 3). The titles and abstracts of 51 studies were screened by 1 reviewer, and articles were excluded with reasons. Of these 51 articles, 19 were selected for the full-text review. Subsequently, 1 study was identified as a duplicate upon full-text review [28], and 1 study was a secondary data analysis of an RCT, so the original RCT was identified and included [29], resulting in a final set of 19 included studies. The reasons for exclusion in the full-text review stage are detailed in Figure 1. The table with the extracted data is included as Multimedia Appendix 4.

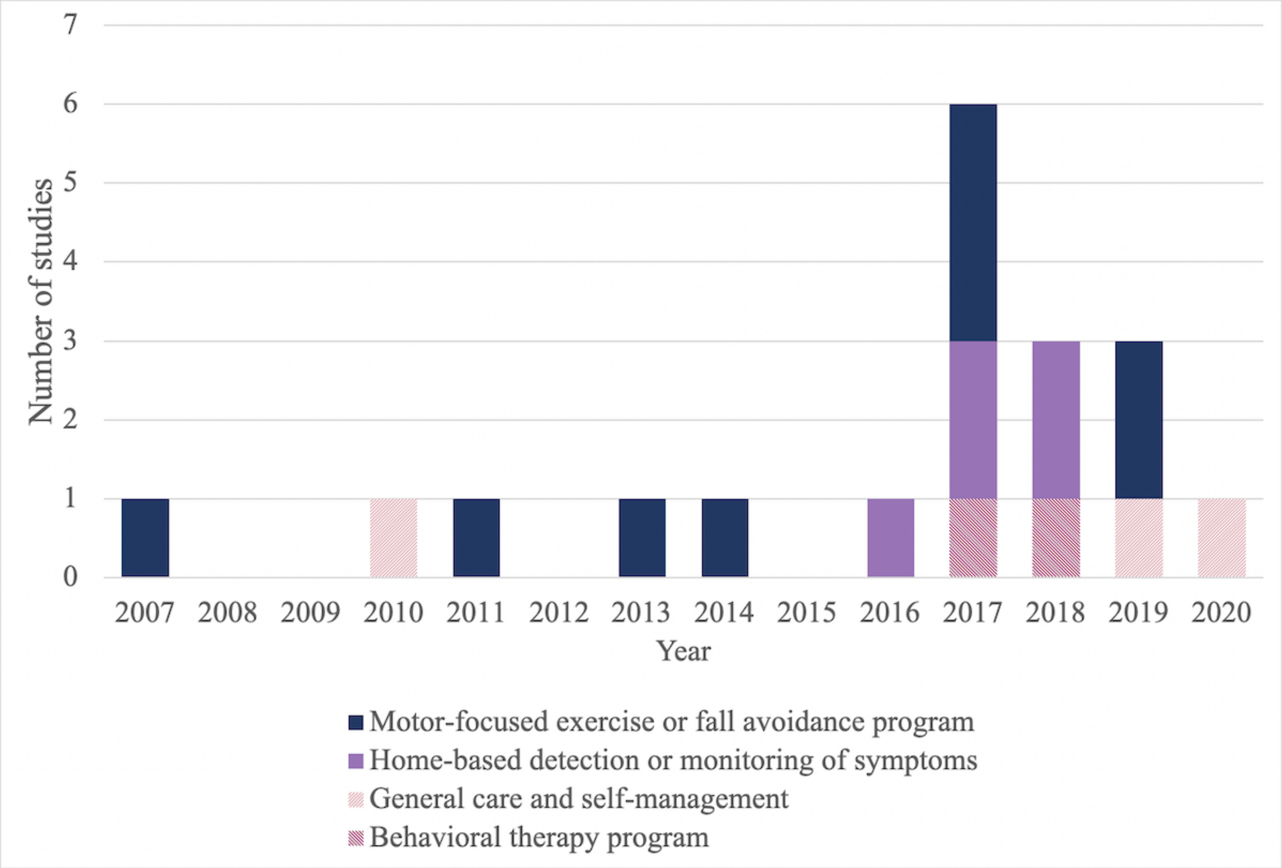
Study Characteristics

The largest proportion (8/19, 42% or 7/18, 39%, as 2 of the studies referred to the same trial) of the studies included in the review used an RCT methodology [29-36]. The remaining studies used a variety of study types, including 3 feasibility studies evaluating algorithms [37-39], 2 one-arm pre-post trials [40,41], 2 pilot studies [42,43], 2 mixed methods acceptability studies [44,45], a randomized case-control study [46], and a secondary data analysis of program adherence [47].

The studies also had a wide range of sample sizes, from 11 participants [42] to 474 participants [31]. All of the 8 studies with the largest sample sizes (greater than 100 participants) were RCTs [29-36].

The earliest study included in the review was published in 2007 [30] and the latest in 2020 [45], but two-thirds (13/19, 68%) of the studies were published in 2017 or later (see Figure 2).

Figure 2. Number of included studies by year and category of Parkinson disease self-management intervention. The intervention types included home-based detection or monitoring of symptoms [37-39,44,46], general care and self-management [29,35,45], behavioral therapy program [40,43], and motor-focused exercise or fall avoidance program [30-34,36,41,42,47]. Note that 2 of the papers included in general care and self-management relate to the same study [29,45].



Types of Interventions

A variety of different types of self-management intervention were described and evaluated by the included studies. The most common (9/19, 47%) type of intervention was home-based exercise or fall prevention programs [30-34,36,41,42,47]; within this category, two-thirds (6/9, 67%) were motor-related exercises or fall avoidance programs, and the remaining interventions were 1 that used sensor-based feedback [42], a community-based exercise program [32], and a handwriting program [33]. Several (5/19, 26%) interventions provided a means of home-based detection or monitoring of symptoms, which were also primarily focused on motor symptoms [37-39,44,46]. Of the remaining studies, 2 delivered behavioral therapy–type interventions to address mental health outcomes [40,43], 1 (addressed in 2 papers) delivered a nurse-led care management program [29,45], and 1 examined a rehabilitation program specifically focusing on self-management skills [35].

Evidence of Effectiveness

Summary

A variety of different outcome measures were used by the studies to evaluate the interventions, given their different aims and types. The strongest evidence of effectiveness came from the 8 RCTs [29-36]. In all, 5 of the 8 RCTs found significant evidence that the intervention was more effective than the control at achieving its respective primary outcome: a 2-minute walk [32], reducing fall rates [30], self-perceived performance of daily activities [34], motor score on the Movement Disorders Society–Unified Parkinson's Disease Rating Scale [36], and health-related quality of life [35]. The other 3 RCTs [29,31,33], a pre-post trial [41], and 1 of the pilot studies [42] found significant differences (between groups or over time) for some, but not all, of the outcomes measured.

Motor-Focused Program Outcomes

Several of the studies examined interventions that included sessions with physiotherapists or occupational therapists combined with independent practice aimed at preventing falls or improving physical activity function [30,31,34,41]. Of the 2 studies that focused specifically on fall prevention, 1 found a significantly reduced rate of falls in the intervention group compared to the control group [30]. The other study (PDSAFE personalized fall prevention program) did not find a significant difference in repeated falling between groups but did observe better balance, functional strength, and fall efficacy and reduced near-falls in the intervention group compared to the control group [31]. The remaining 2 studies focused on exercise; 1 found significant improvements in outcome expectations for exercise and time spent exercising and on the Unified Parkinson’s Disease Rating Scale and the Parkinson’s Disease Questionnaire-39 but not self-efficacy, outcome expectations for functional ability, depression, or timed chair rise scores [41], whereas the other found a significant difference on self-perceived performance in daily activity but not perceived capacity, daily activity performance, effect of fatigue, coping skills, mood, or quality of life measures [34].

There were 2 studies of motor-focused interventions that used digital technology to provide self-management support [36,42]. These studies found that a virtual reality home-trainer stationary cycle resulted in a significant difference between intervention and control groups on the Unified Parkinson’s Disease Rating Scale in favor of the intervention [36] and that a sensor-based auditory feedback device to improve stepping automaticity while dual-tasking had a significant difference for step automaticity but not for fear of falling, cognitive functioning, or self-reported gait freezing [42].

The remaining 2 studies of the motor-focused interventions delivered a community exercise intervention with aerobic and resistance training twice weekly, which found significant effects for 2-minute walk scores and the Unified Parkinson’s Disease Rating Scale over 12 months [32], and a control intervention focused on handwriting, which found some effect on self-reported difficulty [33].

Mental Health Outcomes

There were 2 studies that examined the impact of interventions on mental health outcomes related to Parkinson disease [40,43]. Both trials examined the effect of an intervention over time and found significant improvements with large effect sizes. One of the studies, which was pilot-testing a 10-week cognitive-behavioral telemedicine program (a self-help workbook combined with occasional telephone sessions), found a significant improvement in depression and anxiety over the 4-month study period [43]. The other study, which evaluated a 6-week telephone-based behavioral activation intervention, found a significant, medium-to-large effect size of the intervention on apathy (d=0.77), depression (d=0.70), and quality of life (d=0.50) [40].

Algorithm Evaluations

There were 2 studies that evaluated a classification algorithm by measuring area under the receiving operator curve (AUC) as their primary outcome. The AUC represents how well the model can differentiate between 2 conditions, with a general understanding that scores of 0.7-0.8 are acceptable, 0.8-0.9 are excellent, and 0.9 and higher are outstanding [48]. The first study found AUCs of 0.88 and 0.91 for the best models [37], and the other validated that the model performed similarly on data collected in clinic (AUC 0.83) as on data collected at home (AUC 0.76) [38]. A third study evaluating an algorithm reported that it compared favorably to similar systems and that it could replicate clinical decisions; however, the supervised machine learning process was based on the decisions of 1 neurologist, so the algorithm had learned to replicate those decisions [39]. The authors recognized this as a limitation of the study.

Patient Perceptions, Acceptability, and Usability

Of the 19 studies, 2 used mixed methods to examine the user perceptions of the interventions [44,45]. The first was a companion study to a multisite RCT of a nurse-led care management intervention for Parkinson disease [29]. The study found that after the intervention, people with Parkinson disease rated their medication self-management highly and found the nurse care managers to be helpful, although some usability issues with the program were reported by participants and nurse care managers. Likewise, the nurses found the program to be helpful, the Parkinson disease specialists found the nurse care manager’s role to be helpful, and both reported seeing improvements in the self-management of people with Parkinson disease [45].

The other study assessed the acceptability of a wrist-worn sensor [44]. Participants identified discomfort after long periods of use and problems with the strap; however, there was high compliance with wearing the sensor, and participants reported a preference for the sensor over symptom diaries [44].

Adherence

There was 1 article [47] that reported a secondary data analysis for 1 of the RCTs included in the review [30]. It examined the adherence of the 70 participants in the intervention group of the study to the home-based exercise program. Patients reported completing a high percentage (79%) of the recommended number of repetitions of their exercises. Adherence varied depending on participant characteristics; specifically, older age, worse physical condition, pain, anxiety, and depression were all associated with reduced adherence to the prescribed exercises [47].

Cost-effectiveness

Only 1 study focused on assessing cost-effectiveness [46]. Cubo et al [46] conducted a randomized, case-control study comparing home-based motor monitoring (using wireless motion sensors) with in-office monitoring. They reported that the home-based monitoring was cost-effective but found no significant differences between the groups for symptoms or quality of life [46].

Discussion

Summary of Findings

The studies included in the review varied widely in terms of the study and intervention types, the number of participants, and the outcomes assessed. Approximately two-thirds (13/19) of the studies included examined interventions that focused primarily on motor-related outcomes. Almost 40% (7/18) of the studies included were RCTs; these trials had the largest sample sizes (ranging from 105 to 474 participants) compared to the other study types (ranging from 11 to 82 participants).

The RCTs also had the strongest evidence of effectiveness for their interventions, with almost two-thirds (5/8) finding significant evidence of effectiveness for their primary outcome compared to the control group, and the remaining 3 studies finding significant evidence for some of their outcome variables. However, the non-RCT studies all had at least some evidence to support their intervention, including evidence of an effect of the intervention over time, good model fit, adherence, generally positive acceptability and user perceptions, or cost-effectiveness. However, several of these studies reported limitations in their design and emphasized the need for further investigations to address unanswered questions.

Limitations

A limitation of this scoping review is that only 1 researcher performed the article selection, data extraction, and data analysis. The PRISMA-ScR framework was used to guide the review [27] and ensure that the requirements for a scoping review were reported, but we could not prevent any potential bias due to the lack of validation from a second, independent reviewer.

Another potential source of selection bias is that no manual searches were conducted in the references of any of the included articles or reviews retrieved in the initial search. Due to time constraints, this search was not feasible but increases the possibility that relevant studies may not have been included in the review.

Meaning and Future Research

The volume of studies retrieved during the initial search and the variety of intervention types included in the review demonstrate the breadth of research on technological support and home-based care for Parkinson disease. The research into supporting self-management for people with Parkinson disease addressed several different aspects of management: home-based symptom monitoring that aimed to improve data collection and better inform health care professionals' care decisions; behavioral therapy that aimed to improve mental health; and independent, supervised, or community programs that aimed to increase mobility and strength, reduce the risk of falls, and improve the quality of life.

In addition to the variety of the research, this review identified some trends in the interventions being developed and evaluated to support self-management in people with Parkinson disease. The most prominent trend was the focus of the interventions on motor-related monitoring and care. This was an interesting observation, because the nonmotor symptoms of Parkinson disease can have a substantial impact on disease burden and the quality of life [49-52]. Only 2 of the reviewed interventions focused primarily on the mental health aspects of Parkinson disease [40,43]. Although some (6/17) of the other studies did include an assessment of at least one nonmotor symptom as a secondary outcome (most frequently depression) [29,31,34,36,41,46], there was a surprising lack of interventions that aimed to improve the self-management of nonmotor symptoms. Given the impact of nonmotor symptoms on people with Parkinson disease, this is an important gap that should be further investigated and addressed.

Another key area for future research would be the cost-effectiveness of self-management interventions. Only 1 of the included studies examined cost-effectiveness, which also identified a lack of the resources needed to conduct high-quality, cost-effectiveness evaluations (eg, the lack of a specific "cost-of-illness" questionnaire for people with Parkinson disease) [46]. This will be an important area to explore, as several of the interventions appeared to be resource-intensive, especially the interventions that involved home visits by therapists. Although there is likely to be an offset of costs if these interventions improve the quality of life, slow the deterioration of health, and reduce the need for expensive treatments, these interventions will need to be rigorously evaluated to demonstrate the potential benefit of their widespread adoption.

Conclusion

This scoping review aimed to examine and provide an overview of the state of the literature on self-management interventions for people with Parkinson disease. There is a large amount of research in this area, including several RCTs, that focus on a variety of types of self-management intervention. Most of the studies reported at least some evidence of effectiveness or positive effect of the intervention examined, with the best evidence of effectiveness coming from the RCTs. However, the majority of the studies reviewed focused on motor-related interventions and outcomes, with few interventions aimed at addressing the nonmotor aspects of Parkinson disease. There was also an apparent lack of consideration of the cost-effectiveness of the interventions. Further research will be needed to compare the potential health and economic benefits of implementing interventions to support self-management in people with Parkinson disease with the costs of delivering the interventions. Although some of the studies examined interventions that used digital technologies to monitor symptoms or provide feedback, many of the interventions had substantial in-person time commitments. Future investigations could compare the effectiveness of delivering interventions in person or through digital technologies to potentially improve the cost-effectiveness and availability of self-management support.

Acknowledgments

This research was supported by funding from Parkinson’s UK Non-Drug Approaches Award (grant H-2101) and the Centre for Health Technology at the University of Plymouth. The views expressed in the paper belong to the authors and do not necessarily reflect those of Parkinson’s UK or the University of Plymouth. The funding bodies were not involved in the study design, data collection or analysis, or the writing and decision to submit the article for publication.

Authors' Contributions

MMI and EM conceived the key research questions and scope. The scoping review was executed and drafted by MMI with revisions from CC and EM.

Conflicts of Interest

EM is the director of CM Digital Health Solutions Ltd (13570320). CC is the inventor of NMS Assist (a digital system to help people with Parkinson disease monitor and manage their nonmotor symptoms at home). CC receives salary from University Hospitals Plymouth National Health Service Trust and National Institute of Health and Care Research and has received advisory, consulting, or lecture fees from AbbVie, Bial, Scient, Orkyn, Abidetex, UCB, Pfizer, EverPharma, Lundbeck, Global Kinetics, Kyowa Kirin, Britannia, and Medscape and research funding from Parkinson’s UK, Edmond J Safra Foundation, National Institute of Health and Care Research, and Cure Parkinson's.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 108 KB - jmir_v24i8e40181_app1.docx]

Multimedia Appendix 2
Search record.
[DOCX File , 13 KB - jmir_v24i8e40181_app2.docx]

Multimedia Appendix 3
EndNote search criteria.
[DOCX File , 13 KB - jmir_v24i8e40181_app3.docx]

Multimedia Appendix 4
Data extraction table.
[XLSX File (Microsoft Excel File), 66 KB - jmir_v24i8e40181_app4.xlsx]

References

1. Yang W, Hamilton JL, Kopil C, Beck JC, Tanner CM, Albin RL, et al. Current and projected future economic burden of Parkinson's disease in the U.S. NPJ Parkinsons Dis 2020;6:15 [FREE Full text] [doi: 10.1038/s41531-020-0117-1] [Medline: 32665974]
2. Weir S, Samnaliev M, Kuo T, Tierney TS, Walleser Autiero S, Taylor RS, et al. Short- and long-term cost and utilization of health care resources in Parkinson's disease in the UK. Mov Disord 2018 Jul;33(6):974-981. [doi: 10.1002/mds.27302] [Medline: 29603405]
3. The incidence and prevalence of Parkinson's in the UK: results from the Clinical Practice Research Datalink Summary report. Parkinson's UK. 2018 Jan. URL: https://www.parkinsons.org.uk/sites/default/files/2018-01/CS2960%20Incidence%20and%20prevalence%20report%20branding%20summary%20report.pdf [accessed 2022-07-27]
4. NHS Long Term Plan. NHS. 2019 Jan. URL: https://www.longtermplan.nhs.uk [accessed 2022-07-27]
5. Panagioti M, Richardson G, Small N, Murray E, Rogers A, Kennedy A, et al. Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. BMC Health Serv Res 2014 Aug 27;14:356 [FREE Full text] [doi: 10.1186/1472-6963-14-356] [Medline: 25164529]
6. Feng S, Mäntymäki M, Dhir A, Salmela H. How self-tracking and the quantified self promote health and well-being: systematic review. J Med Internet Res 2021 Sep 21;23(9):e25171 [FREE Full text] [doi: 10.2196/25171] [Medline: 34546176]
7. Hibbard J, Gilburt H. Supporting people to manage their health: an introduction to patient activation. The King's Fund. 2014 May. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/supporting-people-manage-health-patient-activation-may14.pdf [accessed 2021-04-20]

8. Kessler D, Hauteclocque J, Grimes D, Mestre T, Côtéd D, Liddy C. Development of the Integrated Parkinson's Care Network (IPCN): using co-design to plan collaborative care for people with Parkinson's disease. Qual Life Res 2019 May;28(5):1355-1364. [doi: 10.1007/s11136-018-2092-0] [Medline: 30600492]
9. Chaudhuri KR, Prieto-Jurcynska C, Naidu Y, Mitra T, Frades-Payo B, Tluk S, et al. The nondeclaration of nonmotor symptoms of Parkinson's disease to health care professionals: an international study using the nonmotor symptoms questionnaire. Mov Disord 2010 Apr 30;25(6):704-709. [doi: 10.1002/mds.22868] [Medline: 20437539]
10. Non-motor symptoms of Parkinson's. Parkinson's UK. URL: https://www.parkinsons.org.uk/information-and-support/non-motor-symptoms-parkinsons [accessed 2021-02-16]
11. Marinus J, Zhu K, Marras C, Aarsland D, van Hilten JJ. Risk factors for non-motor symptoms in Parkinson's disease. Lancet Neurol 2018 Jun;17(6):559-568. [doi: 10.1016/S1474-4422(18)30127-3] [Medline: 29699914]
12. Minar M, Dragasek J, Valkovic P. Comorbidities in Parkinson's disease – the results from national epidemiological study cosmos. J Neurol Sci 2019 Oct;405:211. [doi: 10.1016/j.jns.2019.10.1193]
13. Findley LJ. The economic impact of Parkinson's disease. Parkinsonism Relat Disord 2007 Sep;13 Suppl:S8-S12. [doi: 10.1016/j.parkreldis.2007.06.003] [Medline: 17702630]
14. Gumber A, Ramaswamy B, Thongchundee O. Effects of Parkinson's on employment, cost of care, and quality of life of people with condition and family caregivers in the UK: a systematic literature review. Patient Relat Outcome Meas 2019;10:321-333 [FREE Full text] [doi: 10.2147/PROM.S160843] [Medline: 31695537]
15. Helping people share decision making. The Health Foundation. 2012 Jul. URL: https://www.health.org.uk/publications/helping-people-share-decision-making [accessed 2021-11-15]
16. Lin CE, Wood JJ. Self-management interventions. In: Volkmar FR, editor. Encyclopedia of Autism Spectrum Disorders. New York, NY: Springer; 2013:2735-2742.
17. Jonkman NH, Schuurmans MJ, Jaarsma T, Shortridge-Baggett LM, Hoes AW, Trappenburg JCA. Self-management interventions: proposal and validation of a new operational definition. J Clin Epidemiol 2016 Dec;80:34-42. [doi: 10.1016/j.jclinepi.2016.08.001] [Medline: 27531245]
18. Morgan C, Rolinski M, McNaney R, Jones B, Rochester L, Maetzler W, et al. Systematic review looking at the use of technology to measure free-living symptom and activity outcomes in Parkinson's disease in the home or a home-like environment. J Parkinsons Dis 2020;10(2):429-454 [FREE Full text] [doi: 10.3233/JPD-191781] [Medline: 32250314]
19. Channa A, Popescu N, Ciobanu V. Wearable solutions for patients with Parkinson's disease and neurocognitive disorder: a systematic review. Sensors (Basel) 2020 May 09;20(9):2713 [FREE Full text] [doi: 10.3390/s20092713] [Medline: 32397516]
20. Deb R, Bhat G, An S, Shill H, Ogras U. Trends in technology usage for Parkinson’s disease assessment: a systematic review. medRxiv Preprint posted online on February 04, 2021. [doi: 10.1101/2021.02.01.21250939]
21. Vaartio-Rajalin H, Rauhala A, Fagerström L. Person-centered home-based rehabilitation for persons with Parkinson's disease: a scoping review. Int J Nurs Stud 2019 Nov;99:103395. [doi: 10.1016/j.ijnurstu.2019.103395] [Medline: 31525645]
22. Tuijt R, Tan A, Armstrong M, Pigott J, Read J, Davies N, et al. Self-management components as experienced by people with Parkinson's disease and their carers: a systematic review and synthesis of the qualitative literature. Parkinsons Dis 2020;2020:8857385 [FREE Full text] [doi: 10.1155/2020/8857385] [Medline: 33489082]
23. Kessler D, Liddy C. Self-management support programs for persons with Parkinson's disease: an integrative review. Patient Educ Couns 2017 Oct;100(10):1787-1795. [doi: 10.1016/j.pec.2017.04.011] [Medline: 28465112]
24. Kane E, Schrag A, Walters K, Pigott J. Self-management interventions in Parkinson's disease: a systematic review and meta-analysis. PROSPERO. 2019 Apr 15. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019117183 [accessed 2021-04-07]
25. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
26. Wannheden C, Revenäs Å. How people with Parkinson's disease and health care professionals wish to partner in care using eHealth: co-design study. J Med Internet Res 2020 Sep 21;22(9):e19195 [FREE Full text] [doi: 10.2196/19195] [Medline: 32955448]
27. 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]
28. Chivers Seymour K, Pickering R, Rochester L, Roberts HC, Ballinger C, Hulbert S, et al. Multicentre, randomised controlled trial of PDSAFE, a physiotherapist-delivered fall prevention programme for people with Parkinson's. J Neurol Neurosurg Psychiatry 2019 Jul;90(7):774-782 [FREE Full text] [doi: 10.1136/jnnp-2018-319448] [Medline: 30944149]
29. Connor KI, Cheng EM, Barry F, Siebens HC, Lee ML, Ganz DA, et al. Randomized trial of care management to improve Parkinson disease care quality. Neurology 2019 Apr 16;92(16):e1831-e1842 [FREE Full text] [doi: 10.1212/WNL.0000000000007324] [Medline: 30902908]
30. Ashburn A, Fazakarley L, Ballinger C, Pickering R, McLellan LD, Fitton C. A randomised controlled trial of a home based exercise programme to reduce the risk of falling among people with Parkinson's disease. J Neurol Neurosurg Psychiatry 2007 Jul;78(7):678-684 [FREE Full text] [doi: 10.1136/jnnp.2006.099333] [Medline: 17119004]

31. Ashburn A, Pickering R, McIntosh E, Hulbert S, Rochester L, Roberts HC, et al. Exercise- and strategy-based physiotherapy-delivered intervention for preventing repeat falls in people with Parkinson's: the PDSAFE RCT. Health Technol Assess 2019 Jul;23(36):1-150 [FREE Full text] [doi: 10.3310/hta23360] [Medline: 31339100]
32. Collett J, Franssen M, Meaney A, Wade D, Izadi H, Tims M, et al. Phase II randomised controlled trial of a 6-month self-managed community exercise programme for people with Parkinson's disease. J Neurol Neurosurg Psychiatry 2017 Mar;88(3):204-211. [doi: 10.1136/jnnp-2016-314508] [Medline: 27837101]
33. Collett J, Franssen M, Winward C, Izadi H, Meaney A, Mahmoud W, et al. A long-term self-managed handwriting intervention for people with Parkinson's disease: results from the control group of a phase II randomized controlled trial. Clin Rehabil 2017 Dec;31(12):1636-1645. [doi: 10.1177/0269215517711232] [Medline: 28547999]
34. Sturkenboom IHWM, Graff MJL, Hendriks JCM, Veenhuizen Y, Munneke M, Bloem BR, OTiP study group. Efficacy of occupational therapy for patients with Parkinson's disease: a randomised controlled trial. Lancet Neurol 2014 Jun;13(6):557-566. [doi: 10.1016/S1474-4422(14)70055-9] [Medline: 24726066]
35. Tickle-Degnen L, Ellis T, Saint-Hilaire MH, Thomas CA, Wagenaar RC. Self-management rehabilitation and health-related quality of life in Parkinson's disease: a randomized controlled trial. Mov Disord 2010 Jan 30;25(2):194-204 [FREE Full text] [doi: 10.1002/mds.22940] [Medline: 20077478]
36. van der Kolk NM, de Vries NM, Kessels RPC, Joosten H, Zwinderman AH, Post B, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. Lancet Neurol 2019 Nov;18(11):998-1008. [doi: 10.1016/S1474-4422(19)30285-6] [Medline: 31521532]
37. Arroyo-Gallego T, Ledesma-Carbayo MJ, Sanchez-Ferro A, Butterworth I, Mendoza CS, Matarazzo M, et al. Detection of motor impairment in Parkinson's disease via mobile touchscreen typing. IEEE Trans Biomed Eng 2017 Sep;64(9):1994-2002. [doi: 10.1109/tbme.2017.2664802]
38. Arroyo-Gallego T, Ledesma-Carbayo MJ, Butterworth I, Matarazzo M, Montero-Escribano P, Puertas-Martín V, et al. Detecting motor impairment in early Parkinson's disease via natural typing interaction with keyboards: validation of the neuroQWERTY approach in an uncontrolled at-home setting. J Med Internet Res 2018 Mar 26;20(3):e89 [FREE Full text] [doi: 10.2196/jmir.9462] [Medline: 29581092]
39. Ferraris C, Nerino R, Chimienti A, Pettiti G, Cau N, Cimolin V, et al. A self-managed system for automated assessment of UPDRS upper limb tasks in Parkinson's disease. Sensors (Basel) 2018 Oct 18;18(10):3523 [FREE Full text] [doi: 10.3390/s18103523] [Medline: 30340420]
40. Butterfield LC, Cimino CR, Salazar R, Sanchez-Ramos J, Bowers D, Okun MS. The Parkinson's Active Living (PAL) program. J Geriatr Psychiatry Neurol 2017 Jan;30(1):11-25. [doi: 10.1177/0891988716673467] [Medline: 28248557]
41. Pretzer-Aboff I, Galik E, Resnick B. Feasibility and impact of a function focused care intervention for Parkinson's disease in the community. Nurs Res 2011;60(4):276-283. [doi: 10.1097/NNR.0b013e318221bb0f] [Medline: 21691241]
42. Chomiak T, Watts A, Meyer N, Pereira FV, Hu B. A training approach to improve stepping automaticity while dual-tasking in Parkinson's disease: a prospective pilot study. Medicine (Baltimore) 2017 Feb;96(5):e5934 [FREE Full text] [doi: 10.1097/MD.0000000000005934] [Medline: 28151878]
43. Dobkin RD, Interian A, Durland JL, Gara MA, Menza MA. Personalized telemedicine for depression in Parkinson's disease: a pilot trial. J Geriatr Psychiatry Neurol 2018 Jul;31(4):171-176. [doi: 10.1177/0891988718783274] [Medline: 29945467]
44. Fisher JM, Hammerla NY, Rochester L, Andras P, Walker RW. Body-worn sensors in Parkinson's disease: evaluating their acceptability to patients. Telemed J E Health 2016 Jan;22(1):63-69 [FREE Full text] [doi: 10.1089/tmj.2015.0026] [Medline: 26186307]
45. Connor KI, Siebens HC, Mittman BS, McNeese-Smith DK, Ganz DA, Barry F, et al. Stakeholder perceptions of components of a Parkinson disease care management intervention, care coordination for health promotion and activities in Parkinson's disease (CHAPS). BMC Neurol 2020 Dec 02;20(1):437 [FREE Full text] [doi: 10.1186/s12883-020-02011-9] [Medline: 33267827]
46. Cubo E, Mariscal N, Solano B, Becerra V, Armesto D, Calvo S, et al. Prospective study on cost-effectiveness of home-based motor assessment in Parkinson's disease. J Telemed Telecare 2017 Feb;23(2):328-338. [doi: 10.1177/1357633X16638971] [Medline: 27000142]
47. Pickering RM, Fitton C, Ballinger C, Fazakarley L, Ashburn A. Self reported adherence to a home-based exercise programme among people with Parkinson's disease. Parkinsonism Relat Disord 2013 Jan;19(1):66-71. [doi: 10.1016/j.parkreldis.2012.07.006] [Medline: 22878222]
48. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010 Sep;5(9):1315-1316 [FREE Full text] [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]
49. Todorova A, Jenner P, Ray Chaudhuri K. Non-motor Parkinson's: integral to motor Parkinson's, yet often neglected. Pract Neurol 2014 Oct;14(5):310-322 [FREE Full text] [doi: 10.1136/practneurol-2013-000741] [Medline: 24699931]
50. Hermanowicz N, Jones SA, Hauser RA. Impact of non-motor symptoms in Parkinson's disease: a PMDAlliance survey. Neuropsychiatr Dis Treat 2019;15:2205-2212 [FREE Full text] [doi: 10.2147/NDT.S213917] [Medline: 31496703]
51. Gökçal E, Gür VE, Selvitop R, Babacan Yildiz G, Asil T. Motor and non-motor symptoms in Parkinson's disease: effects on quality of life. Noro Psikiyatr Ars 2017 Jun;54(2):143-148 [FREE Full text] [doi: 10.5152/npa.2016.12758] [Medline: 28680312]

52. Pfeiffer RF. Non-motor symptoms in Parkinson's disease. Parkinsonism Relat Disord 2016 Jan;22 Suppl 1:S119-S122. [doi: 10.1016/j.parkreldis.2015.09.004] [Medline: 26372623]

Abbreviations

AUC: area under the receiving operator curve
CP: care partner
MeSH: Medical Subject Headings
NHS: National Health Service
PICOS: Population, Intervention, Comparator, Outcome, and Study
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial

Edited by R Kukafka; submitted 09.06.22; peer-reviewed by , R Damaševičius, PV Eswaradass, AA Seid; comments to author 02.07.22; revised version received 07.07.22; accepted 20.07.22; published 05.08.22.

Please cite as:
Milne-Ives M, Carroll C, Meinert E
Self-management Interventions for People With Parkinson Disease: Scoping Review
J Med Internet Res 2022;24(8):e40181
URL: https://www.jmir.org/2022/8/e40181
doi:10.2196/40181
PMID:35930315

Review

Technology-Delivered Adaptations of Motivational Interviewing for the Prevention and Management of Chronic Diseases: Scoping Review

Havisha Pedamallu[1], BA; Matthew J Ehrhardt[2], MD, MS; Julia Maki[1], PhD; April Idalski Carcone[3], MSW, PhD; Melissa M Hudson[2], MD; Erika A Waters[4], MPH, PhD

[1]Division of Public Health Sciences, Department of Surgery, Washington University in St Louis, St Louis, MO, United States
[2]Department of Oncology, Department of Epidemiology and Cancer Control, St. Jude Children’s Research Hospital, Memphis, TN, United States
[3]Department of Family Medicine and Public Health Sciences, Wayne State University School of Medicine, Detroit, MI, United States
[4]Division of Public Health Sciences, Department of Surgery, Washington University in St Louis, School of Medicine, St Louis, MO, United States

Corresponding Author:
Erika A Waters, MPH, PhD
Division of Public Health Sciences
Department of Surgery
Washington University in St Louis, School of Medicine
600 S Euclid Ave
Campus Box 8100
St Louis, MO, 63110
United States
Phone: 1 314 747 5705
Email: waterse@wustl.edu

Abstract

Background: Motivational interviewing (MI) can increase health-promoting behaviors and decrease health-damaging behaviors. However, MI is often resource intensive, precluding its use with people with limited financial or time resources. Mobile health–based versions of MI interventions or technology-delivered adaptations of MI (TAMIs) might increase reach.

Objective: We aimed to understand the characteristics of existing TAMIs. We were particularly interested in the inclusion of people from marginalized sociodemographic groups, whether the TAMI addressed sociocontextual factors, and how behavioral and health outcomes were reported.

Methods: We employed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for scoping reviews to conduct our scoping review. We searched PubMed, CINAHL, and PsycInfo from January 1, 1996, to April 6, 2022, to identify studies that described interventions incorporating MI into a mobile or electronic health platform. For inclusion, the study was required to (1) describe methods/outcomes of an MI intervention, (2) feature an intervention delivered automatically via a mobile or electronic health platform, and (3) report a behavioral or health outcome. The exclusion criteria were (1) publication in a language other than English and (2) description of only in-person intervention delivery (ie, no TAMI). We charted results using Excel (Microsoft Corp).

Results: Thirty-four studies reported the use of TAMIs. Sample sizes ranged from 10 to 2069 participants aged 13 to 70 years. Most studies (n=27) directed interventions toward individuals engaging in behaviors that increased chronic disease risk. Most studies (n=22) oversampled individuals from marginalized sociodemographic groups, but few (n=3) were designed specifically with marginalized groups in mind. TAMIs used text messaging (n=8), web-based intervention (n=22), app + text messaging (n=1), and web-based intervention + text messaging (n=3) as delivery platforms. Of the 34 studies, 30 (88%) were randomized controlled trials reporting behavioral and health-related outcomes, 23 of which reported statistically significant improvements in targeted behaviors with TAMI use. TAMIs improved targeted health behaviors in the remaining 4 studies. Moreover, 11 (32%) studies assessed TAMI feasibility, acceptability, or satisfaction, and all rated TAMIs highly in this regard. Among 20 studies with a disproportionately high number of people from marginalized racial or ethnic groups compared with the general US population, 16 (80%) reported increased engagement in health behaviors or better health outcomes. However, no TAMIs included elements that addressed sociocontextual influences on behavior or health outcomes.

Conclusions: Our findings suggest that TAMIs may improve some health promotion and disease management behaviors. However, few TAMIs were designed specifically for people from marginalized sociodemographic groups, and none included elements to help address sociocontextual challenges. Research is needed to determine how TAMIs affect individual health outcomes and how to incorporate elements that address sociocontextual factors, and to identify the best practices for implementing TAMIs into clinical practice.

(J Med Internet Res 2022;24(8):e35283) doi:10.2196/35283

KEYWORDS

motivational interviewing; technology; telehealth; health behavior; chronic disease; socioeconomic factors; health promotion; disease management; primary prevention; secondary prevention; minority health

Introduction

Background

Chronic diseases, such as heart disease, cancer, and diabetes, are the leading causes of death in the United States, affecting 6 in 10 adults [1]. The risk of developing many chronic diseases and their corresponding complications, if diagnosed, can be reduced by avoiding the following 4 key health behaviors: tobacco use, poor nutrition, physical inactivity, and excessive alcohol use [1]. However, reducing these harmful behaviors can be exceptionally difficult. As a result, a multitude of interventions have been developed that attempt to facilitate health behavior change in these domains [2].

Using eHealth Technologies to Improve Health

The use of computing and internet technologies generally (eHealth), and smartphone and texting technologies specifically (mobile health [mHealth]), in health behavior change interventions has greatly increased over the past 3 decades [3]. These technologies can be effective tools for delivering health behavior interventions to diverse populations for a variety of behavior change goals. To illustrate, a systematic review of mobile apps and text messaging interventions demonstrated improvement across a variety of physical and mental health outcomes, including weight loss, smoking cessation, medication adherence, and depression and anxiety symptoms [4]. Similar findings were reported in a systematic review of text messaging health promotion interventions [5]. mHealth interventions may also be useful for facilitating self-management in patients with chronic diseases, such as improving medication adherence and control of chronic disease indicators like BMI, and activating and empowering patients [4-7].

mHealth technologies in particular have the potential to greatly improve health care access, improve delivery processes, and reduce chronic care costs [6], especially in areas that are medically underserved and underresourced, and in areas where internet access is limited to personal mobile phones [8]. mHealth technologies are widely accessible, popular across sociodemographic groups, and portable, and can facilitate timely interventions for patients [8,9]. These features are especially important for encouraging chronic disease prevention and management behaviors, which require timely and frequent reminders and interactions with patients that are impractical for health care practitioners to provide in office settings.

The accessibility of mHealth technologies may be especially important for people from sociodemographic groups that have been underserved, mistreated, or marginalized by biomedical research and practice (hereafter *marginalized*). For example, patients who live in rural areas face long travel and wait times, which may limit how often and for how long they can meet with their providers [10]. People from marginalized racial or ethnic groups face additional barriers, including medical distrust, which stems in large part from experiencing stigma, discrimination, and racism from health care systems and providers [11].

The effectiveness of eHealth and mHealth technologies for health promotion and management behaviors, combined with their availability and accessibility to the public, suggests that these platforms are effective tools for increasing the reach of interventions to greater numbers of individuals with sociocontextual challenges. However, the potential benefits of technology-delivered interventions may be offset by the fact that standard eHealth and mHealth technologies might not motivate patients in a way that fosters autonomy, which is critical to maintaining behavior change over time [12].

Motivational Interviewing and Its Use in eHealth Interventions

Motivational interviewing (MI) is a method of talking to people about changing their behavior [13]. The goal of MI is to increase intrinsic motivation and self-efficacy for engaging in health promoting behaviors using patient-centered yet directive communication techniques [14]. Specifically, MI counselors rely on reflexive listening, strategic questions, affirmations of character strengths, and statements emphasizing patients' decision-making autonomy to elicit "change talk" [15]. Change talk involves statements expressing patients' own desires, abilities, reasons, needs, and commitments to change their behavior while embodying "MI spirit," an empathetic, collaborative, and nonjudgmental demeanor. There is strong evidence supporting MI as a strategy to address barriers to effective health behavior counseling [16]. While health care professionals may attempt to persuade patients to adhere to recommended health behaviors, MI encourages personal decision-making about change and provides guidance and support about potential mechanisms of change [17,18]. During an MI session, a counselor uses the principles of the self-actualization theory and free choice to help individuals identify and work toward their goals[19,20]. MI can include health education and address sociocontextual factors that constrain an individual's choices. MI has been effective in promoting health behavior change for individuals with and without chronic diseases [21,22]. However, while MI offers many benefits to patients, its reach may be limited as employing trained counselors is expensive and training other health care providers may be time consuming and resource intensive.

Technology-delivered adaptations of MI (TAMIs) have been developed to combine the useful features of MI interventions (eg, promoting patient autonomy) with the benefits of mHealth interventions (eg, increase accessibility while limiting costs to patients and the health care system). Despite the complexity of developing TAMIs for intervention studies seeking to improve health behaviors, a 2016 systematic review reported that they are feasible to implement and well accepted by patients [23].

Objective and Research Questions

The objective of this research was to gain an understanding of the characteristics and outcomes of TAMIs that were not addressed in a previous review [23]. Specifically, we asked the following questions: (1) To what extent do TAMIs include individuals from marginalized sociodemographic groups? (2) How do TAMIs address sociocontextual influences on health (eg, the built environment and financial barriers)? (3) How do studies that report TAMIs describe their effects on behavioral and health outcomes?

This work is important for several reasons. First, the answers to the first 2 research questions (ie, including individuals from marginalized sociodemographic groups and addressing sociocontextual influences on health) have not been previously addressed, yet they are critical for expanding access of these potentially useful interventions to individuals and groups who have been socially marginalized or systematically and intentionally excluded from or underrepresented in biomedical research. Second, there have been many technological advancements made over the past 6 years that might shape the nature and effectiveness of TAMIs (eg, increased sophistication of tailoring interactions to participant responses) [24]. Although the previous review evaluated the effects of TAMIs on behavioral and psychological outcomes [23], examination of more modern interventions is informative considering rapidly emerging novel technologies.

We sought to map key concepts and knowledge gaps about TAMIs, including identifying the number of studies that include participants from underrepresented populations and the number of studies that address sociocontextual factors. Because these goals are more consistent with the goals of a scoping review than a systematic review [25], we conducted a scoping review.

Methods

We employed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for scoping reviews [26,27]. We did not preregister the review protocol.

JM and HP searched the PubMed, CINAHL, and PsycInfo databases for articles published from January 1, 1996, through April 6, 2022, that met the following inclusion criteria: (1) the publication described conducting an MI intervention, (2) the intervention was incorporated into an automated mobile or electronic health platform, and (3) the article reported behavioral or health outcomes.

The exclusion criteria were as follows: (1) the article was published in a language other than English and (2) the MI intervention was administered only in-person (ie, no mHealth or eHealth element).

We used the following search string to identify potentially eligible articles: ((“mHealth” OR (“m-Health”)) OR ((“text message”) OR (“text-message”) OR (“text messaging”) OR (“text-messaging”) OR (“ehealth”) OR (“e-health”) OR (“web-based”) OR (“electronic health”) OR (“technology-based”)) AND ((“motivational interviewing”) OR (“motivational interview”) OR (“intrinsic motivation”)) AND (“intervention”).

We reviewed the reference lists of identified articles and authors’ files for additional studies missed by the search criteria. HP reviewed abstracts and full-length articles. HP last searched the literature for articles to include in this review on April 6, 2022.

HP used Microsoft Excel (Microsoft Corp) to abstract the following data for each study: type of data collected (quantitative, qualitative, or mixed), study design, theoretical or conceptual model, study population, intervention target, eligibility criteria, intervention description, mHealth tool details, measures, sample size, sample characteristics, results, conclusions, and key limitations. Not all articles reported all data elements; we noted such instances as “NR” (not reported). The rows in the spreadsheet were individual articles, and the columns were individual data elements.

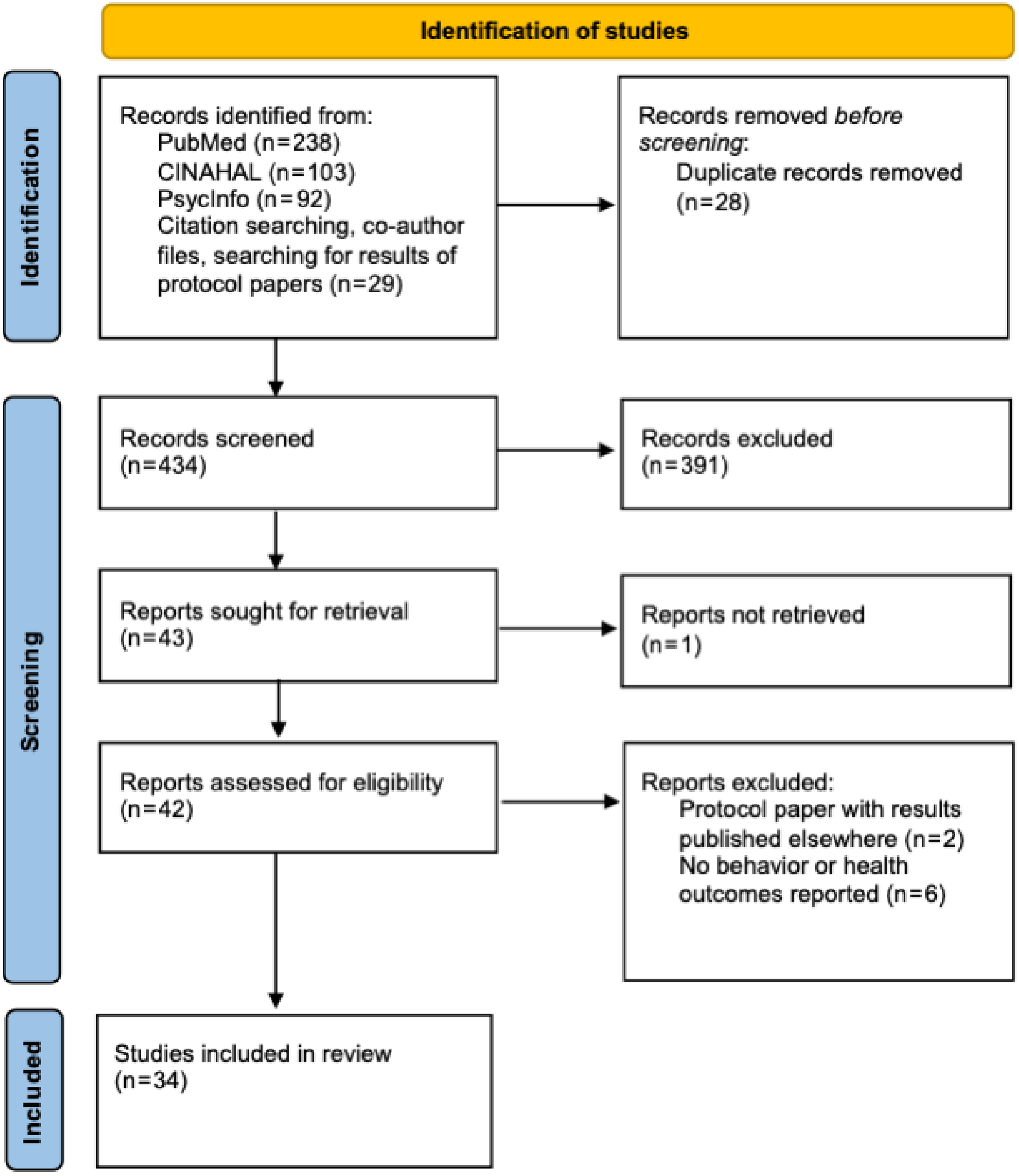
We have summarized the data using tables (Multimedia Appendix 1 and Multimedia Appendix 2) that are separated by the age category of participants (ie, younger than 18 years of age vs 18 years or older) and by the type of outcome (ie, behavioral vs health outcome). We have also described in the text the number and percentage of articles that had different methodological characteristics.

Results

Overview

We identified 34 studies reporting unique TAMIs (Figure 1). Most were conducted in the United States (n=22) [28-49]. Others were conducted in the Netherlands (n=4) [50-53], Switzerland (n=2) [54,55], Germany (n=1) [56], Korea (n=1) [57], Austria (n=1) [58], Canada (n=1) [59], and New Zealand (n=1) [60]. One study included participants from Germany, Sweden, Belgium, and the Czech Republic [61]. Detailed information about each study’s design, population, and outcomes is provided in Multimedia Appendix 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [27] for the scoping review.



General Study Information

Conceptual Framework

Studies were based on a variety of conceptual frameworks. Of the 34 studies, 24 used only MI [28-32,34,35,37-39,41-44,46,47,49,51,53,54,56,59-61] and the remaining 10 used alternative frameworks in conjunction with MI [33,36,40,45,48,50,52,55,57,58] (Multimedia Appendix 1). The most commonly used conceptual framework other than MI was cognitive behavioral therapy (CBT; n=3) [45,50,55].

Study Design

The studies used a variety of designs. There were 30 randomized controlled trials (RCTs), 3 studies that used a nonrandomized pretest-posttest design with no control group [33,36,59], and 1 randomized experiment whose purpose was to refine the contents of the TAMI [58] (Multimedia Appendix 1). Control groups included active control (n=14) [29,31,34,35,37,39-41,43,45,47,53,54,57], treatment as usual (n=7) [28,30,42,44,48,49,60], wait list (n=5) [32,50,52,55,56], or no treatment (n=4) [38,46,51,61].

Study Population

The study populations among the selected studies varied widely (Multimedia Appendix 1). Studies ranged in size from 10 to 2069 participants (mean 341, median 136). Six studies included participants between 13 and 18 years of age. One study recruited from the general population [51], and 2 specifically recruited individuals with mental health diagnoses [31,32]. The 31 remaining studies focused on individuals who were engaging in behaviors that increased their risk of either developing new or exacerbating existing chronic health conditions.

Purpose

The purpose of the health behavior interventions fell into 2 broad categories (Multimedia Appendix 1). Twenty-seven interventions were designed to prevent disease among individuals in the general population or among those engaging in behaviors that put them at a higher risk of being diagnosed with a chronic disease [29-31,33,34,36,38-40,42-48,50-57,59-61]. The remaining 7 interventions were for encouraging chronic disease management [28,32,35,37,41,49,58].

Behaviors and Health Outcomes

Targeted health behaviors included the following: substance, alcohol, or tobacco use (n=24) [29-32,34,36,38-40,42-44,46-48,50,53-57,59-61]; diet or physical activity (n=2) [51,52]; treatment and medication adherence (n=4) [28,35,37,41]; mental health (n=1) [45]; and risky sexual behaviors (n=2) [33,47] (Multimedia Appendix 1). One study addressed both substance, alcohol, or tobacco use and risky sexual behaviors [47]. The most common chronic disease targeted for self-management interventions was diabetes (n=2) [28,35]. Targeted health outcomes included control of diabetes (n=2) [28,35], reduction of HIV viral load (n=1) [41], reduction in depressive symptoms (n=1) [45], reduction in BMI (n=1) [49], improvement in asthma symptoms (n=1) [37], and reduction in anorexia symptoms (n=1) [58].

Delivery

Most studies relied on text messaging or web-based tools to deliver their intervention (Multimedia Appendix 1). Eight used automated text messaging [28,33,38-40,46,49,60], 22 used a web-based intervention [30-32,34,35,41-44,47,48,50-59,61], 1 used an app combined with automated text messaging [36], and 3 used a web-based intervention and automated text messaging [29,37,45]. None used human-generated or chat bot–generated text messaging. Of the 25 studies that used a web-based intervention or a web-based intervention + text messaging, 8 used the Computerized Intervention Authorizing Software (CIAS; [62]) [29,30,34,35,37,41,42,47]. Studies with long-term engagement often used text message–based interventions.

CIAS is a mobile or web-based platform designed to create and launch behavioral health interventions, including those based on MI. CIAS includes an avatar that is capable of over 50 animated expressions and guides participants through questions, allowing the TAMI to more accurately mimic one-on-one conversations [29]. All 8 studies that used CIAS and 3 other studies included an avatar that interacted with participants and delivered the intervention [29,30,34,35,37,41-43,47,51,53]. In 1 study that included an avatar but did not use CIAS [51], participants could select either a male or female avatar whose appearance was designed based on focus group interviews asking the target population about how they believed a motivating and reliable avatar would look. The avatar communicated with speech movements and text displayed in balloons, and featured a limited number of nonverbal expressions. Although the avatar + intervention group did not show a significant improvement in the target outcome compared to the content-identical intervention without the avatar, the authors noted that the avatar had limited relational skills that may have precluded the development of a strong relationship with the user.

Inclusion of Individuals From Marginalized Populations

There was considerable variation in the extent to which studies included members of populations that have been marginalized (Multimedia Appendix 1). Among the 34 studies, 2 interventions were specifically designed for use by African American young people [35,37]. Moreover, 22 studies had samples that included an overrepresentation of people from marginalized sociodemographic groups in the general US population. Specifically, 14 studies recruited ≥22% African American participants [29,31,32,35,37-39,41-43,46-49], whereas African Americans represent 13.4% of the US population [63]. Five studies included ≥30% Hispanic/Latino participants [28,29,33,44,47], whereas 18.5% of the US population identifies as Hispanic/Latino [63]. One study reported that 50% of its participants identified with racial or ethnic groups other than white [34]. Three studies reported at least 67% of participants as having limited incomes [42,43,45], although TAMIs were not specifically designed for use by these populations. The remaining 14 studies had samples that either included fewer people from marginalized racial or ethnic groups than was representative in the US population or did not provide participants' racial and ethnic information.

Only 3 studies indicated an intent to develop content and interventions optimized for use by people from marginalized populations (Multimedia Appendix 1). Two studies described developing interventions with input from African American youth or adolescents [35,37]. However, neither of these studies commented on how the content of their interventions was adapted to fit the needs of African Americans. One study was designed to be culturally relevant, appropriate, accessible, and engaging (ie, for Māori people, New Zealand's indigenous population) [60].

Addressing Sociocontextual Influences on Health

No studies reported that their intervention was designed to accommodate challenging sociocontextual factors, such has having too little income to be able to afford products or services designed to improve health (eg, gym membership, fresh fruits, and vegetables).

Description of How TAMIs Affect Behavioral and Health Outcomes

Of the 34 studies, 30 were RCTs and reported outcomes, including alcohol consumption, tobacco use, hemoglobin A_{1c} levels, and human immunodeficiency virus viral load (see Multimedia Appendix 1 for a complete list). Of these 30 studies, 23 reported that the TAMI resulted in statistically significant improvements in the target behavioral and health outcomes [29,32,34,35,37-42,44-48,50,51,53,55-57,60,61] and the remaining 7 stated that the TAMI had no significant effect on behavior compared to the control group [28,30,31,43,49,52,54]. No RCTs reported that the TAMI resulted in worse outcomes compared to controls. The last 4 (out of 34) studies reporting behavioral and health outcomes used either a pretest-posttest design with no control group or a 1 group posttest only design [33,36,58,59]. All reported that the TAMI had a beneficial impact on promoting the desired health behavior. Eleven of the 34 studies examined feasibility, acceptability, or satisfaction, in addition to the behavioral or health outcomes [29,33,34,36,37,41,44,45,47,49,59]. All reported high participant ratings for these outcomes.

There were inconsistent reports about the value of adding CBT or other therapeutic components to an MI intervention. Some studies that reported combining MI and social determination theory (SDT) or MI and CBT did not achieve statistically significant increases in health behavior engagement compared to traditional web-based interventions [50,52,54]. One study reported that MI + CBT had a significant beneficial effect on alcohol consumption after 6 months [50]. Another study found that a web-based MI intervention with and without a trained psychologist coach was equally effective in decreasing the weekly standard units of alcohol at 6 weeks and 6 months compared to a wait-list control [56]. Two studies included in-person MI and reported a decrease in the number of drinking days in the TAMI group compared to the in-person MI and treatment as usual groups [44,48]. Finally, 1 study reported that adding an avatar to a web-based intervention did not significantly increase self-reported physical activity compared to the web-based intervention without an avatar [51], and the avatar did not create a stronger therapeutic relationship with participants.

Many studies noted that TAMIs could be useful in communities that have limited access to health care. Studies that focused on these communities or had many participants from these communities reported generally positive results, that is, the TAMI produced statistically significantly higher engagement in health promotion behaviors than the control treatment. Three studies designed specifically for underserved populations all reported that the TAMI intervention was more successful at promoting the desired target behavior than the control [35,37,60]. Specifically, 1 study noted that the intervention, which was designed specifically for the needs of Māori (New Zealand's indigenous population) and non-Māori Pacific audiences, was equally effective for both populations and across different age groups [60]. Of the 18 studies reporting behavioral or health outcomes in a large proportion of participants who were racial or ethnic minorities or had limited incomes, 15 showed increases in engagement in healthy behaviors or better health outcomes [29,32-35,37-39,41,42,44-47,60]. Studies designed specifically for underinvested communities reported more success in promoting behavior change compared to studies that were not designed specifically for underinvested communities.

Discussion

Principal Findings

This scoping review extends prior work [23] by examining the following: (1) to what extent have individuals from marginalized sociodemographic groups been included in research on TAMIs; (2) how TAMIs offset sociocontextual influences on health, such as challenges imposed by the built environment or economics; and (3) how TAMIs may affect behavioral and health outcomes. The 34 studies reviewed here suggest that TAMIs likely improve health promotion and disease management behaviors, and health outcomes. However, the diversity in study designs, populations, and target behavioral or health outcomes preclude a formal meta-analysis at this point in time. The impact of TAMIs on health may be stronger among marginalized sociodemographic groups, including people from racial and ethnic minority backgrounds and those with low incomes. Although TAMIs have generally led to improvements in health promotion behaviors compared to control conditions, to date, TAMIs have been developed for a limited scope of health behaviors, the inclusion of individuals from marginalized sociodemographic groups has been minimal, and their impact on sociocontextual factors is not well understood.

Twenty-three studies reviewed reported that TAMIs were associated with statistically significant improvements in health promotion and disease management behaviors. However, 10 studies combined MI with other therapeutic approaches, like CBT, or intervention elements, such as counselor-mediated chat, that prevented an assessment of the unique impact of the TAMI. It is important to understand how these various therapeutic frameworks affected the results to identify what is necessary for a successful eHealth intervention. An experiment using the multiphase optimization strategy (MOST) design would allow investigators to parse which combination of several intervention components yields the greatest benefit for patients [64].

With the advent of more sophisticated technologies, such as machine learning and artificial intelligence, there have been many technological advancements in recent years, including the development of avatars capable of more personalized interactions and greater relational skills [24]. In fact, study participants who interact with a human-like virtual character may feel stronger social relations compared to interacting only with a plain text-based interface [65]. Yet, no study has specifically examined whether avatars improve an intervention's target health behavior, despite calls [51] for future research to examine the effects of an avatar with more complex relational features. Such research may have implications for future uses of virtual reality, automated counseling or counselors, and other technological advancements for behavioral counseling, and offer a unique perspective on the importance of replicating human conversation in mHealth counseling technologies [66].

Scope of Health Behaviors and Outcomes Addressed by TAMIs

A majority of studies focused on substance, alcohol, or tobacco use (n=24), with the next most common behavioral targets being treatment or medication adherence (n=5) and diet or physical activity (n=2). This is consistent with the key health behaviors identified by the Centers for Disease Control and Prevention as being critical to promoting health and preventing chronic diseases [1]. However, there is potential to increase the scope of TAMI interventions. Health screenings can be critically important for detecting early disease in healthy populations and for preventing disease progression in populations with chronic health conditions [1,67]. For example, some studies screened for health metrics, such as BMI and hemoglobin A_{1c}, which are very important for preventive screening [28,49]. In addition, health screenings are particularly important for high-risk populations, such as the nearly 17 million cancer survivors living in the United States [68], many of whom are vulnerable to experiencing the late effects of toxic treatments [69] but who may not be aware of the need to be screened. TAMIs that promote screening behaviors, particularly among cancer survivors, should be designed and tested. Such interventions may be particularly impactful for survivors of childhood cancers, who experience significant premature morbidity and mortality [70].

Inclusion of Individuals From Populations That Have Been Marginalized

Few of the studies reviewed were intentionally designed for marginalized sociodemographic groups. Although 59% (20/34) of the studies reviewed included an overrepresentation of individuals from marginalized groups, only 2 studies [37,60] explicitly stated that they received input from the marginalized communities for which the intervention was designed (ie, African American [37], and Māori and non-Māori Pacific individuals [60]). Only 2 studies included over 67% of individuals with low incomes [43,45]. However, no study addressed other factors, such as socioeconomic status, age, and sexual or gender identity. This is important because many of the studies reviewed suggested that TAMIs have the potential to be particularly beneficial for marginalized communities. For example, a study involving an MI intervention that was culturally tailored for Hispanic/Latino participants reported that “nearly all participants (95%) said that understanding their culture was important to understanding their drinking behavior” [71]. This suggests that TAMIs that consider the unique interests and needs of marginalized populations may have a more beneficial and sustainable impact than standard MI [20]. In addition, MI may be highly efficacious in minority populations due to its autonomy supportive approach rather than the authoritarian approach, which can trigger feelings of stigmatization and marginalization, that is commonly found in behavior change interventions [72,73]. However, research is needed to confirm this hypothesis. It also suggests that further studies that are more inclusive of marginalized participants are needed, so that the benefits of MI for groups that are at the greatest risk of experiencing health disparities can be better evaluated and, if appropriate, translated into practice and policy change.

Address Strategies for Overcoming Sociocontextual Barriers

In the studies we identified, there was scant attention to sociocontextual factors that shape and constrain people’s ability to engage in health promotion behaviors outside of race and ethnicity. This dearth of research is concerning, because it represents a missed opportunity to understand the role of TAMIs in overcoming powerful barriers to health behavior change [74]. For example, a TAMI focused on improving dietary behavior among people with low incomes might be more effective if information and other resources are provided to enable people living in food deserts to access fresh fruits and vegetables more easily. Similarly, a TAMI might offer information about strategies for exercising safely to people who indicate that they live in a neighborhood that is not conducive to outdoor exercise and cannot afford a gym membership.

Comparison With Prior Work

A previous systematic review of TAMIs did not examine health outcomes or the potential relevance of TAMIs to people from marginalized or underserved sociodemographic groups [23]. Our research extends that work by evaluating these important characteristics. In addition, our findings indicating that TAMIs may be effective for improving health behaviors are consistent with the findings of previous reports [23].

Limitations

This review should be considered in light of the following limitations. First, our analysis may have been limited by publication bias. While we found that a majority of studies reported positive results, those with negative results might remain disproportionately unpublished. The identified studies were also quite heterogeneous in populations, interventions, and outcomes assessed, restricting the conclusions that can be drawn when considering the results in aggregate. In addition, similar to a 2016 systematic review [23], only 1 study included a counselor-mediated MI group, so we are unable to report the efficacy of TAMIs compared to traditional MI interventions [23]. However, reports that TAMIs produce more beneficial changes in health promotion and disease management behaviors compared to no treatment or treatment as usual are critical, because traditional MI interventions are very costly and difficult to disseminate widely.

Conclusions

MI has been largely successful in influencing positive behavioral change. Given the rapid increase in technological advancements in recent decades, TAMIs offer a low-cost and accessible platform to help patients improve their health. The results of this scoping review suggest that TAMIs are likely an effective way to promote positive behavioral change for the prevention and management of chronic diseases. TAMIs may hold particular promise for improving the health of marginalized communities, but few studies have described tailoring TAMIs in a culturally relevant, appropriate, or meaningful manner. In addition, no studies we reviewed addressed major sociocontextual factors that shape and constrain people’s ability to initiate and maintain changes in health behaviors. Studies that are more inclusive of communities that have been marginalized or underserved and that adequately address the sociocontextual factors shaping health could help reduce health disparities.

Acknowledgments

This work was supported by the St. Jude Children's Research Hospital – Washington University St. Louis Implementation Sciences Collaborative. MJE and MMH are employed by St. Jude Children's Research Hospital.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the studies included in the scoping review.
[XLSX File (Microsoft Excel File), 18 KB - jmir_v24i8e35283_app1.xlsx]

Multimedia Appendix 2
Additional characteristics of the studies included in the scoping review.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v24i8e35283_app2.xlsx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 108 KB - jmir_v24i8e35283_app3.docx]

References

1. About Chronic Diseases. National Center for Chronic Disease Prevention and Health Promotion. URL: https://www.cdc.gov/chronicdisease/about/index.htm [accessed 2022-07-18]
2. Johnson BT, Scott-Sheldon LAJ, Carey MP. Meta-synthesis of health behavior change meta-analyses. Am J Public Health 2010 Nov;100(11):2193-2198 [FREE Full text] [doi: 10.2105/AJPH.2008.155200] [Medline: 20167901]
3. Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of healthcare. Eur Heart J 2016 May 07;37(18):1428-1438 [FREE Full text] [doi: 10.1093/eurheartj/ehv770] [Medline: 26873093]
4. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: Systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
5. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
6. Amdie F, Woo K. The use of mHealth technology for chronic disease management: the challenges and opportunities for practical application. Wounds International 2020;11(2):32-38 [FREE Full text]
7. Brath H, Morak J, Kästenbauer T, Modre-Osprian R, Strohner-Kästenbauer H, Schwarz M, et al. Mobile health (mHealth) based medication adherence measurement - a pilot trial using electronic blisters in diabetes patients. Br J Clin Pharmacol 2013 Sep 06;76 Suppl 1:47-55 [FREE Full text] [doi: 10.1111/bcp.12184] [Medline: 24007452]
8. Mobile Fact Sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-07-18]
9. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
10. Connell CL, Wang SC, Crook L, Yadrick K. Barriers to healthcare seeking and provision among African American adults in the rural Mississippi Delta region: Community and provider perspectives. J Community Health 2019 Aug;44(4):636-645 [FREE Full text] [doi: 10.1007/s10900-019-00620-1] [Medline: 30661152]
11. Washington HA. Medical Apartheid: The Dark History of Medical Experimentation on Black Americans from Colonial Times to the Present. Hamburg, Germany: Anchor; 2007.
12. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
13. Miller W, Rollnick S. Motivational Interviewing, Third Edition: Helping People Change. New York, NY: The Guilford Press; 2012.
14. Miller WR, Rollnick S. Ten things that motivational interviewing is not. Behav Cogn Psychother 2009 Mar;37(2):129-140. [doi: 10.1017/S1352465809005128] [Medline: 19364414]
15. Miller WR, Rose GS. Toward a theory of motivational interviewing. Am Psychol 2009 Sep;64(6):527-537 [FREE Full text] [doi: 10.1037/a0016830] [Medline: 19739882]
16. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312 [FREE Full text] [Medline: 15826439]
17. Whitlock EP, Orleans CT, Pender N, Allan J. Evaluating primary care behavioral counseling interventions: an evidence-based approach. Am J Prev Med 2002 May;22(4):267-284. [doi: 10.1016/s0749-3797(02)00415-4] [Medline: 11988383]

18. Markland D, Ryan RM, Tobin VJ, Rollnick S. Motivational interviewing and self–determination theory. Journal of Social and Clinical Psychology 2005 Sep;24(6):811-831. [doi: 10.1521/jscp.2005.24.6.811]
19. Hettema J, Steele J, Miller W. Motivational interviewing. Annu Rev Clin Psychol 2005;1:91-111 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.143833] [Medline: 17716083]
20. Stanton M. Motivational interviewing and the social context. Am Psychol 2010;65(4):297-8; discussion 298. [doi: 10.1037/a0019276] [Medline: 20455630]
21. Coyne N, Correnti D. Effectiveness of motivational interviewing to improve chronic condition self-management: what does the research show us? Home Healthc Nurse 2014 Jan;32(1):56-63. [doi: 10.1097/NHH.0000000000000001] [Medline: 24326477]
22. Wu L, Gao X, Lo ECM, Ho SMY, McGrath C, Wong MCM. Motivational interviewing to promote oral health in adolescents. J Adolesc Health 2017 Sep;61(3):378-384. [doi: 10.1016/j.jadohealth.2017.03.010] [Medline: 28532895]
23. Shingleton RM, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: A systematic review of the current research. Patient Educ Couns 2016 Jan;99(1):17-35 [FREE Full text] [doi: 10.1016/j.pec.2015.08.005] [Medline: 26298219]
24. Park S, Choi J, Lee S, Oh C, Kim C, La S, et al. Designing a chatbot for a brief motivational interview on stress management: Qualitative case study. J Med Internet Res 2019 Apr 16;21(4):e12231 [FREE Full text] [doi: 10.2196/12231] [Medline: 30990463]
25. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
27. 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]
28. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
29. Braciszewski JM, Tzilos Wernette GK, Moore RS, Bock BC, Stout RL, Chamberlain P. A pilot randomized controlled trial of a technology-based substance use intervention for youth exiting foster care. Child Youth Serv Rev 2018 Nov;94:466-476 [FREE Full text] [doi: 10.1016/j.childyouth.2018.08.011] [Medline: 31435121]
30. Breland AB, Almond L, Kienzle J, Ondersma SJ, Hart A, Weaver M, et al. Targeting tobacco in a community-based addiction recovery cohort: results from a computerized, brief, randomized intervention trial. Contemp Clin Trials 2014 May;38(1):113-120. [doi: 10.1016/j.cct.2014.03.008] [Medline: 24721481]
31. Brunette MF, Ferron JC, McGurk SR, Williams JM, Harrington A, Devitt T, et al. Brief, web-based interventions to motivate smokers with schizophrenia: Randomized trial. JMIR Ment Health 2020 Feb 08;7(2):e16524 [FREE Full text] [doi: 10.2196/16524] [Medline: 32039811]
32. Brunette MF, Ferron JC, McHugo GJ, Davis KE, Devitt TS, Wilkness SM, et al. An electronic decision support system to motivate people with severe mental illnesses to quit smoking. Psychiatr Serv 2011 Apr;62(4):360-366. [doi: 10.1176/ps.62.4.pss6204_0360] [Medline: 21459986]
33. Chernick LS, Stockwell MS, Gonzalez A, Mitchell J, Ehrhardt A, Bakken S, et al. A user-informed, theory-based pregnancy prevention intervention for adolescents in the emergency department: A prospective cohort study. J Adolesc Health 2021 Apr;68(4):705-712 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.020] [Medline: 32948403]
34. Choo EK, Zlotnick C, Strong DR, Squires DD, Tapé C, Mello MJ. BSAFER: A Web-based intervention for drug use and intimate partner violence demonstrates feasibility and acceptability among women in the emergency department. Subst Abus 2016;37(3):441-449 [FREE Full text] [doi: 10.1080/08897077.2015.1134755] [Medline: 26714233]
35. Ellis DA, Idalski Carcone A, Ondersma SJ, Naar-King S, Dekelbab B, Moltz K. Brief computer-delivered intervention to increase parental monitoring in families of African American adolescents with type 1 diabetes: A randomized controlled trial. Telemed J E Health 2017 Jun;23(6):493-502 [FREE Full text] [doi: 10.1089/tmj.2016.0182] [Medline: 28061319]
36. Kazemi DM, Borsari B, Levine MJ, Shehab M, Nelson M, Dooley B, et al. Real-time demonstration of a mHealth app designed to reduce college students hazardous drinking. Psychol Serv 2019 May;16(2):255-259 [FREE Full text] [doi: 10.1037/ser0000310] [Medline: 30407059]
37. Kolmodin MacDonell K, Naar S, Gibson-Scipio W, Lam P, Secord E. The Detroit Young Adult Asthma Project: Pilot of a technology-based medication adherence intervention for African-American emerging adults. J Adolesc Health 2016 Oct;59(4):465-471 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.05.016] [Medline: 27475032]
38. Mason M, Benotsch EG, Way T, Kim H, Snipes D. Text messaging to increase readiness to change alcohol use in college students. J Prim Prev 2014 Feb;35(1):47-52. [doi: 10.1007/s10935-013-0329-9] [Medline: 24114551]

39. Mason M, Mennis J, Way T, Zaharakis N, Campbell LF, Benotsch EG, et al. Text message delivered peer network counseling for adolescent smokers: A randomized controlled trial. J Prim Prev 2016 Oct;37(5):403-420. [doi: 10.1007/s10935-016-0439-2] [Medline: 27388626]
40. Mason MJ, Campbell L, Way T, Keyser-Marcus L, Benotsch E, Mennis J, et al. Development and outcomes of a text messaging tobacco cessation intervention with urban adolescents. Subst Abus 2015;36(4):500-506. [doi: 10.1080/08897077.2014.987946] [Medline: 25551337]
41. Naar-King S, Outlaw AY, Sarr M, Parsons JT, Belzer M, Macdonell K, Adolescent Medicine Network for HIV/AIDS Interventions. Motivational Enhancement System for Adherence (MESA): pilot randomized trial of a brief computer-delivered prevention intervention for youth initiating antiretroviral treatment. J Pediatr Psychol 2013 Jul;38(6):638-648 [FREE Full text] [doi: 10.1093/jpepsy/jss132] [Medline: 23359664]
42. Ondersma SJ, Svikis DS, Schuster CR. Computer-based brief intervention a randomized trial with postpartum women. Am J Prev Med 2007 Mar;32(3):231-238 [FREE Full text] [doi: 10.1016/j.amepre.2006.11.003] [Medline: 17236741]
43. Ondersma SJ, Svikis DS, Thacker LR, Beatty JR, Lockhart N. Computer-delivered screening and brief intervention (e-SBI) for postpartum drug use: a randomized trial. J Subst Abuse Treat 2014 Jan;46(1):52-59 [FREE Full text] [doi: 10.1016/j.jsat.2013.07.013] [Medline: 24051077]
44. Osilla KC, Paddock SM, Leininger TJ, D'Amico EJ, Ewing BA, Watkins KE. A pilot study comparing in-person and web-based motivational interviewing among adults with a first-time DUI offense. Addict Sci Clin Pract 2015 Sep 03;10:18 [FREE Full text] [doi: 10.1186/s13722-015-0039-0] [Medline: 26334629]
45. Ranney ML, Pittman SK, Dunsiger S, Guthrie KM, Spirito A, Boyer EW, et al. Emergency department text messaging for adolescent violence and depression prevention: A pilot randomized controlled trial. Psychol Serv 2018 Nov;15(4):419-428 [FREE Full text] [doi: 10.1037/ser0000193] [Medline: 30382737]
46. Suffoletto B, Kristan J, Callaway C, Kim KH, Chung T, Monti PM, et al. A text message alcohol intervention for young adult emergency department patients: a randomized clinical trial. Ann Emerg Med 2014 Dec;64(6):664-72.e4 [FREE Full text] [doi: 10.1016/j.annemergmed.2014.06.010] [Medline: 25017822]
47. Tzilos Wernette G, Plegue M, Kahler CW, Sen A, Zlotnick C. A pilot randomized controlled trial of a computer-delivered brief intervention for substance use and risky sex during pregnancy. J Womens Health (Larchmt) 2018 Jan;27(1):83-92 [FREE Full text] [doi: 10.1089/jwh.2017.6408] [Medline: 28981379]
48. Lerch J, Walters ST, Tang L, Taxman FS. Effectiveness of a computerized motivational intervention on treatment initiation and substance use: Results from a randomized trial. J Subst Abuse Treat 2017 Sep;80:59-66 [FREE Full text] [doi: 10.1016/j.jsat.2017.07.002] [Medline: 28755774]
49. Armstrong S, Mendelsohn A, Bennett G, Taveras EM, Kimberg A, Kemper AR. Texting motivational interviewing: A randomized controlled trial of motivational interviewing text messages designed to augment childhood obesity treatment. Child Obes 2018 Jan;14(1):4-10. [doi: 10.1089/chi.2017.0089] [Medline: 29019418]
50. Blankers M, Koeter MWJ, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: A randomized controlled trial. J Consult Clin Psychol 2011 Jun;79(3):330-341. [doi: 10.1037/a0023498] [Medline: 21534652]
51. 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]
52. Friederichs SAH, Oenema A, Bolman C, Lechner L. Motivational interviewing and self-determination theory in a web-based computer tailored physical activity intervention: A randomized controlled trial. Psychol Health 2016 Aug;31(8):907-930. [doi: 10.1080/08870446.2016.1151018] [Medline: 26849996]
53. Bommelé J, Schoenmakers TM, Kleinjan M, Peters GY, Dijkstra A, van de Mheen D. Targeting hardcore smokers: The effects of an online tailored intervention, based on motivational interviewing techniques. Br J Health Psychol 2017 Sep;22(3):644-660. [doi: 10.1111/bjhp.12256] [Medline: 28703470]
54. Becker J, Haug S, Sullivan R, Schaub MP. Effectiveness of different Web-based interventions to prepare co-smokers of cigarettes and cannabis for double cessation: a three-arm randomized controlled trial. J Med Internet Res 2014 Dec 05;16(12):e273 [FREE Full text] [doi: 10.2196/jmir.3246] [Medline: 25486674]
55. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: Three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
56. Boß L, Lehr D, Schaub MP, Paz Castro R, Riper H, Berking M, et al. Efficacy of a web-based intervention with and without guidance for employees with risky drinking: results of a three-arm randomized controlled trial. Addiction 2018 Apr;113(4):635-646 [FREE Full text] [doi: 10.1111/add.14085] [Medline: 29105879]
57. Jo S, Lee HK, Kang K, Joe KH, Lee S. Efficacy of a web-based screening and brief intervention to prevent problematic alcohol use in Korea: Results of a randomized controlled trial. Alcohol Clin Exp Res 2019 Oct;43(10):2196-2202. [doi: 10.1111/acer.14169] [Medline: 31386203]

58. Philipp J, Franta C, Zeiler M, Truttmann S, Wittek T, Imgart H, et al. Does a skills intervention for parents have a positive impact on adolescents' anorexia nervosa outcome? Answers from a quasi-randomised feasibility trial of SUCCEAT. Int J Environ Res Public Health 2021 Apr 27;18(9):4656 [FREE Full text] [doi: 10.3390/ijerph18094656] [Medline: 33925694]
59. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: Design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020 Nov 03;22(11):e20251 [FREE Full text] [doi: 10.2196/20251] [Medline: 33141095]
60. Sharpe S, Kool B, Whittaker R, Lee AC, Reid P, Civil I, et al. Effect of a text message intervention to reduce hazardous drinking among injured patients discharged from a trauma ward: a randomized controlled trial. NPJ Digit Med 2018 May 2;1(1):13 [FREE Full text] [doi: 10.1038/s41746-018-0019-3] [Medline: 31304298]
61. Arnaud N, Baldus C, Elgán TH, De Paepe N, Tønnesen H, Csémy L, et al. Effectiveness of a web-based screening and fully automated brief motivational intervention for adolescent substance use: A randomized controlled trial. J Med Internet Res 2016 May 24;18(5):e103 [FREE Full text] [doi: 10.2196/jmir.4643] [Medline: 27220276]
62. Welcome to CIAS. Intervention Authoring. URL: https://www.interventionauthoring.net/ [accessed 2022-07-18]
63. QuickFacts. United States Census. URL: https://www.census.gov/quickfacts/fact/table/US/PST045221 [accessed 2022-07-18]
64. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
65. Appel J, von der Pütten A, Krämer NC, Gratch J. Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. Advances in Human-Computer Interaction 2012;2012:1-10. [doi: 10.1155/2012/324694]
66. Boeldt D, McMahon E, McFaul M, Greenleaf W. Using virtual reality exposure therapy to enhance treatment of anxiety disorders: Identifying areas of clinical adoption and potential obstacles. Front Psychiatry 2019 Oct 25;10:773 [FREE Full text] [doi: 10.3389/fpsyt.2019.00773] [Medline: 31708821]
67. Long-Term Follow-Up Guidelines for Survivors of Childhood, Adolescent, and Young Adult Cancers. Children's Oncology Group. URL: http://www.survivorshipguidelines.org [accessed 2022-07-18]
68. Cancer Statistics. National Cancer Institute. URL: https://www.cancer.gov/about-cancer/understanding/statistics [accessed 2022-07-18]
69. Late Effects of Cancer Treatment. National Cancer Institute. URL: https://www.cancer.gov/about-cancer/coping/survivorship/late-effects [accessed 2022-07-18]
70. Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: life-long risks and responsibilities. Nat Rev Cancer 2014 Jan;14(1):61-70 [FREE Full text] [doi: 10.1038/nrc3634] [Medline: 24304873]
71. Lee CS, López SR, Hernández L, Colby SM, Caetano R, Borrelli B, et al. A cultural adaptation of motivational interviewing to address heavy drinking among Hispanics. Cultur Divers Ethnic Minor Psychol 2011 Jul;17(3):317-324 [FREE Full text] [doi: 10.1037/a0024035] [Medline: 21787064]
72. Lundahl BW, Kunz C, Brownell C, Tollefson D, Burke BL. A meta-analysis of motivational interviewing: Twenty-five years of empirical studies. Research on Social Work Practice 2010 Jan 11;20(2):137-160. [doi: 10.1177/1049731509347850]
73. Hettema J, Steele J, Miller WR. Motivational interviewing. Annu Rev Clin Psychol 2005 Apr 01;1(1):91-111. [doi: 10.1146/annurev.clinpsy.1.102803.143833] [Medline: 17716083]
74. Stokols D. Translating social ecological theory into guidelines for community health promotion. Am J Health Promot 1996;10(4):282-298. [doi: 10.4278/0890-1171-10.4.282] [Medline: 10159709]

Abbreviations

CBT: cognitive behavioral therapy
CIAS: Computerized Intervention Authorizing Software
MI: motivational interviewing
RCT: randomized controlled trial
TAMI: technology-delivered adaptation of motivational interviewing

Edited by A Mavragani; submitted 29.11.21; peer-reviewed by H Mehdizadeh, H Ayatollahi, B Borsari, A Oenema; comments to author 31.01.22; revised version received 01.06.22; accepted 27.06.22; published 09.08.22.

Please cite as:
Pedamallu H, Ehrhardt MJ, Maki J, Carcone AI, Hudson MM, Waters EA
Technology-Delivered Adaptations of Motivational Interviewing for the Prevention and Management of Chronic Diseases: Scoping Review
J Med Internet Res 2022;24(8):e35283
URL: https://www.jmir.org/2022/8/e35283
doi:10.2196/35283
PMID:35943775

Review

Overview of Artificial Intelligence–Driven Wearable Devices for Diabetes: Scoping Review

Arfan Ahmed[1], PhD; Sarah Aziz[1], MSc; Alaa Abd-alrazaq[1], PhD; Faisal Farooq[2], PhD; Javaid Sheikh[1], MD

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

[2]Center for Digital Health and Precision Medicine, Qatar Computing Research Institute, Doha, Qatar

Corresponding Author:
Arfan Ahmed, PhD
AI Center for Precision Health
Weill Cornell Medicine-Qatar
Education City, Qatar Foundation
PO Box 24144
Doha
Qatar
Phone: 974 4492 8826
Email: ara4013@qatar-med.cornell.edu

Abstract

Background: Prevalence of diabetes has steadily increased over the last few decades with 1.5 million deaths reported in 2012 alone. Traditionally, analyzing patients with diabetes has remained a largely invasive approach. Wearable devices (WDs) make use of sensors historically reserved for hospital settings. WDs coupled with artificial intelligence (AI) algorithms show promise to help understand and conclude meaningful information from the gathered data and provide advanced and clinically meaningful analytics.

Objective: This review aimed to provide an overview of AI-driven WD features for diabetes and their use in monitoring diabetes-related parameters.

Methods: We searched 7 of the most popular bibliographic databases using 3 groups of search terms related to diabetes, WDs, and AI. A 2-stage process was followed for study selection: reading abstracts and titles followed by full-text screening. Two reviewers independently performed study selection and data extraction, and disagreements were resolved by consensus. A narrative approach was used to synthesize the data.

Results: From an initial 3872 studies, we report the features from 37 studies post filtering according to our predefined inclusion criteria. Most of the studies targeted type 1 diabetes, type 2 diabetes, or both (21/37, 57%). Many studies (15/37, 41%) reported blood glucose as their main measurement. More than half of the studies (21/37, 57%) had the aim of estimation and prediction of glucose or glucose level monitoring. Over half of the reviewed studies looked at wrist-worn devices. Only 41% of the study devices were commercially available. We observed the use of multiple sensors with photoplethysmography sensors being most prevalent in 32% (12/37) of studies. Studies reported and compared >1 machine learning (ML) model with high levels of accuracy. Support vector machine was the most reported (13/37, 35%), followed by random forest (12/37, 32%).

Conclusions: This review is the most extensive work, to date, summarizing WDs that use ML for people with diabetes, and provides research direction to those wanting to further contribute to this emerging field. Given the advancements in WD technologies replacing the need for invasive hospital setting devices, we see great advancement potential in this domain. Further work is needed to validate the ML approaches on clinical data from WDs and provide meaningful analytics that could serve as data gathering, monitoring, prediction, classification, and recommendation devices in the context of diabetes.

(J Med Internet Res 2022;24(8):e36010) doi:10.2196/36010

KEYWORDS

diabetes; artificial intelligence; wearable devices; machine learning; mobile phone

Introduction

Background

Diabetes, also known as diabetes mellitus, is a metabolic disease characterized by elevated blood glucose levels, which can ultimately result in many complications such as heart attack, stroke, kidney failure, leg amputation, vision loss, and nerve damage [1]. As the world embarks on a centennial anniversary since the development of insulin to manage glucose levels of people with diabetes, we have seen remarkable advances during these 100 years, with improved life expectancy and quality of life [2]. Noncommunicable diseases such as metabolic syndrome and diabetes continue to be among the leading causes of disability and mortality [3]. The number of cases and their prevalence have steadily increased over the last few decades. According to the World Health Organization, 1.5 million people died in 2012 alone because of diabetes, with an additional 2.1 million deaths caused by a higher than optimal blood glucose level, resulting in increased risks of cardiovascular and other diseases. A total of 463 million people, globally, were affected by type 2 diabetes (T2D) mellitus in 2019. Furthermore, it is predicted that 700 million individuals would develop diabetes by 2045 [4]. Although the World Health Organization acknowledges that there is no one fixed solution and that a coordinated multicomponent intervention is needed, it outlines technology as one of the key stakeholders in reducing the impact of diabetes in addition to input from governments, health care providers, people with diabetes, civil society, food producers and manufacturers, and suppliers of medicine [1].

Despite the advancements in blood glucose monitoring techniques, the mainstream detection technology remains largely invasive. The commonly used home electronic glucometers involve people with diabetes invasively self-pricking to draw blood from fingertips, opening them up to infections as well as stress and pain caused by the procedure that is often expected multiple times a day.

The availability and advancements of smart devices, such as smartphones, have made the monitoring of diabetes-related features more accessible. Many studies have examined this much welcomed technology [5,6]. These normally require the use of an external attachable sensor, and monitoring is then delivered via an app or a separate continuous glucose monitoring (CGM) device, which can still be semi-invasive and require a connection range via Bluetooth or Wi-Fi signals. The use of completely noninvasive technology in the form of wearable devices (WDs) for regulating and monitoring glucose levels for people with diabetes is a fairly new concept and is in its infancy. Commercially available devices, such as smart watches and smart bands, can take measurements using sensors that researchers have reported on their usefulness in diabetes monitoring [7,8]. Such technologies can be affordable and easily accessible, and when used properly, can improve the quality of life of patients in a noninvasive manner. With their widespread commercial use and acceptance owing to their fashionable nature, globally researchers have a unique opportunity to provide medical care away from hospital settings and bulky invasive hardware in an affordable manner without requiring expert assistance. WDs have an increasing capacity, although not at the level of smartphones, to gather, store, transmit, and process data; the features can then be used for management, treatment, assessment, and sometimes even prediction. Furthermore, many WDs are normally connected via Wi-Fi or Bluetooth to external devices, such as a smartphone, where computationally expensive processing is performed for the simple purpose of storage or as a gateway to cloud spaces. Cloud storage can facilitate monitoring by clinicians without the need of hospitalization. Several useful sensors already exist incorporated into WDs similar to those of smartphones, including electrocardiogram (ECG), photoplethysmography, galvanic skin response, near infrared, and accelerometer sensors. WDs have additional advantages when it comes to sensing physiological signs, such as heart rate, ECG, and skin temperature. This is largely owing to their close contact with the wearer, which is of particular interest when monitoring diabetes-related metrics.

Artificial intelligence (AI) is a broader term that encompasses machine learning (ML). Technically, ML is a subset of AI, often loosely used interchangeable buzzwords. As a high-level definition, AI is anything related to making machines smarter (eg, computational search algorithms). ML, on the other hand, is an AI system that can self-learn via an algorithm, and as a result, such a system becomes smarter without human intervention over time (eg, classifying an outcome) [9]. Deep learning, on the other hand, is another branch of AI that attempts to mimic the human brain in terms of how it processes large amounts of data and has already shown success rates in areas such as diabetic retinopathy screening [10]. ML principles have been applied in clinical settings to build algorithms to support predictive models for the risk of development of diabetes [11]. AI has also been shown to provide useful management tools to deal with large amounts of data [12]. Owing to the large amount of data measurable through continuous monitoring via wearables, AI can be used to further analyze the acquired data. This can help to understand and draw meaningful information from the gathered data and provide advanced and clinically meaningful analytics. Many researchers have adapted existing WDs not originally intended for diabetes management and adapted the sensory information for use in diabetes-related metrics, and some have created prototypes especially designed for diabetes [13,14]. WDs are used for a variety of reasons, including monitoring, prevention, glucose estimation, diagnostics, classification, and prevention, but the number of studies that are reported are low in comparison with those that make use of smartphones for example. With the increased potential outreach of WDs globally, especially when combined with the ever-expanding field of AI-incorporating ML algorithms, the correct management of large amounts of data and processing with ML algorithms holds great potential for quality-of-life improvement in people with diabetes [15].

Research Problem and Aim

Many studies have been conducted on AI-based WDs for diabetes. Exploring the features of AI-based WDs reported in these studies is important for developers, patients, health care providers, and researchers to identify the recent advances and challenges in this area. Although several reviews were conducted in this area, (1) they were focused on smartphones and AI for diabetes [16-18], (2) they were focused on WDs in general rather than AI-based WDs [17,19], and (3) they did not summarize the features of AI-based WDs in a thorough manner [16-19]. Therefore, we aimed to explore the features of AI-based WDs for diabetes as reported in previous studies. We believe that this review will allow developers and researchers to advance further in this field by highlighting the gaps and opportunities.

Methods

Overview

This scoping review was carried out to satisfy this study's goals of exploring features of AI-driven wearable technologies for diabetes. In order to construct a complete scoping review, the *PRISMA-ScR* (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [20] was used as a guiding approach. The PRISMA-ScR checklist is shown in Multimedia Appendix 1.

Search Strategy

Search Sources

The article search for this review began by identifying all relevant studies using 7 electronic databases: MEDLINE, PsycINFO, EMBASE, IEEE Xplore, ACM Digital Library, Web of Science, and Google Scholar. We scanned the first 100 hits retrieved by searching Google Scholar. The reason being Google Scholar typically returns several items that are sorted by relevance to the search topic. Bibliographic collection was conducted from October 25 to October 30, 2021. The reference lists of the included articles were then searched for additional sources. We also checked relevant articles that cited the included studies using Google Scholar's "cited by" tool (forward reference list checking).

Search Terms

A number of different sets of keywords were designed to search databases depending on each database's search term limit; as IEEE and Google Scholar have term limits, search queries were truncated based on the required limit. We considered the research topics included in the database to complete our search queries. We combined *Diabetic* OR *Diabetes* keywords describing the relevant population (people with diabetes), with each kind of relevant intervention to wearables (*wearable** OR *smart watch** OR *smart** OR *smartwatch** OR *fitness band* OR flexible band* OR wristband* OR smart insole* OR bracelet**) and AI (*Artificial Intelligence OR Machine Learning OR Deep Learning OR Decision tree OR K-Nearest Neighbor* OR Support vector machine* OR Recurrent neural network* OR convolutional neural network* OR Artificial neural network* OR Naïve Bayes OR Naive Bayes OR Fuzzy Logic OR K-Means OR Random Forest OR LSTM OR autoencoder OR boltzmann machine OR deep belief network*). For example, the following search terms were applied in Google Scholar: (*Artificial Intelligence OR Machine Learning OR Deep Learning OR convolutional neural network* OR Artificial neural network*) AND (wearable* OR smart watch* OR smart*) AND (Diabetic OR Diabetes)*. All the databases had search time period criteria that were enabled and set with the search query from 2015 to present; in addition, the language checkbox in each database was set to English only. Full search terms for each electronic database searched are available in Multimedia Appendix 2.

Studies were chosen based on the criteria in Textbox 1. Peer-reviewed articles and published protocols were included only if they were related to wearables that could be used by an individual outside of a clinical setting. They also had to use AI for the purpose of diabetes and be classified as noninvasive. For full inclusion and exclusion criteria refer to Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

1. Publications that are in the English language.
2. Peer-reviewed articles including proposals.
3. Population with or suspected to have diabetes. No restrictions regarding their age, gender, and ethnicity.
4. Commercial, medical, or prototypes but with condition wearable device and uses artificial intelligence (AI).
5. Wearable usable by individual person not with help of clinical staff or plugged in to hospital setting.
6. Wearables using methods for diabetes analysis are to be noninvasive.

Exclusion criteria

1. Any study that does not contain AI as an intervention.
2. People with other diseases, health care providers, and caregivers as population.
3. Not a wearable device (example artificial implant or body infused).
4. Studies opting statistical measures only, for analysis of collected data.
5. Sensors or tracking devices infused inside a person's body.
6. Wearable devices that need professional sittings or hospital sittings.

Study Selection

This review's studies were selected in 2 steps. In the first stage, 2 reviewers (AA and SA) independently reviewed the titles and abstracts of all retrieved papers. In the second phase, the same reviewers individually read the whole texts of the papers included in the first step. Rayyan (Qatar Computing Research Institute, Hamad Bin Khalifa University) [21], a web-based tool developed for data management for systematic and scoping reviews, was used to upload all the articles acquired from databases in a Research Information Systems format; then, filtering and citations were managed. During the first and second steps of the selection process, any disagreements between the 2 reviewers were resolved through conversation and decisions were made based on consensus.

Data Extraction

AA and SA constructed the data extraction form, as shown in Multimedia Appendix 3. The data extraction technique was carried out independently by 2 reviewers (AA and SA), and any discrepancies were resolved by discussion and consensus. Microsoft Excel was used to record the data extracted.

Data Synthesis

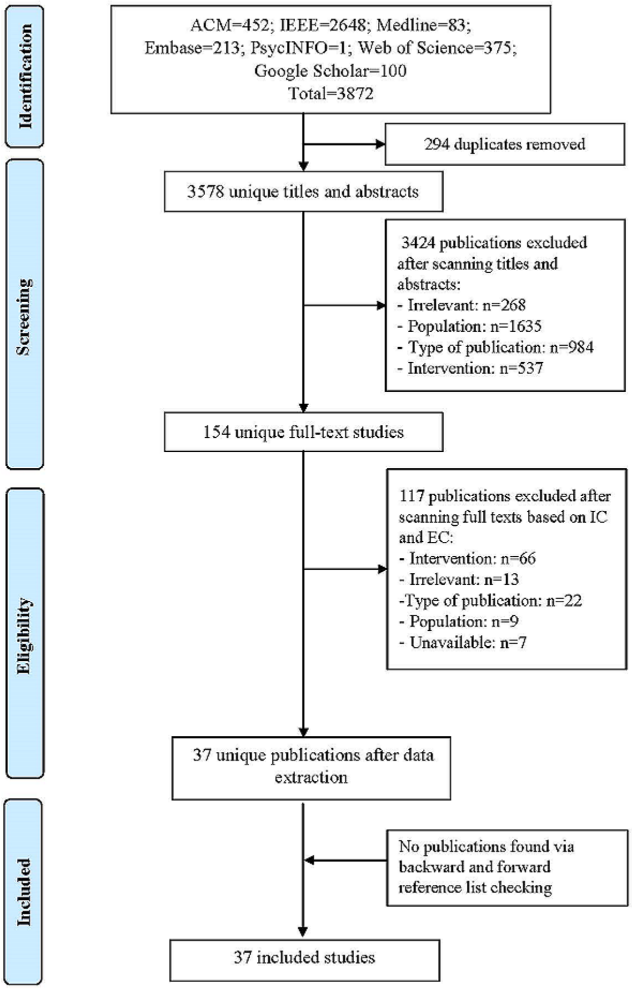
SA synthesized the extracted data using the narrative approach, aggregating the data using tables and text and nonstatistical techniques. For being more precise, we presented the search results followed by general features of the studies, finally describing characteristics of the WDs and AI technologies. We described the general features of WDs (eg, device placement, type, and operating system [OS]) and their technical features (ie, features of sensors, such as sensors used, sensing approach, and primary measurements). The AI features were addressed based on the models used, the evaluation metrics, and their applications.

Results

Search Results

Having searched 7 bibliographic databases, this study returned 3872 citations. As shown in Figure 1, a total of 294 duplicates were subsequently removed, leaving 3578 unique titles, and abstracts; publications that did not make use of AI technologies via WDs for diabetes management were considered irrelevant. Of these, we further excluded 3424 citations after screening their titles and abstracts. Of the remaining 154 references, 117 publications were excluded during the full-text screening. We were left with 37 studies, and this number remained unchanged even after performing backward and forward reference list checking. The synthesis included a total of 37 articles (Multimedia Appendix 4 [7,8,13,14,22-54]).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart of the study selection process. EC: exclusion criteria; IC: inclusion criteria.



General Description of Included Studies

Table 1 highlights the general features of the studies included. Many of the included studies (27/37, 73%) were published between 2019 and 2021, with the remaining 27% (10/37) published between 2016 and 2018. Most of the included studies were published in IEEE (21/37, 57%). A large proportion of the studies were authored by institutes in the United States (7/37, 19%), China (5/37, 14%), and India (5/37, 14%). A total of 26 of the 37 studies (70%) were journal articles, and the remainder were conference proceedings (11/37, 30%). Most of the studies targeted type 1 diabetes (T1D), T2D, or both (21/37, 57%), whereas 32% (12/37) did not specify the type of diabetes and mentioned diabetes in general. The remainder targeted prediabetes or a combination of T1D, T2D, and prediabetes (4/37, 11%). Features of each included study are shown in Multimedia Appendix 5.

Table 1. General features of included studies (n=37).

Features	Studies, n (%)	Study ID
Year		
2019	10 (27)	S4, S8, S10, S16, S18, S20, S21, S24, S29, S30
2020	9 (24)	S3, S7, S11, S13, S15, S17, S19, S22, S35
2021	8 (22)	S5, S9, S12, S14, S25, S27, S28, S33
2018	6 (16)	S1, S6, S23, S34, S36, S37
2017	3 (8)	S2, S26, S32
2016	1 (3)	S31
Publisher		
IEEE	21 (57)	S1, S3, S5, S9-S11, S13-S18, S20, S24, S26, S28, S29, S31, S32, S36, S37
Elsevier	3 (8)	S2, S12, S22
MDPI	3 (8)	S6-S8
ACM	2 (5)	S21, S35
Other (JMIR, IET, ICST, Confluence, BMJ Publishing Group, SPIE, Telemedicine and e-Health, SAGE)	8[a] (22)	S4, S19, S23, S25, S27, S30, S33, S34
Country		
United States	7 (19)	S13, S14, S21, S27, S30, S31, S34
China	5 (14)	S5, S15, S18, S19, S37
India	5 (14)	S12, S17, S23, S25, S32
Pakistan	2 (5)	S3, S20
Switzerland	2 (5)	S6, S35
Bangladesh	2 (5)	S10, S24
Other (Korea, Colombia, Canada, Morocco, Mexico, Italy, Macedonia, Sri Lanka, United Kingdom, Russia, Taiwan, Philippines, Saudi Arabia, Germany)	14[b] (38)	S1, S2, S4, S7, S8, S9, S11, S16, S22, S26, S28, S29, S33, S36
Publication type		
Journal articles	26 (70)	S1-S20, S22, S23, S27, S28, S33, S34
Conference proceedings	11 (30)	S21, S24-S26, S20-S32, S35-S37
Diabetes type studied		
Both T1D[c] and T2D[d]	9 (24)	S1, S5, S6, S8, S10, S11, S14, S24, S36
T2D	7 (19)	S2-S4, S15, S16, S21, S29
T1D	5 (14)	S13, S22, S30, S34, S35
T1D, T2D, and prediabetes	2 (5)	S12, S25
T1D and prediabetes	1 (3)	S17
Prediabetes	1 (3)	S27
Not specified	12 (32)	S7, S9, S18, S19, S20, S23, S26, S28, S31, S32, S33, S37

[a] 1 study for each publication.

[b] 1 study for each country.

[c] T1D: type 1 diabetes.

[d] T2D: type 2 diabetes.

Study Design Features

Table 2 outlines details about the studies associated with this review. More than half of the studies (21/37, 57%) had the aim of estimation and prediction of glucose (10/37, 27%) or glucose level monitoring (11/37, 30%). A couple of studies had multiple aims, and the remainder aimed to provide diabetes classification (4/37, 11%), diagnostic solutions (5/37, 14%), self-administration and monitoring (4/37, 11%), and prevention (2/37, 5%). Most of the studies did not mention anything about security (31/37, 84%); the remainder did specify security measures taken (6/37, 16%). Participants' demographics depicted most of them being adult (18/37, 49%) with both genders considered in equal proportional in most of the studies (10/37, 27%). Approximately 41% (15/37) of the studies used diverse populations by separating them into people with diabetes and people without diabetes. Study design features of each included study are shown in Multimedia Appendix 5.

Table 2. Study design features (n=37).

Features	Studies, n (%)	Study ID
Study aim		
Blood glucose estimation (predictions)	10 (27)	S3, S18, S19, S21, S23, S25, S27, S28, S32, S33, S37
Glucose level monitoring	10 (27)	S7-S11, S15, S20, S24, S26, S30
Diagnostic solution	5 (14)	S4, S29, S33, S34, S35
Diabetes classification	4 (11)	S12-S14, S17
Self-administration and monitoring	4 (11)	S1, S5, S6, S31
Prevention	2 (5)	S2, S16
Other disease predictions, detection, and monitoring (hypoglycemia and foot temperature)	2 (5)	S22, S36
Privacy and security		
Not mentioned	31 (84)	S1-S22, S24-S26, S29, S30, S34-S37
Mentioned	6 (16)	S23, S27, S28, S31, S32, S33
Data source		
Private	25 (68)	S2, S3, S5, S7, S8-S19, S21, S22, S24, S26, S29, S34-S37
Public	4 (11)	S1, S4, S6, S25
Not mentioned	2 (5)	S20, S30
Participant demographics		
Age group (years)[a]		
Children and young adults (≤18)	1 (3)	S8
Adult (19-65)	18 (49)	S2-S5, S8, S10, S13, S15, S16, S17, S19, S21, S22, S27, S29, S31, S33, S34
Older adult (>65)	6 (16)	S2, S4, S15, S21, S22, S33
Not mentioned	19 (51)	S1, S6, S7, S9, S11, S12, S14, S18, S20, S23-S26, S28, S30, S32, S35-S37
Gender		
Male	10 (27)	S2, S3, S5, S13, S15, S17, S18, S27, S29, S34
Female	10 (27)	S2, S3, S5, S13, S15, S17, S18, S27, S29, S34
Not mentioned	27 (73)	S1, S4, S6-S12, S14, S16, S19-S26, S28, S30-S33, S35-S37
Diabetes[b]		
Yes	14 (38)	S1, S4, S5-S7, S10, S12, S14, S15, S18, S19, S21, S27, S34, S36
No	15 (41)	S1, S5, S6, S8-S10, S12, S14, S18, S19, S27, S29, S31, S33, S36
Not mentioned	17 (46)	S2, S3, S11, S13, S16, S17, S20, S22-S26, S29, S30, S32, S35, S37

[a] Numbers do not add up as participants in some studies belong to more than one age group.

[b] Numbers do not add up as participants in some studies were diabetic and nondiabetic.

Features of WDs

General Features of Wearables

Table 3 highlights the general features of WDs; some studies used multiple devices. Only 41% (15/37) of the studies used commercially available WDs, whereas 59% (22/37) used prototypes. Most of the studies (22/37, 59%) included wrist-worn devices. In most of the studies, the device type was in the form of a wearable sensor (14/37, 38%), followed by smartwatch (8/37, 22%) and smart wristband (9/37, 24%). Only one study reported a smart sock and another reported smart clothes. Among the developed wearable technologies used, Empatica E4 was the most cited (6/37, 16%), followed by Glutrac (3/37, 8%). For OSs, most of the studies reported devices either directly or indirectly compatible with iPhone OS and Android OS 43% (16/37); OS was not mentioned in a large number of studies (11/37, 28%), and 8% (3/37) mentioned Android only and 5% (2/37) mentioned iPhone OS only. For gateway (ie, a hardware that acts as a “gate” between 2 networks or any device that enables traffic to flow in and out of the network), many of the studies did not mention any sort of gateway (17/37, 46%). Most of the studies that mentioned a gateway used a smartphone (16/37, 43%). Host devices (devices where the actual manipulation of collected data was performed, ie, processing) were used in many of the studies, the most popular being cloud-based (18/37, 49%), many did not report any use of a host device (8/37, 22%), and smart devices were mentioned in 16% (6/37) of studies. For the purpose of transferring data (mode of data transfer) from the WD, the majority of devices reported Bluetooth as the means of transfer (19/37, 51%); 19% (7/37) of studies did not mention the mode of transfer. A total of 16% (6/37) of studies reported the use of some form of internet connection as the mode of transfer (ie, Wi-Fi or mobile network). Features of WDs for each included study are shown in Multimedia Appendix 5.

Table 3. General features of wearable devices (n=37).

Features	Studies, n (%)	Study ID
Technology status		
Prototype	22 (59)	S1, S3-S5, S8-S11, S16, S17, S20, S23, S24, S26, S28-S33, S36, S37
Commercial	15 (41)	S2, S6, S7, S12-S15, S18, S19, S21, S22, S25, S27, S34, S35
Device type		
Smart clothes	1 (3)	S1
Smart socks	1 (3)	S31
Smart watch	8 (22)	S2, S7, S14, S15, S18, S19, S28, S35
Smart watch and wearable sensor	2 (5)	S21, S24
Smart wristband	9 (24)	S4, S6, S12, S13, S25, S27, S30, S33, S34
Smart wristband, smart footwear, and smart neckband	2 (5)	S23, S32
Wearable sensor	14 (38)	S3, S5, S8-S11, S16, S17, S20, S22, S26, S36, S37
Placement		
Body	1 (3)	S1
Chest	1 (3)	S11
Finger	5 (14)	S3, S8, S17, S20, S26
Foot	6 (16)	S5, S9, S16, S29, S31, S36
Hand	1 (3)	S10
Wrist	18 (49)	S4, S6, S7, S12-S15, S18, S19, S24, S25, S27, S28, S30, S33-S35, S37
Wrist and arm	1 (3)	S21
Wrist or thigh	1 (3)	S2
Wrist, foot, and neck	2 (5)	S23, S32
Arm and body	1 (3)	S22
Device technology[a]		
Actigraph	1 (3)	S21
Arduino Nano	1 (3)	S24
Basis Peak	1 (3)	S34
FreeStyle Libre Flash	2 (5)	S22, S35
Medtronic Zephyr	1 (3)	S22
Dexcom G4 Platinum (Professional)	1 (3)	S21
Empatica E4	6 (16)	S12, S13, S14, S25, S27, S35
Glutrac	3 (8)	S15, S18, S19
Mi band 2	1 (3)	S6
Raspberry Pi Zero	2 (5)	S8, S16
Pebble	1 (3)	S2
Custom	2 (5)	S26, S28
Not mentioned	18 (49)	S1, S3-S5, S7, S9-S11, S17, S20, S23, S29-S33, S36, S37
Operating system[b]		
Android	3 (8)	S2, S8, S16
iOS[c]	2 (5)	S9, S11
Microsoft	1 (3)	S31

Features	Studies, n (%)	Study ID
Raspberry Pi OS[d]	1 (3)	S24
iOS and Android	16 (43)	S6, S7, S12-S15, S17-S20, S22, S23, S26, S28, S30, S32
Any desktop OS	3 (8)	S25, S27, S29
Any smartphone OS	1 (3)	S29
Not mentioned	11 (30)	S1, S3-S5, S10, S21, S33-S37
Gateway		
Smartphone	16 (43)	S1, S6, S7, S11-S15, S17-S20, S23, S28, S30, S32
Database servers (Hbase and Hadoop or Spark)	1 (3)	S33
Adapter	1 (3)	S4
Smartphone or PC	2 (5)	S25, S27
None	17 (46)	S2, S3, S5, S8-S10, S16, S21, S22, S24, S26, S29, S31, S34-S37
Host device		
Cloud (MongoDb, Database server, Google)	18 (49)	S1, S6, S7, S11-S15, S17-S19, S23, S25, S27, S28, S30, S32, S33
PC (laptop, desktop, or Microsoft Surface)	4 (11)	S4, S20, S29, S31
Raspberry Pi	1 (3)	S24
Smart devices (smartphone, tablet, or PC)	6 (16)	S5, S8, S9, S16, S22, S26
None	8 (22)	S2, S3, S10, S21, S34-S37
Mode of data transfer		
Bluetooth	19 (51)	S2, S5, S6, S9, S11-S15, S18-S20, S22, S25-S28, S30, S31
Internet (Wi-Fi or cellular or mobile network)	6 (16)	S1, S7, S8, S16, S17, S33
Internet (Wi-Fi or cellular or mobile network) and Bluetooth	2 (5)	S23, S32
Wired	2 (5)	S24, S29
Removable media	1 (3)	S4
N/A[e]	7 (19)	S3, S10, S21, S34-S37

[a] Numbers do not add up as some studies used more than one wearable device.

[b] Numbers do not add up as the WD in one study worked on 2 operating systems.

[c] iOS: iPhone operating system.

[d] OS: operating system.

[e] N/A: not applicable.

Technical Features of Wearables

Table 4 shows an overview of the technical features of the WDs associated with the studies in this review. Devices were often reported as having >1 device measure, and we reported these primary measures along with the measurements used for the respective studies. We observed that many studies reported blood glucose (15/37, 41%) followed by temperature (10/37, 27%), heart rate (9/37, 24%) and galvanic skin response (9/37, 24%) as their top primary device measures. Regarding the second feature shown in Table 4, the majority of the studies reported on blood glucose as the main measurement studied (27/37, 73%), followed by heart rate or variability (4/37, 11%).

Most of the studies (28/37, 76%) reported an opportunistic approach (ie, no input required from the participant) when obtaining data using the WDs, whereas the remaining (9/37, 24%) used a participatory approach (ie, input required from the participants). For sensing technologies, various sensors were used, either built-in to the WD or as wearable sensors, often reported as >1 sensor per device. We observed a large number of devices in the studies reviewed reporting photoplethysmography sensor use (12/37, 32%), while optical heart rate was only seen in 5% (2/37) of studies among some of the other less-reported sensors. Features of WDs for each included study are shown in Multimedia Appendix 5.

Table 4. Technical features of wearables (n=37).

Feature	Studies, n (%)	Study ID
Primary device measure[a]		
Blood glucose	15 (41)	S3, S8, S10, S15, S17-S22, S24, S26, S28, S30, S37
Physiological	2 (5)	S1, S10
Heart rate, heart rate variability, or interbeat interval of the heart	9 (24)	S6, S11, S14, S22, S23, S32-S35
Galvanic skin response	9 (24)	S12-S14, S23, S25, S27, S32, S34, S35
Blood volume pulse	6 (16)	S12-S14, S25, S27, S35
Acceleration	6 (16)	S12-S14, S25, S27, S35
Plantar pressure	5 (14)	S5, S9, S23, S29, S32, S33
Temperature (skin, foot, shoe, air, or ambient)	10 (27)	S12, S13, S16, S23, S25, S27, S32, S34-S36
Step count	2 (5)	S7, S16
Other (sedentary behaviors, pulse wave information, inertial data, weight, humidity, activity patterns, frequency of food intake and water, and ankle edema quantification)	8 (22)	S2, S4, S9, S16, S21, S23, S31, S32
Measurement studied[b]		
Blood glucose	27 (73)	S1, S3, S6-S8, S10-S28, S30, S33, S37
Plantar pressure	3 (8)	S5, S9, S29
Heart rate or heart rate variability	4 (11)	S28, S33, S34, S35
Other (sedentary behavior, pulse wave, edema, general diabetes symptoms, temperature, sleep quality, step counts, and GSR)	7 (19)	S2, S4, S31, S32, S34-S36
Sensing approach		
Opportunistic	28 (76)	S1, S2, S5, S7, S11-S14, S16-S29, S31-S33, S35-S37
Participatory	9 (24)	S3, S4, S6, S8-S10, S15, S30, S34
Sensing technology[c]		
Accelerometer	5 (14)	S2, S13, S14, S21, S27
Photoplethysmography	12 (32)	S3, S10, S12-S15, S19, S20, S24, S25, S27, S28
Galvanic skin response	8 (22)	S10, S13, S14, S23, S24, S27, S32, S34
Near infrared	5 (14)	S3, S17, S18, S28, S37
Electrocardiography	3 (8)	S11, S18, S22
Continuous glucose monitoring	2 (5)	S21, S22
Bluetooth	1 (3)	S6
Pressure sensors	7 (19)	S5, S9, S23, S29, S32, S33, S36
Infrared thermopile	3 (8)	S13, S14, S27
Temperature sensor	6 (16)	S7, S16, S23, S24, S32, S36
Optical heart rate sensor	2 (5)	S23, S32
Vibration sensor and flex sensor	2 (5)	S23, S32
Motion sensor	2 (5)	S7, S31
Others (physiological sensors, pulse sensor, blood glucose level sensor, Raspberry Pi camera, humidity sensor, step count sensor, weight sensor, stretch sensor, and optical sensor)	6 (16)	S1, S4, S7, S8, S16, S31

[a] Numbers do not add up as WDs in many studies were used to measures many biomarkers.

[b] Numbers do not add up as some studies used more than one measure.

[c] Numbers do not add up as WDs in most studies used more than one sensor.

Wearables Characteristics With Regard to Diabetes Measurements

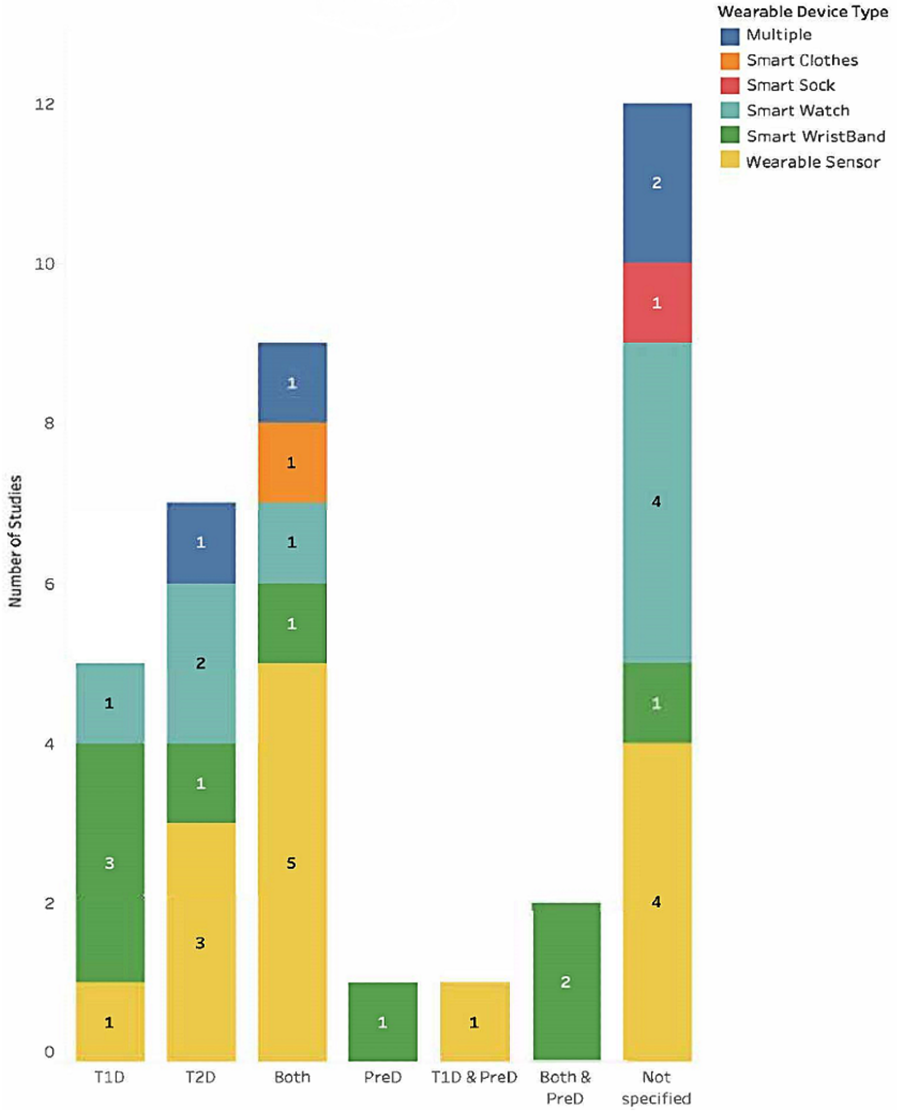
Wearable Technology Status Versus WD Type

Multimedia Appendix 6 further visualizes the data highlighting the WD type and whether they are commercial or prototypes. Wearable sensors were the most prominent as a prototype while smartwatches and smart wristbands were the most common as commercial.

Diabetes Types Versus WDs

Figure 2 shows the type of diabetes and number of studies related to each WD type. While most studies did not specify the type, T1D (as a smart wristband), T2D (as a wearable sensor), or both (as a wearable sensor or smartwatch) seem to be the most targeted types.

Figure 2. Diabetes type with regards to wearable device type. PreD: prediabetes; T1D: type 1 diabetes; T2D: type 2 diabetes.



AI and ML Technologies

For the purpose of this study, we categorized the ML algorithms into 4 categories (classification models, regression models, neural network–based models, and optimization algorithms) and those that were not clearly specified by the study authors were categorized as black boxes (ie, studies that mention they make use of ML or AI but do not specify any further details of algorithms used). Many ML technologies were reported that come under these headings (refer to Table 5 for a full list), and some studies reported and compared >1 model. Support vector machine (SVM) was the most reported (13/37, 35%), followed by random forest (12/37, 32%), k-nearest neighbor (7/37, 19%), Naive Bayes (5/37, 14%), and decision trees (4/37, 11%) among the most used models from classification models. From the regression models, only linear regression (2/37, 5%) was reported in a couple of studies, whereas all others were reported by single studies only. Artificial neural networks were reported in 14% (5/37) of the studies in neural network–based models, followed by long short-term memory (4/37, 11%), convolutional neural networks (3/37, 8%), and deep neural networks (3/37, 8%); these networks were used for both classification and regression purposes. Table 5 also highlights that the majority of the studies applied the AI and ML technologies for either the purpose of blood glucose level forecasting (12/37, 32%) or classifying the participants as normal, diabetic, or prediabetic

(12/37, 32%). Table 6 highlights some of the statistical measures used to evaluate the ML algorithms within the reported studies. Some studies used multiple statistical techniques for this purpose, among them were reports of accuracy (20/37, 54%) and sensitivity (9/37, 24%). While some studies did not mention which was the best ML model identified (6/37, 16%), random forest was reported as the best identified model (7/37, 19%), followed by SVM (6/37, 16%).

Table 5. Artificial intelligence (AI)– and machine learning (ML)–related features (n=37).

Features	Studies, n (%)	Study ID
AI or ML technologies used[a]		
Classification models		
Support vector machine	13 (35)	S1, S2, S4, S5, S9, S12, S13, S25, S29, S30, S33, S34, S36
Random forest	12 (32)	S2, S4, S5, S7, S11, S14, S15, S18, S27, S29, S36, S37
K-nearest neighbors	7 (19)	S5, S9, S12, S13, S25, S29, S31
Naive Bayes	5 (14)	S2, S7, S13, S31, S36
Decision tree	4 (11)	S1, S13, S31, S35
Ensemble learning or ensemble—boosted trees	2 (5)	S1, S13
Logistic regression	2 (5)	S2, S11
J48	2 (5)	S2, S7
Linear discriminant analysis or linear discriminant	2 (5)	S4, S13
Gradient boosting decision trees	2 (5)	S5, S35
AdaBoost classifier	1 (3)	S5
ZeroR	1 (3)	S7
OneR	1 (3)	S7
Simple logistic regression	1 (3)	S7
Gaussian Process classifier	1 (3)	S29
C4.5	1 (3)	S33
Linear ridge Classifier	1 (3)	S14
Extreme gradient boost	1 (3)	S12
Regression models		
Linear regression	2 (5)	S3, S16
Support vector regression or Fine Gaussian support vector regression	1 (3)	S3
Random Forest regression	1 (3)	S15
AdaBoost regression	1 (3)	S15
Multilayer Polynomial regression	1 (3)	S17
Ensemble—boosted trees	1 (3)	S3
Exponential Gaussian process regression	1 (3)	S20
Neural network–based models		
Artificial Neural Network	5 (14)	S1, S2, S8, S26, S36
Long short-term memory	4 (11)	S6, S13, S21, S34
Convolutional Neural Network	3 (8)	S10, S22, S24
Deep neural networks	3 (8)	S11, S13, S22
Recurrent Neural Network	2 (5)	S21, S34
Multilayer Perceptron	2 (5)	S6, S29
Optimization algorithm		
Sequential minimal optimization	1 (3)	S7
L1 norm optimization	1 (3)	S19
Particle swarm optimization	1 (3)	S23
ML[a] black box	3 (8)	S19, S23, S32
Application of AI technology used[b]		

Features	Studies, n (%)	Study ID
Blood glucose level forecasting	12	S6, S8, S16, S18, S20, S22, S24, S25-28, S34
Blood glucose monitoring	4	11, S30, S32, S37
Classify patients with diabetes (normal, diabetic, and prediabetic)	12	S3, S4, S5, S6, S7, S12, S14, S21, S23, S29, S32, S36
Classify other diseases (patients with hypertension or hypoglycemia)	2	S33, S35
Evaluation of a developed system	3	S2, S10, S13
Feature selection	2	S3, S5
Performance validation	3	S1, S9, S15
Optimize sensors results	3	S16, S17, S19
Predictions for step count, shoe removal time, or serum glucose	2	S16, S17
Edema monitoring	1	S31

[a] Numbers do not add up as most studies developed more than one AI algorithms.

[b] Numbers do not add up as AI algorithms in some studies were used for more than one application.

Table 6. Statistical evaluation of artificial intelligence and machine learning algorithm (n=37).

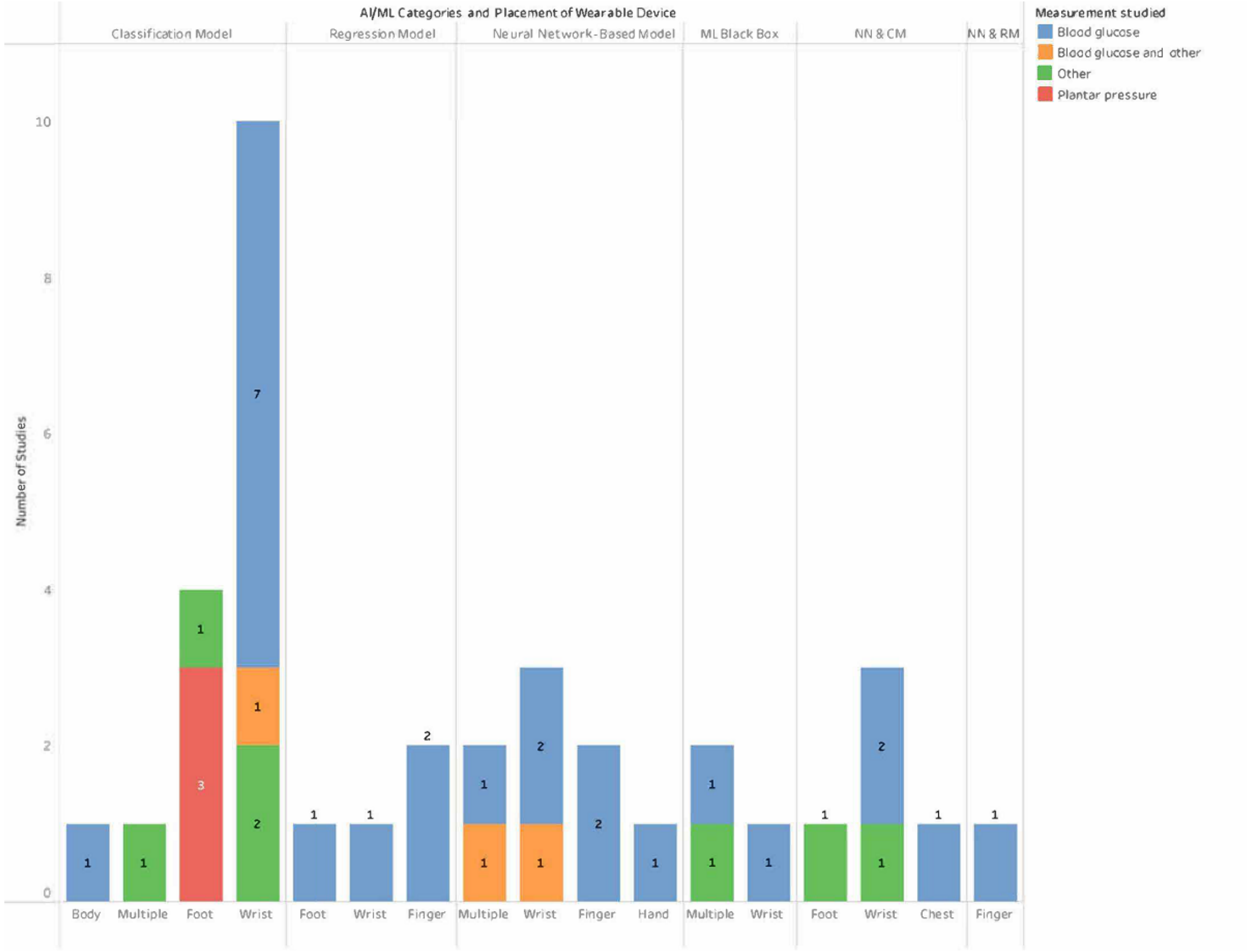
Characteristic	Value	Study ID
Accuracy (%; n=20)		
	≤80	S6, S33
	81-90	S15, S21, S28, S35, S36
	91-95	S1, S9, S13, S15, S22, S25, S29
	>95	S4, S5, S7, S12, S14, S30, S31
Sensitivity (%; n=9)		
	≤80	S35
	81-90	S4, S6, S25, S33
	91-95	S9, S22
	>95	S5, S7
Specificity (%; n=7)		
	≤85	S35
	86-90	S9, S22
	91-95	S5, S25
	>95	S4, S7
Area under the curve (%; n=2)		
	≤91	S22
	>91	S35
Clarke Error Grid zone A (%; n=8)		
	≤74	S37
	75-80	S19, S10
	81-90	S18, S28
	>90	S3, S8
	Not mentioned	S24
Precision (%; n=6)		
	≤80	S6, S33
	81-90	S9
	91-95	S25
	>95	S2, S7
Root mean square error (%; n=4)		
	<5	S19, S21
	5-15	S17
	>15	S27
Others, n (%)		
	8 (22)	S3, S8, S16, S17, S19, S21, S27, S37
Best model identified, n (%)		
Artificial Neural Network	2 (5)	S8, S26
Convolutional Neural Network	3 (8)	S10, S22, S24
Deep Neural Networks	4 (11)	S14, S17, S21, S28
Support Vector Machine	6 (16)	S4, S9, S25, S29, S30, S33
Random Forest	7 (19)	S2, S5, S15, S18, S27, S36, S37
Long Short-Term Memory	1 (3)	S13

Characteristic	Value	Study ID
Decision Trees or Gradient Boosting Decision Trees	2 (5)	S31, S35
K-Nearest Neighbors	1 (3)	S31
Multilayer Perceptron	1 (3)	S6
OneR	1 (3)	S7
Ensemble	1 (3)	S1
Support Vector Regression	1 (3)	S3
Not mentioned	6 (16)	S11, S19, S20, S23, S32, S34

AI and ML Versus Wearables Versus Diabetes

Figure 3 shows the category of the ML algorithm used according to each WD placement and measurement. Most devices that made use of classification models among the wrist-worn devices were the most prominent. Neural network and regression model were the least used.

Figure 3. Artificial intelligence (AI) or machine learning (ML) models used with regard to wearable device placement and measurement studied. CM: classification model; NN: neural network; RM: regression model.



Discussion

Principal Findings

This was the first study of its kind to the best of our knowledge, considering the amount of features we were able to extract from each publication. The features extracted should give researchers insight not only into the technologies that are readily available commercially but also into what is possible in the future with studies we identified that developed prototypes. Our findings shed light on this emerging field, which is still in its infancy. This is further highlighted by the fact that 59% (22/37) of the studies that met our inclusion criteria were prototypes; we were only able to identify 41% as commercially available (as demonstrated in Multimedia Appendix 6) devices, of which only (7/15, 46%) studies performed some sort of ML classification on the extracted data directly from WDs, whereas (6/15, 40%) studies made use of neural network–based models with classification to make out of already collected data. Most of these measured blood glucose on wrist-worn devices and used a classification algorithm (Figure 3). Classification models were widely used (Figure 3) in the reviewed studies, largely owing to studies attempting to classify types of diabetes (T1D, T2D, etc). SVM and random forest were the most prevalent classifiers and exhibited the highest performance. SVM [55] is extensively used because of its superiority in generalization and nonlinear function fitting, and it also has a number of advantages when dealing with small-sample studies [56]. Furthermore, SVM is a binary classifier, and we observed that it is mostly used on blood glucose level data to determine levels for diabetes categorization. Aside from the accuracy for demonstrating efficacy, the Clarke Error Grid was the most commonly used performance metric, possibly because of its popularity as a performance metric for assessing blood glucose estimation. The grid was split into 5 zones, each with varied prediction accuracy between the estimated and reference blood glucose readings. The data fell within zone A, which pertains to precise glucose calculations, where each consecutive zone is thought to have progressively substantial erroneous estimations [57,58]. Most of the sensory data being collected especially when looking at commercially available devices, did not require any further or minimal input from the user, meaning the person with diabetes can get on with day-to-day tasks without having to worry about taking regular invasive finger pricks for monitoring glucose levels; for example, while still feeling that they are wearing a stylish item such as a smartwatch. We specifically examined studies after 2015, as previous studies related to the use of WDs found that most wearables were used in this range [59]. One of the reasons may be that Fitbit released its first device in 2009 and the Apple Watch followed in 2015; both these devices set the tone for WDs, and it is not surprising that 59% (22/37) from our review were wrist-worn. A total of 78% (29/37) of the devices were connected to either a gateway or host device, usually a smartphone (16/37, 43%) via either Bluetooth (19/37, 51%) or Wi-Fi or internet (6/37, 16%); this is likely owing to the fact that web-based data are now more affordable and the availability of low energy connectivity technology such as Bluetooth. This ability to connect has resulted in more analytics and data storage being possible on host devices than on smartphones or directly on the cloud (18/37, 49%) of studies in this review, compared with limited computing power on the WD itself. One of the limitations of this is that devices need continuous connections, which can be an issue, as reported data can be lost if the connection is not maintained for long periods [60]. We also observed that many devices used gateways or host devices, which we believe to be largely because of the limited computing power of WDs.

Strengths

This review was conducted according to the PRISMA-ScR; therefore, it can be considered a high standard. Two reviewers independently conducted the study selection and data extraction. We believe this to be the first of its kind study focusing on WDs targeting diabetes using AI approaches and were unable to identify previous scoping reviews in the literature that has as an exhaustive list of features extracted in this field. A combination of expert research computer scientists and research medical practitioners allowed us to explore the current technologies in depth and highlight gaps in the research community. The most popular databases in the health care and information technology fields were searched; furthermore, Google Scholar with forward and backward reference list checking allowed an exhaustive search of the literature, reducing the risk of publication bias.

Limitations

Only studies published between 2015 and 2021 in the English language were included. Furthermore, we did not use Medical Subject Headings terms in our search; therefore, we may have overlooked some relevant studies. We excluded devices that could be classified as WDs, such as electroencephalogram and ECG machines, which limited their use in hospital settings. As our focus was AI, we excluded any study of WDs and diabetes that had a statistical measurement not considered an AI approach. Although we included a large number of features and some effectiveness measures, we fall short of critically assessing the quality of each of the included studies—this goes beyond the scope of our review—and we hope to cover this in a full systematic review in the near future on the same topic.

Practical and Research Implications

WDs hold great potential for the self-monitoring of diabetes-related parameters, and their ability to be paired with a range of smart devices, including smartphones and general connectivity to clouds, allows the continuous collection of data from many biosensors that measure vitals and biosignals without user interference. The fact that they can be worn in a stylish and fashionable manner has potential for wider acceptance than other technologies, such as CGMs. Although many studies have used WDs for diabetes, we found that ML is still lacking in a sizable number of these studies. With the limited number of studies that reported the use of ML, we see great promise, largely owing to the accuracy levels of the ML algorithms reported in Table 6. Engineering and data science research experts need to come together and identify the most common sensors and technologies and study their effectiveness when combined with ML approaches. In addition, commercially available WDs are readily available and therefore sit in waiting for researchers to conduct studies and apply ML and report further in scientific journals to prove validity and instill consumer confidence. Most of the papers identified in this study used AI or ML algorithms for testing the validity of the system functioning rather than identifying the approaches that could be used for the development of such intelligent devices. More work needs to concentrate on applying known ML algorithms for the purpose of making more accurate diabetes-related measurement calculations. Currently, the number of commercial devices associated with studies are still very low, a quick search on retail sites such as Amazon reveals many commercial devices claiming diabetes-related measurements, which have still yet to be validated with related studies, and this is one area where researchers could get involved. Researchers need to make more use of purpose-shifted devices as they are lying in wait as opposed to creating prototypes and testing the effectiveness of the many commercially available devices. We encourage researchers to perform systematic reviews to assess the efficacy of AI-based and non–AI-based WDs compared with traditional medical devices. Some technologies that are classified as WDs such as CGMs are still classified as semi-invasive as they allow the embedding of a sensor partially into participants' skin, we feel for wider acceptance especially for home use products the technology really needs to move away from such sensors and more studies now need to focus on how measurements can be obtained from noninvasive sensors such as those available on commercial smart watches. Further work is also required on ML algorithms used for diabetes data that can be used on the WDs as opposed to on host devices, as this would reduce some issues reported such as loss of data owing to WD out of range with the host device, which will become easier with time as the technology advances and WD memories are no longer a limitation. We suspect there would be less reliance on host devices for some of the ML computations.

Another area for exploration is the use of the internet of things (IoT); in our search, we found a handful of studies making use of IoT. Most IoT papers describe the IoT architecture for diabetes management without specifying the sensors or WDs actually used or implemented, and do not go into much (if any) detail about any ML deployed. There are many opportunities in this domain; none of the studies were found to make good use of developed commercial technologies such as Alexa, Google Home, and Apple watches, which are readily available. The possibilities here are endless, using a combination of data gathered from sensors at the WDs with other patients and personal data in real-time with IoT. This brings along with its own caveats and the need to incorporate questions of privacy and data sovereignty arising from the mass data storage in cloud-based systems and the many interconnected devices and hospital datacenters; there are issues that need to be considered with the use of data and individual consent. There are also problems regarding the scope of an individual's consent to use their data, as well as potential accountability if the data are mishandled. There are dangers associated with AI algorithms and their misdiagnoses, dangerous advice, or recommendations that do not correspond to the required standard of care. Data security breaches or the reidentification of previously deidentified data may have unintended repercussions. Furthermore, other ethical issues need to be considered, such as accessibility, although commercial WDs that are easily and cheaply available may not be affordable for the masses in low-income countries. A multidisciplinary effort is required, including but not limited to engineers, medical practitioners, and legal experts.

Conclusions

We investigated and reported the current state of WDs and their features for the purpose of diabetes that use ML approaches. Considering the availability of consumer-grade biosensors, we see great advancement potential in this domain, replacing hospital setting, invasive devices, especially when it comes to monitoring glucose levels. Further clinically significant studies are needed to instill confidence and validate WD use as well as the application of ML algorithms on WD data. Researchers and those wanting to develop AI-based WDs can use our review to understand where the gaps are in this emerging field. We encourage readers to use more data and delve deeper into the studies we have identified in order to establish, validate, and repeat studies that showed high accuracy. There is still much work needed, and we feel our review has provided the most extensive work so far summarizing WDs that use ML for people with diabetes to date. Finally, researchers will also benefit from our study as they can embark on longer and better populated systematic studies scrutinizing the benefits of WDs as data gathering, monitoring, prediction, classification, and recommendation devices in the context of diabetes. We envisage several follow-up studies, starting with a full systematic review from our own group.

Acknowledgments

The authors would like to thank Weill Cornell Medicine-Qatar for making this study possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 84 KB - jmir_v24i8e36010_app1.docx]

Multimedia Appendix 2
Full search terms and strings table.
[DOCX File , 14 KB - jmir_v24i8e36010_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File , 17 KB - jmir_v24i8e36010_app3.docx]

Multimedia Appendix 4
Study reference table.

[DOCX File , 33 KB - jmir_v24i8e36010_app4.docx]

Multimedia Appendix 5
Data extraction sheet.
[XLSX File (Microsoft Excel File), 25 KB - jmir_v24i8e36010_app5.xlsx]

Multimedia Appendix 6
Wearable technologies status with regard to wearable device technologies.
[PNG File , 93 KB - jmir_v24i8e36010_app6.png]

References

1. Global report on diabetes. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf) [accessed 2021-11-02]
2. Sims EK, Carr AL, Oram RA, DiMeglio LA, Evans-Molina C. 100 years of insulin: celebrating the past, present and future of diabetes therapy. Nat Med 2021 Jul 15;27(7):1154-1164 [FREE Full text] [doi: 10.1038/s41591-021-01418-2] [Medline: 34267380]
3. Wang L, Li X, Wang Z, Bancks MP, Carnethon MR, Greenland P, et al. Trends in prevalence of diabetes and control of risk factors in diabetes among US adults, 1999-2018. JAMA 2021 Jun 25;326(8):704 [FREE Full text] [doi: 10.1001/jama.2021.9883] [Medline: 34170288]
4. About diabetes. International Diabetes Federation. URL: https://tinyurl.com/349dbcau [accessed 2022-07-20]
5. Alves S, Matos AG, Almeida JS, Benevides CA, Cunha CC, Santiago RV, et al. A new strategy for the detection of diabetic retinopathy using a smartphone app and machine learning methods embedded on cloud computer. In: Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS). 2020 Presented at: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); Jul 28-30, 2020; Rochester, MN, USA. [doi: 10.1109/cbms49503.2020.00108]
6. Chakravadhanula K. A smartphone-based test and predictive models for rapid, non-invasive, and point-of-care monitoring of ocular and cardiovascular complications related to diabetes. Informatics Med Unlocked 2021;24:100485. [doi: 10.1016/j.imu.2020.100485]
7. Ramazi R, Perndorfer C, Soriano E, Laurenceau JP, Beheshti R. Multi-modal predictive models of diabetes progression. In: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. 2019 Presented at: BCB '19: 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; Sep 7 - 10, 2019; Niagara Falls NY USA. [doi: 10.1145/3307339.3342177]
8. Prabha A, Yadav J, Rani A, Singh V. Non-invasive diabetes mellitus detection system using machine learning techniques. In: Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). 2021 Presented at: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence); Jan 28-29, 2021; Noida, India. [doi: 10.1109/confluence51648.2021.9377138]
9. Calderaro J, Seraphin TP, Luedde T, Simon TG. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. J Hepatol 2022 Jun;76(6):1348-1361. [doi: 10.1016/j.jhep.2022.01.014] [Medline: 35589255]
10. Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan GS, Abramoff M, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Lond) 2020 Mar 05;34(3):451-460 [FREE Full text] [doi: 10.1038/s41433-019-0566-0] [Medline: 31488886]
11. Ellahham S. Artificial intelligence: the future for diabetes care. Am J Med 2020 Aug;133(8):895-900. [doi: 10.1016/j.amjmed.2020.03.033] [Medline: 32325045]
12. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
13. Islam M, Manjur SM. Design and implementation of a wearable system for non-invasive glucose level monitoring. In: Proceedings of the 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON). 2019 Presented at: 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON); Nov 28-30, 2019; Dhaka, Bangladesh. [doi: 10.1109/becithcon48839.2019.9063193]
14. Lee E, Lee C. Ppg-based smart wearable device with energy-efficient computing for mobile health-care applications. IEEE Sensors J 2021 Jun 15;21(12):13564-13573. [doi: 10.1109/jsen.2021.3069460]
15. Xie P, Li Y, Deng B, Du C, Rui S, Deng W, et al. An explainable machine learning model for predicting in-hospital amputation rate of patients with diabetic foot ulcer. Int Wound J 2022 May;19(4):910-918 [FREE Full text] [doi: 10.1111/iwj.13691] [Medline: 34520110]
16. Daley BJ, Ni'Man M, Neves MR, Bobby Huda MS, Marsh W, Fenton NE, et al. mHealth apps for gestational diabetes mellitus that provide clinical decision support or artificial intelligence: a scoping review. Diabet Med 2022 Jan 16;39(1):e14735. [doi: 10.1111/dme.14735] [Medline: 34726798]

17. Rodriguez-León C, Villalonga C, Munoz-Torres M, Ruiz JR, Banos O. Mobile and wearable technology for the monitoring of diabetes-related parameters: systematic review. JMIR Mhealth Uhealth 2021 Jun 03;9(6):e25138 [FREE Full text] [doi: 10.2196/25138] [Medline: 34081010]
18. Tan CH, Kyaw BM, Smith H, Tan CS, Tudor Car L. Use of smartphones to detect diabetic retinopathy: scoping review and meta-analysis of diagnostic test accuracy studies. J Med Internet Res 2020 May 15;22(5):e16658 [FREE Full text] [doi: 10.2196/16658] [Medline: 32347810]
19. Fagherazzi G, Ravaud P. Digital diabetes: perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
20. 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]
21. 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]
22. Chen M, Yang J, Zhou J, Hao Y, Zhang J, Youn C. 5G-smart diabetes: toward personalized diabetes diagnosis with healthcare big data clouds. IEEE Commun Mag 2018 Apr;56(4):16-23. [doi: 10.1109/mcom.2018.1700788]
23. Ceron JD, Lopez DM, Ramirez GA. A mobile system for sedentary behaviors classification based on accelerometer and location data. Comput Industry 2017 Nov;92-93:25-31. [doi: 10.1016/j.compind.2017.06.005]
24. Hina A, Saadeh W. A noninvasive glucose monitoring SoC based on single wavelength photoplethysmography. IEEE Trans Biomed Circuits Syst 2020 Jun;14(3):504-515. [doi: 10.1109/tbcas.2020.2979514]
25. Hao Y, Cheng F, Pham M, Rein H, Patel D, Fang Y, et al. A noninvasive, economical, and instant-result method to diagnose and monitor type 2 diabetes using pulse wave: case-control study. JMIR Mhealth Uhealth 2019 Apr 23;7(4):e11959 [FREE Full text] [doi: 10.2196/11959] [Medline: 31012863]
26. Wang D, Ouyang J, Zhou P, Yan J, Shu L, Xu X. A novel low-cost wireless footwear system for monitoring diabetic foot patients. IEEE Trans Biomed Circuits Syst 2021 Feb;15(1):43-54. [doi: 10.1109/tbcas.2020.3043538]
27. Alfian G, Syafrudin M, Ijaz M, Syaekhoni M, Fitriyani N, Rhee J. A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing. Sensors (Basel) 2018 Jul 06;18(7):2183 [FREE Full text] [doi: 10.3390/s18072183] [Medline: 29986473]
28. Rghioui A, Lloret J, Harane M, Oumnad A. A smart glucose monitoring system for diabetic patient. Electronics 2020 Apr 22;9(4):678. [doi: 10.3390/electronics9040678]
29. Alarcón-Paredes A, Francisco-García V, Guzmán-Guzmán IP, Cantillo-Negrete J, Cuevas-Valencia RE, Alonso-Silverio GA. An IoT-based non-invasive glucose level monitoring system using raspberry Pi. Applied Sci 2019 Jul 28;9(15):3046. [doi: 10.3390/app9153046]
30. D'Angelantonio E, Lucangeli L, Camomilla V, Mari F, Mascia G, Pallotti A. Classification-based screening of phlebopathic patients using smart socks. In: Proceedings of the 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA). 2021 Presented at: 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA); Jun 23-25, 2021; Lausanne, Switzerland. [doi: 10.1109/memea52024.2021.9478688]
31. Gusev M, Guseva E, Poposka L. Design of a non-invasive ECG-based glucose measurement system. In: Proceedings of the 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO). 2020 Presented at: 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO); Sep 28 -Oct 02, 2020; Opatija, Croatia. [doi: 10.23919/mipro48935.2020.9245403]
32. Prabha A, Yadav J, Rani A, Singh V. Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier. Comput Biol Med 2021 Sep;136:104664. [doi: 10.1016/j.compbiomed.2021.104664] [Medline: 34329866]
33. Sevil M, Rashid M, Maloney Z, Hajizadeh I, Samadi S, Askari MR, et al. Determining physical activity characteristics from wristband data for use in automated insulin delivery systems. IEEE Sens J 2020 Nov 1;20(21):12859-12870 [FREE Full text] [doi: 10.1109/jsen.2020.3000772] [Medline: 33100923]
34. Yin H, Mukadam B, Dai X, Jha NK. DiabDeep: pervasive diabetes diagnosis based on wearable medical sensors and efficient neural networks. IEEE Trans Emerg Topics Comput 2021 Jul 1;9(3):1139-1150. [doi: 10.1109/tetc.2019.2958946]
35. Tsai C, Li C, Lam RW, Li C, Ho S. Diabetes care in motion: blood glucose estimation using wearable devices. IEEE Consumer Electron Mag 2020 Jan 1;9(1):30-34. [doi: 10.1109/mce.2019.2941461]
36. Kularathne N, Wijayathilaka U, Kottawaththa N, Hewakoralage S, Thelijjagoda S. Dia-Shoe: a smart diabetic shoe to monitor and prevent diabetic foot ulcers. In: Proceedings of the 2019 International Conference on Advancements in Computing (ICAC). 2019 Presented at: 2019 International Conference on Advancements in Computing (ICAC); Dec 05-07, 2019; Malabe, Sri Lanka. [doi: 10.1109/icac49085.2019.9103408]
37. Joshi AM, Jain P, Mohanty SP, Agrawal N. iGLU 2.0: a new wearable for accurate non-invasive continuous serum glucose measurement in IoMT framework. IEEE Trans Consumer Electron 2020 Nov;66(4):327-335. [doi: 10.1109/tce.2020.3011966]
38. Zhou X, Lei R, Ling B, Li C. Joint empirical mode decomposition and singular spectrum analysis based pre-processing method for wearable non-invasive blood glucose estimation. In: Proceedings of the 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP). 2019 Presented at: 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP); Jul 19-21, 2019; Wuxi, China. [doi: 10.1109/siprocess.2019.8868879]

39. Zhou X, Ling BW, Tian Z, Ho Y, Teo K. Joint empirical mode decomposition, exponential function estimation and norm approach for estimating mean value of photoplethysmogram and blood glucose level. IET Signal Process 2020 Dec;14(9):652-665. [doi: 10.1049/iet-spr.2020.0096]
40. Hina A, Nadeem H, Aslam AR, Saadeh W. Live demonstration: a single LED PPG-based noninvasive glucose monitoring prototype system. In: Proceedings of the 2019 IEEE International Symposium on Circuits and Systems (ISCAS). 2019 Presented at: 2019 IEEE International Symposium on Circuits and Systems (ISCAS); May 26-29, 2019; Sapporo, Japan. [doi: 10.1109/iscas.2019.8702241]
41. Porumb M, Griffen C, Hattersley J, Pecchia L. Nocturnal low glucose detection in healthy elderly from one-lead ECG using convolutional denoising autoencoders. Biomedical Signal Process Control 2020 Sep;62:102054. [doi: 10.1016/j.bspc.2020.102054]
42. Non-invasive analytics based smart system for diabetes monitoring. In: Internet of Things (IoT) Technologies for HealthCare. Cham: Springer; 2018.
43. Mahmud T, Limon MH, Ahmed S, Rafi MZ, Ahamed B, Nitol NN, et al. Non-invasive blood glucose estimation using multi-sensor based portable and wearable system. In: Proceedings of the 2019 IEEE Global Humanitarian Technology Conference (GHTC). 2019 Presented at: 2019 IEEE Global Humanitarian Technology Conference (GHTC); Oct 17-20, 2019; Seattle, WA, USA. [doi: 10.1109/ghtc46095.2019.9033119]
44. Kascheev N, Kozyrev O, Leykin M, Vanyagin A. Non-invasive monitoring of blood glucose by means of wearable tracking technology. In: Proceedings of the 2017 IEEE East-West Design & Test Symposium (EWDTS). 2017 Presented at: 2017 IEEE East-West Design & Test Symposium (EWDTS); Sep 29 -Oct 02, 2017; Novi Sad, Serbia. [doi: 10.1109/ewdts.2017.8110137]
45. Bent B, Cho PJ, Wittmann A, Thacker C, Muppidi S, Snyder M, et al. Non-invasive wearables for remote monitoring of HbA1c and glucose variability: proof of concept. BMJ Open Diab Res Care 2021 Jun 15;9(1):e002027. [doi: 10.1136/bmjdrc-2020-002027]
46. Corpin R, Guingab HA, Manalo AN, Sampana ML, Abello AN, dela Cruz AR, et al. Prediction of diabetic peripheral neuropathy (DPN) using plantar pressure analysis and learning models. In: Proceedings of the 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM). 2019 Presented at: 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM); Nov 29 - Dec 01, 2019; Laoag, Philippines. [doi: 10.1109/hnicem48295.2019.9072889]
47. Smart wristband with integrated chemical sensors for detecting glucose levels using breath volatile organic compounds. Proc SPIE. 2019. URL: https://tinyurl.com/2mva76z8 [accessed 2022-07-20]
48. Fallahzadeh R, Pedram M, Ghasemzadeh H. SmartSock: a wearable platform for context-aware assessment of ankle edema. In: Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 Presented at: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 16-20, 2016; Orlando, FL, USA. [doi: 10.1109/embc.2016.7592169]
49. Saravanan M, Shubha R, Marks AM, Iyer V. SMEAD: a secured mobile enabled assisting device for diabetics monitoring. In: Proceedings of the 2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS). 2017 Presented at: 2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS); Dec 17-20, 2017; Bhubaneswar, India. [doi: 10.1109/ants.2017.8384099]
50. Qaffas AA, Hoque R, Almazmomi N. The internet of things and big data analytics for chronic disease monitoring in Saudi Arabia. Telemedicine e-Health 2021 Jan 01;27(1):74-81. [doi: 10.1089/tmj.2019.0289]
51. Maritsch M. Towards wearable-based hypoglycemia detection and warning in diabetes. In: Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3334480.3382808]
52. Ravindran R, Niemann U, Klose S, Walter I, Ming A, Mertens PR, et al. Transformation of temperature timeseries into features that characterize patients with diabetic autonomic nerve disorder. In: Proceedings of the 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS). 2018 Presented at: 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS); Jun 18-21, 2018; Karlstad, Sweden. [doi: 10.1109/cbms.2018.00019]
53. Li Z, Ye Q, Guo Y, Tian Z, Ling B, Lam R. Wearable non-invasive blood glucose estimation via empirical mode decomposition based hierarchical multiresolution analysis and random forest. In: Proceedings of the 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP). 2018 Presented at: 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP); Nov 19-21, 2018; Shanghai, China. [doi: 10.1109/icdsp.2018.8631545]
54. Schwartz FL, Marling CR, Bunescu RC. The promise and perils of wearable physiological sensors for diabetes management. J Diabetes Sci Technol 2018 May;12(3):587-591 [FREE Full text] [doi: 10.1177/1932296818763228] [Medline: 29542348]
55. Vapnik V. The Nature of Statistical Learning Theory. New York: Springer; 1995.
56. Liu W, Wang J, Zhang C. Study of bioelectrical impedance analysis methods for visceral fat estimation using SVM. J Electron Measure Instrument 2011 Jul 20;25(7):648-653. [doi: 10.3724/sp.j.1187.2011.00648]
57. Comparison of noninvasive blood glucose estimation using various regression models. In: Advances in Signal Processing and Intelligent Recognition Systems. Singapore: Springer; 2020.

58. Clarke W, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987;10(5):622-628. [doi: 10.2337/diacare.10.5.622] [Medline: 3677983]
59. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 05;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
60. Schwartz FL, Marling CR, Bunescu RC. The promise and perils of wearable physiological sensors for diabetes management. J Diabetes Sci Technol 2018 May 15;12(3):587-591 [FREE Full text] [doi: 10.1177/1932296818763228] [Medline: 29542348]

Abbreviations

AI: artificial intelligence
CGM: continuous glucose monitoring
ECG: electrocardiogram
IoT: internet of things
ML: machine learning
OS: operating system
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SVM: support vector machine
T1D: type 1 diabetes
T2D: type 2 diabetes
WD: wearable device

Edited by T Leung; submitted 27.12.21; peer-reviewed by P Bhadra, J Sanders, G Al-Hussain, M Kraus; comments to author 23.05.22; revised version received 31.05.22; accepted 20.06.22; published 09.08.22.

Please cite as:
Ahmed A, Aziz S, Abd-alrazaq A, Farooq F, Sheikh J
Overview of Artificial Intelligence–Driven Wearable Devices for Diabetes: Scoping Review
J Med Internet Res 2022;24(8):e36010
URL: https://www.jmir.org/2022/8/e36010
doi:10.2196/36010
PMID:35943772

Review

Understanding Engagement in Digital Mental Health and Well-being Programs for Women in the Perinatal Period: Systematic Review Without Meta-analysis

Jacqueline A Davis[1,2,3], BSc, Grad Dip; Jeneva L Ohan[2], BSc, MA, PhD; Lisa Y Gibson[1,2,4], BA, MPsych, PhD; Susan L Prescott[1,2,5], BMedSci, MBBS, FRACP, PhD; Amy L Finlay-Jones[1,2], BPsych, MClinPsych, MHealthEcon, PhD

[1]Telethon Kids Institute, Nedlands, Australia
[2]School of Medicine, University of Western Australia, Perth, Australia
[3]School of Health Sciences, Curtin University, Bentley, Australia
[4]School of Medical and Health Sciences, Edith Cowan University, Joondalup, Australia
[5]The Nova Institute for Health, Baltimore, MD, United States

Corresponding Author:
Jacqueline A Davis, BSc, Grad Dip
Telethon Kids Institute
15 Hospital Avenue
Nedlands, 6009
Australia
Phone: 61 478173989
Email: jackie.davis@telethonkids.org.au

Abstract

Background: Pregnancy and the postnatal period can be a time of increased psychological distress, which can be detrimental to both the mother and the developing child. Digital interventions are cost-effective and accessible tools to support positive mental health in women during the perinatal period. Although studies report efficacy, a key concern regarding web-based interventions is the lack of engagement leading to drop out, lack of participation, or reduced potential intervention benefits.

Objective: This systematic review aimed to understand the reporting and levels of engagement in studies of digital psychological mental health or well-being interventions administered during the perinatal period. Specific objectives were to understand how studies report engagement across 4 domains specified in the Connect, Attend, Participate, and Enact (CAPE) model, make recommendations on best practices to report engagement in digital mental health interventions (DMHIs), and understand levels of engagement in intervention studies in this area. To maximize the utility of this systematic review, we intended to develop practical tools for public health use: to develop a logic model to reference the theory of change, evaluate the studies using the CAPE framework, and develop a guide for future data collection to enable consistent reporting in digital interventions.

Methods: This systematic review used the Cochrane Synthesis Without Meta-analysis reporting guidelines. This study aimed to identify studies reporting DMHIs delivered during the perinatal period in women with subclinical mood symptoms. A systematic database search was used to identify relevant papers using the Ovid Platform for MEDLINE, PsycINFO, EMBASE, Scopus, Web of Science, and Medical Subject Headings on Demand for all English-language articles published in the past 10 years.

Results: Searches generated a database of 3473 potentially eligible studies, with a final selection of 16 (0.46%) studies grouped by study design. Participant engagement was evaluated using the CAPE framework and comparable variables were described. All studies reported at least one engagement metric. However, the measures used were inconsistent, which may have contributed to the wide-ranging results. There was insufficient reporting for enactment (ie, participants' real-world use of intervention skills), with only 38% (6/16) of studies clearly recording longer-term practice through postintervention interviews. The logic model proposes ways of conceptualizing and reporting engagement details in DMHIs more consistently in the future.

Conclusions: The perinatal period is the optimal time to intervene with strength-based digital tools to build positive mental health. Despite the growing number of studies on digital interventions, few robustly explore engagement, and there is limited evidence of long-term skill use beyond the intervention period. Our results indicate variability in the reporting of both short- and long-term participant engagement behaviors, and we recommend the adoption of standardized reporting metrics in future digital interventions.

Trial Registration: PROSPERO CRD42020162283; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=162283

(J Med Internet Res 2022;24(8):e36620) doi:10.2196/36620

KEYWORDS
digital interventions; perinatal; mental health; well-being; logic model; systematic review; mobile phone

Introduction

The Importance of the Perinatal Period

Pregnancy, delivery, and the postnatal period can be times of increased psychological distress (stress, anxiety, or depression) [1], and up to 20% of women experience depression during the perinatal period [2,3]. There is considerable evidence that psychological distress during this period has detrimental effects on maternal health and can have long-term deleterious effects on the child [4-7], as recognized as part of the Developmental Origins of Health and Disease paradigm [8,9]. In particular, there is growing evidence of intergenerational transmission of poor mental health in utero and the first years of life through these pathways [10]. Therefore, cost-effective, accessible interventions that support lasting positive mental health while also preventing symptoms of mental health problems are of critical importance for public health. Mental health interventions to promote well-being have the potential to not only improve women's outcomes but also minimize the risk of negative health effect transmission to the next generation.

Health promotion strategies aim to enable optimal health and skills to cope with adversity in well subclinical populations. Therefore, it is important that efforts are made not only to deal with illness but also to develop individuals' emotional skills that can be applied in everyday life [11]. Psychological interventions aimed at perinatal women have also been shown to be effective when delivered digitally [12]. Digital interventions—that is, computer- or web-based interventions—can be delivered offline or on the web via a computer, tablet, or smartphone. In this format, interventions can be accessed by numerous people across wide geographical regions in a cost-effective and flexible manner [13]. Web-based interventions may be particularly useful in the perinatal period, given the accessibility issues faced by this population and as many pregnant women search the internet for health information [14,15]. Furthermore, these interventions may help overcome numerous barriers that exist for women who attempt to access traditional perinatal well-being, psychological distress prevention, or treatment programs, especially challenges in navigating psychosocial care systems [13]. Widespread restrictions imposed because of the COVID-19 pandemic have generated additional barriers to accessing mental health and well-being information and services [16].

Digital Mental Health Interventions

Although a recent systematic review provided preliminary evidence that web-based interventions can be a promising and advisable form of intervention during the prenatal period [13], there is a paucity of evidence on the long-term effectiveness of these programs [17]. There are many issues affecting digital mental health intervention (DMHI) implementation, such as availability issues, lack of promotion by health care providers, and lack of long-term outcome data; ultimately, program engagement is key. Low uptake of effective, evidence-based programs could diminish women's and infants' opportunities to enhance their well-being, limiting equitable public health benefits.

Dropout from the intervention and loss to follow-up reduces the treatment effect [18]. Although it has been argued that various strategies, including email prompts, SMS text messages, and *homework* are ways of helping participants develop intervention skills that can be applied, practiced, and sustained [19], it is unclear how frequently studies of DMHIs use or evaluate these strategies. Despite studies demonstrating intervention efficacy for those that remain in the study, we argue that it is just as salient to measure engagement as a benchmark of effectiveness. Web-based interventions provide tools to learn more about participant engagement and, furthermore, how it relates to retention and intervention outcomes, both in the short and the long term. This information can be used to understand the dynamics of engagement [18] and how to strengthen these characteristics in intervention development and delivery.

Assessing Engagement

It is widely accepted that the full benefit of many effective treatments can be achieved only if the prescribed regime is followed reasonably closely [20]. Recent reviews [21,22] have consistently highlighted these challenges with regard to low engagement and retention rates, particularly for digital programs that often experience poor reach and uptake [13]. Sustained engagement is a complex process that has been identified as a crucial factor in intervention success [23]. However, there is a lack of systematic methodologies to assess engagement, particularly in real-world contexts. Comparing program engagement across research studies is difficult because of the wide range of strategies applied to evaluate engagement outcomes [22,24]. Accordingly, applying structured processes to assess engagement can make comparisons more meaningful.

One of the frameworks for evaluating engagement in face-to-face programs, which can be adapted to web-based programs, is the Connect, Attend, Participate, and Enact (CAPE) model [25]. The CAPE model identifies and defines 4 aspects of engagement at various stages of intervention. First, *connect* pertains to how many people express interest in engaging in an intervention out of those eligible. Second, *attend* refers to continuous presence, such as the number of intervention sessions a participant completes. Third, *participate* is the degree to which participants actively engage with the content of the intervention, such as completing intervention tasks and remaining in the program. The final component, *enact*, refers to the participant making use of intervention strategies or knowledge as part of their daily life. Although this was developed to guide face-to-face parenting program engagement research, it can be readily applied to understanding and measuring digital intervention engagement in a research context.

Objectives

At a time when public resources are strained, policy makers and program administrators are looking to invest in effective, engaging prevention programs supported by scientific evidence and delivering long-term benefits. Intervention engagement must be foremost among these considerations, as this will ultimately determine the degree to which the target population takes up and benefits from the intervention when implemented in the community. Systematic reviews are an influential decision-making tool as they summarize a body of scientific research; identify implications for policy and practice [26,27]; and can be used to guide investment decisions, particularly for complex problems, such as poor intergenerational mental health.

This systematic review aimed to understand the reporting and levels of engagement in studies of web-based psychological mental health or well-being interventions administered in the perinatal period to women with subclinical mood symptoms.

Specifically, we aimed to (1) understand how studies report engagement, with engagement defined as containing the 4 steps in the CAPE model; (2) make recommendations on best practices to report engagement in DMHIs based on this; and (3) understand levels of engagement in intervention studies in this area.

To maximize the utility of this systematic review, we intended to develop practical tools for future public health use: to develop a logic model from the literature to reference the theory of change, evaluate the studies using the CAPE framework, and develop a guide for future data collection to enable consistent engagement reporting in web-based (and offline) interventions.

Methods

The methods used in this systematic review combine standard rigorous and transparent review methods using the Cochrane Synthesis Without Meta-analysis (SWiM) reporting guidelines [28] in conjunction with the development of a logic model to understand the theory of change.

Search Strategy

The review question, search strategy, inclusion criteria, and methods were registered in PROSPERO (International Prospective Register of Systematic Reviews; approval number CRD42020162283). The research question was as follows: what is known about engagement in digital mental health and well-being programs for women in the perinatal period? A systematic database search was conducted to identify papers relevant to the aims of this review. The initial search was performed by the first reviewer (JAD), using the Ovid Platform for MEDLINE, PsycINFO, EMBASE on the EBSCO Platform, Scopus, Web of Science, and Medical Subject Headings on Demand for all English-language articles published in the past 10 years (ie, from January 1, 2010, to May 29, 2020). Keywords and index terms identified as relevant in the search strategy were used and individual search criteria were developed for each database. All the database search strategies are provided in Multimedia Appendix 1. The impact of the COVID-19 pandemic delayed this publication; therefore, a subsequent rapid review was undertaken in May 2022 and performed in Google Scholar Advanced search to elicit any further publications since June 2020.

Identification of Studies and Eligibility Criteria

The search strategy aimed to identify studies reporting on engagement and retention in digital mental health and well-being programs for women during and after pregnancy. Clear inclusion and exclusion criteria were developed using the Population, Intervention, Comparison, Outcomes, and Study framework to guide the inclusion criteria for participants, intervention or phenomena of interest, comparators, outcomes, study design, and context (Textbox 1).

Textbox 1. Inclusion and exclusion criteria (based on the Population, Intervention, Comparison, Outcomes, and Study framework).

Inclusion criteria

- Participants
  - Childbearing individuals in the perinatal period (ie, from conception to the first year of the infant's life)
  - Studies focusing predominantly on the childbearing individual but can include partners
  - Studies that include childbearing individuals at moderate risk for psychological distress (ie, with Edinburgh Postnatal Depression Scale score ≤12)
  - Studies that include women at risk of postnatal depression with a history of depression or anxiety (ie, early intervention)
- Intervention
  - Any minimal contact digital interventions provided in the perinatal period aiming to reduce mild to moderate psychological distress or promote psychological well-being (ie, minimal contact as defined by a maximum of <1 hour of direct contact each week)
- Comparators
  - Studies with any form of comparator were considered
- Outcomes
  - None; although the focus of the review was on engagement outcomes, we included any studies of interventions meeting the above criteria to determine the proportion that reported engagement outcomes
- Study design
  - Quantitative (eg, randomized controlled trials, quasi-experimental studies, cohort studies, descriptive studies), and qualitative studies

Exclusion criteria

- Participants
  - Studies considering programs before conception and those specifically targeting the child
  - Studies focusing predominantly on the partner or father
  - Studies that include women at high risk for psychological distress (ie, with Edinburgh Postnatal Depression Scale score ≥13)
- Intervention
  - Interventions with a primary focus other than mental health or well-being (eg, parenting self-efficacy)
  - Interventions delivered face to face or as telehealth or telephone coaching

Selected Studies

All papers that appeared eligible based on their title and abstract were retrieved for screening. The first author (JAD) reviewed the titles and abstracts of all papers, assessed eligibility, and noted any reasons for exclusion. Full-text articles were assessed for eligibility and reviewed independently by both the first author (JAD) and third author (LYG). Once the third author (LYG) had reviewed the papers, any discrepancies were resolved through team discussion. The reference lists of the included studies were examined to identify additional relevant papers.

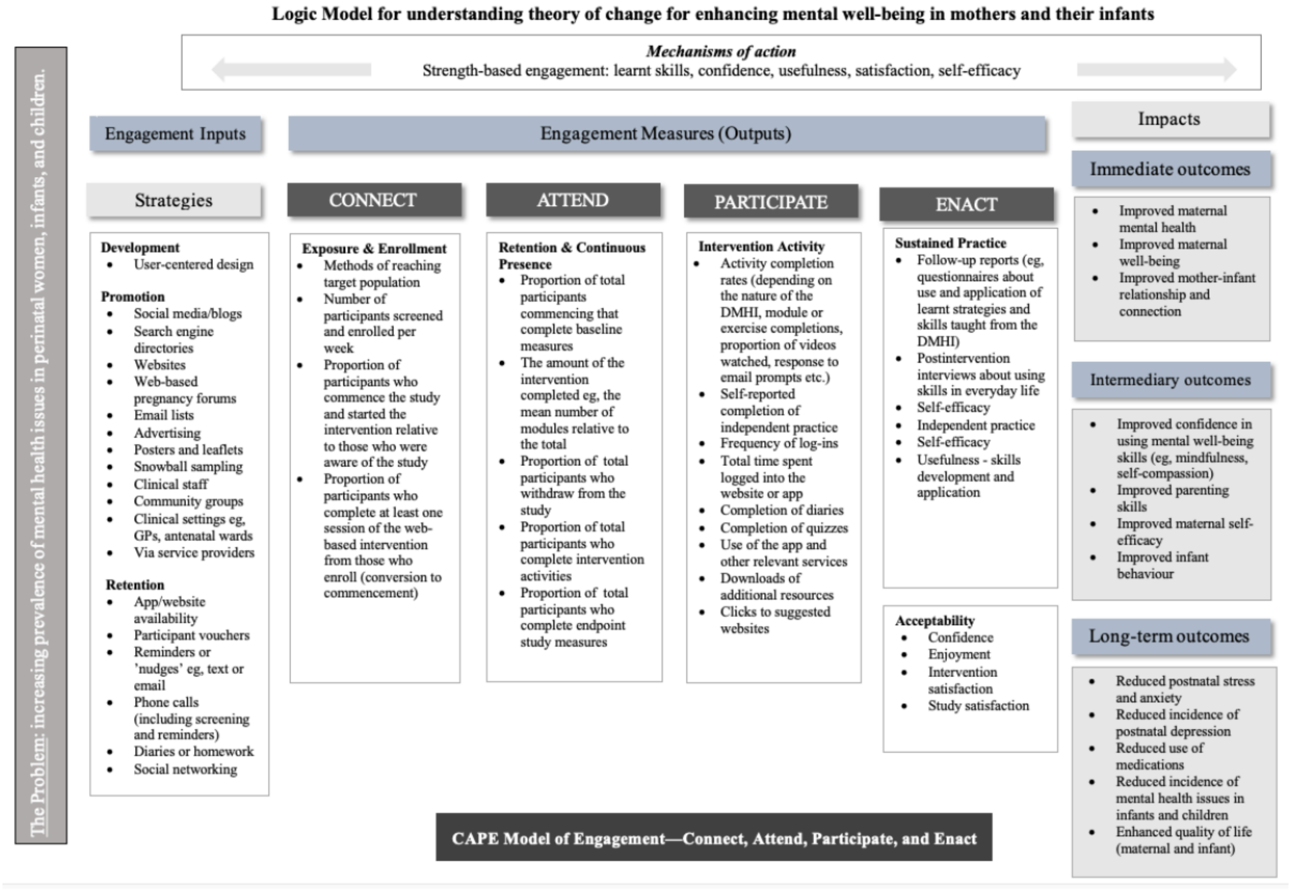
Coding of Study Characteristics and Data Extraction

Key article characteristics were recorded using a Microsoft Excel (version 2020) data extraction table developed for this review. These characteristics included general information about the study, such as the country and author, along with specific information about the study design, comparators, and intervention type. Coding of the study characteristics enabled us to group the studies as part of our synthesis. As our primary aim was to understand the engagement of the study population, we characterized the assessment time points, engagement measures, and reporting of attrition and adherence. Data relevant to engagement were extracted using the CAPE framework; this included variables for recruitment, retention, attrition, and follow-up time points. A framework analysis methodology [29] was used to determine which variables should be included in each step of the CAPE framework.

Development of Logic Model

Logic models can help conceptualize a complex review question and specify analytic links to test the plausibility that a program works as intended [26]. Logic models typically illustrate the chain of reasoning underpinning how interventions lead to immediate (or short-term) outcomes and then to longer-term outcomes and impacts [30]. A key part of the model is detailing the mechanisms of change within the pathway and the moderating and mediating factors that may be associated with or influence outcomes. This is often referred to as the theory of change [30]. In this systematic review, the research team developed a logic model to aid the process of understanding how and when the CAPE framework could be applied to interpret the role of different engagement variables in promoting outcomes in digital perinatal mental health and well-being programs. The project team collaboratively developed the logic model, drawing on themes in the literature and the team's collective knowledge and experience. To develop the logic model, we incorporated the types of engagement metrics found in the selected studies that could be used to assess engagement (Figure 1).

Figure 1. Proposed logic model. GP: general practitioner; DMHI: digital mental health intervention.



Quality Appraisal and Risk of Bias

The risk of bias for studies included in this review was based on the Cochrane Collaboration's tool for assessing the risk of bias for each category of study (ie, randomized controlled trials [RCTs] and non-RCTs), and the risk of bias was adapted for this review and classified as low, uncertain, or high based on the Cochrane risk of bias tool [31] and the primary aim of this systematic review (ie, engagement rather than efficacy). The assessment of study quality was undertaken by the first author (JAD) and reviewed by the project team. Multimedia Appendix 2 [17,32-46] provides the detailed risk of bias assessments of the included studies.

Synthesis of Results

As this systematic review synthesized the results from a diverse range of interventions, we used SWiM guidelines [28] to promote transparent reporting. The SWiM items enable studies to be grouped and provide guidance on the reporting of standardized metrics used for the synthesis of findings. Specifically, we undertook the following steps:

1. Summarized the characteristics of each study and reported intervention implementation, recruitment and engagement activities, study findings, reported attrition, and methodological quality
2. Determined which studies were similar enough to be grouped within each comparison by comparing across studies (eg, types of digital platform and postnatal vs antenatal)
3. Determined which data were available for synthesis
4. Synthesized the characteristics of the studies
5. Performed a statistical synthesis for appropriate quantitative data and comprehensive critical appraisal through a meta-synthesis approach for qualitative data
6. For each trial included in this systematic review, we recorded counts of trial participants who were assessed for eligibility, those who were recruited, and those who were allocated to the intervention and control arms; rates of

recruitment, trial completion, and loss to follow-up were synthesized by evaluating the proportion of recruitment, completeness, and loss to follow-up in base R (R Foundation for Statistical Computing) statistical package [47]; synthesized data were reported as forest plots [48].

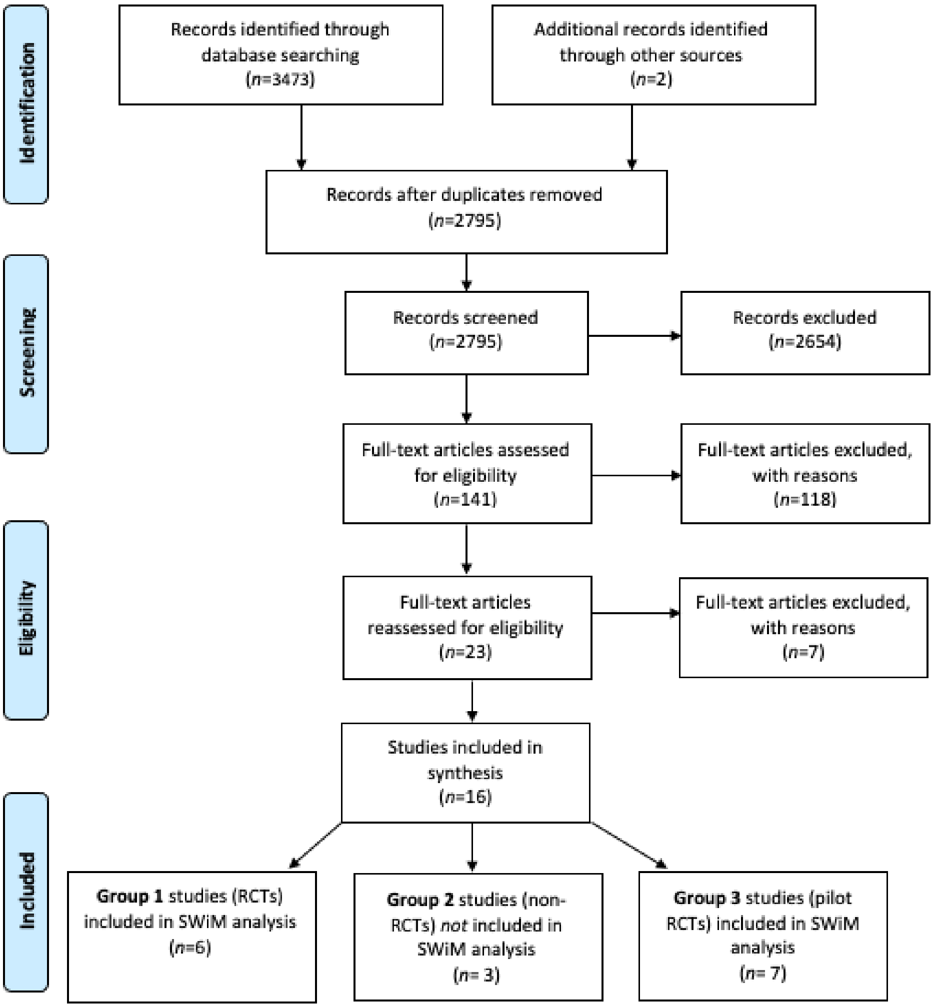
Results

Included Studies

The electronic searches generated a database of 3473 potentially eligible studies that were assessed using the review eligibility criteria. After duplicates were removed (680/3473, 19.58%), all titles and abstracts were screened for eligibility. In total, of the 3473 studies, 2795 (80.48%) records were screened, and 2654 (76.42%) were excluded based on the inclusion or exclusion criteria (Textbox 1). After the first screening, 5.31% (141/2654) of potential studies remained; the full-text articles were assessed for eligibility by the first and third authors. Of the remaining 141 studies, 125 (88.7%) were excluded on consensus by the project team; the first and third authors independently screened the papers that were verified by the team, resulting in a final selection of 16 (11.3%) studies to be included in the synthesis. The final studies were then grouped according to the study design.

The literature search and inclusion processes are detailed in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 2). The search flow diagram indicates the papers that were selected for synthesis using the PRISMA guidelines [49]. A secondary rapid search conducted in 2022 did not yield any additional papers that met our specific inclusion criteria.

Figure 2. Search flow diagram (PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses]). RCT: randomized controlled trial; SWiM: Synthesis Without Meta-analysis.



Synthesis of Results

Overview

The primary aim of this systematic review was to assess the engagement of women participating in digital mental health or well-being interventions; therefore, the standardized metric and transformation method [28] used across all studies were those used to evaluate engagement rather than well-being effect sizes. On analysis of study characteristics, we grouped the 16 interventions according to study design to provide a cohesive comparison in a broad range of study types. Group 1 was RCTs (6/16, 38%), with active and control arms. Generally, RCTs had a longer follow-up period (maximum of 12 months). Group 2 included non-RCTs (3/16, 19%), with no active comparators and brief or no follow-up period. Group 3 comprised pilot RCTs (7/16, 44%) with active and control arms and a follow-up period.

Study Characteristics

In all groups, there was a range of therapeutic approaches, including cognitive behavioral therapy, parenting education, positive psychology, mindfulness, and compassion-based training. Several studies used a psychoeducation approach to build parenting self-efficacy, such as Chan et al [50], Shorey et al [51], Corno et al [32], and Tsai et al [34]. In group 1, overall, 25% (4/16) were delivered as self-help internet interventions and 13% (2/16) as smartphone-based mobile apps; in group 2, all were delivered as self-help internet interventions; in group 3, all were delivered as internet interventions apart from the study by Barrera et al [37], which was delivered as an SMS text messaging program. In addition, 13% (2/16) of studies included the assessment of physiological biomarkers: Cornsweet [38] and Matvienko-Sikar and Dockray [41]. A summary of the study characteristics is reported in Tables 1-3 (a more detailed report is available in Multimedia Appendix 2, including intervention outcomes). For this review, we reported engagement measures collected for each study as CAPE metrics.

Table 1. Group 1: randomized controlled trials (N=6).

Intervention type, format, and duration	Study aims (sample size)	Engagement measures: connect	Engagement measures: attend	Engagement measures: participate	Engagement measures: enact
Self-guided; iWaWa[a]; 9 modules [13]	Assess the feasibility and acceptability of iWaWA among postpartum women with anxiety (89 participants)	• Assessed for eligibility (n=147): recruited via social media, posters, and flyers and numbers recruited • Reasons for exclusion • 89 enrolled and randomized to treatment and control	• Engagement with internet-based components • Attrition and attendance • Participant CONSORT[b] flow diagram (access, allocation, and follow-up)	• Module views, module completion, number and duration of support calls	• Treatment feasibility (engagement and usability) and acceptability (usefulness, satisfaction, and helpfulness) were assessed after treatment through semistructured interviews
Smartphone-based mobile app [50]	Assess the difference in the levels of antenatal and postnatal depression in participants (660 participants)	• Assessed for eligibility (n=803) • Reasons for exclusion • 660 enrolled and randomized (intervention or treatment as usual)	• Participant CONSORT flow diagram (eligibility, enrollment, randomization, follow-up, and analysis) • Retention rates	• The use of the app and other relevant services (eg, antenatal classes and other pregnancy resources: books and websites) documented by self-report	• Postintervention survey included • Use of the app
Web-based compassion-based intervention; *Kindness for Mums Online*; 5-6 weeks [52]	Assess the effect of the intervention on participants' well-being (206 participants)	• Assessed for eligibility (n=310) • Recruitment methods: social media and snowball sampling • Participant vouchers • Accessibility • Reasons for exclusion • 206 enrolled and randomized	• Participant CONSORT flow diagram (enrollment, allocation, follow-up, and analysis)	• Reporting of attrition and engagement (ie, completion of sessions and frequency or program use)	• Acceptability: participants were asked to rate the ease of use and satisfaction after the intervention
A Chinese version of the MBSP[c] program; 10 hours of training with 36 episodes; 6-week internet-based intervention [53]	Assess the effect of the mindful self-compassion intervention on preventing postpartum depression in a group of symptomatic pregnant women (314 participants)	• Assessed for eligibility (n=472) • Screening and baseline assessment (n=344) • Reasons for exclusion • Randomized (n=314)	• Participant CONSORT flow diagram (eligibility, allocation, follow-up, and analysis) • Attendance rates • Reporting of retention	• Reporting of attrition • Feasibility and acceptability • After completing each exercise, participants were instructed to exercise the steps during the day; participants provided a graphical overview and a web-based diary book where they registered their reflections	• Reporting of retention and attrition after the intervention
Condensed web-based version of an 8-week mindfulness course; "Be Mindful Online"; 4 weeks on the web [54]	Evaluate the potential of a web-based mindfulness course for expectant participant women (185 participants)	• Assessed for eligibility (n=237) • Recruitment methods (email lists, social media advertising, and posters in community settings) • Reasons for exclusion • Enrolled and randomization methods	• Participant CONSORT flow diagram (recruitment, allocation, follow-up, and analysis)	• Regular reminders to log on or contact the research team via email • Reporting of retention and attrition	• Postcourse evaluation 45 days after baseline

Intervention type, format, and duration	Study aims (sample size)	Engagement measures: connect	Engagement measures: attend	Engagement measures: participate	Engagement measures: enact
Mobile app for psychoeducation and postnatal depression; “Home-but not Alone” [51]	Examine the effectiveness of the program in improving participant parenting outcomes (250 participants [couples])	• Assessed for eligibility (n=360 couples) • Reasons for exclusion • Recruitment methods • Randomization methods to intervention or control	• Participant CONSORT flow diagram (eligibility, recruitment, allocation, follow-up, and analysis)	• The research team monitored the use of the app and parents received reminders each week	• Intervention posttest

[a] iWaWa: internet-based What Am I Worried About.

[b] CONSORT: Consolidated Standards of Reporting Trials.

[c] MBSP: Mindfulness-Based Strengths Practice.

Table 2. Group 2: non–randomized controlled trials—case series, open trial, and quasi-experimental (N=3).

Intervention type, format, and duration	Study aims (sample size)	Engagement measures: connect	Engagement measures: attend	Engagement measures: participate	Engagement measures: enact
Positive psychology web-based intervention; 5-week web-based self-applied positive psychology intervention specifically adapted for pregnant women; 4 modules [32]	Examine the effect of a positive psychology web-based intervention on indices of participants’ prenatal well-being (6 participants); case series design	• Eligibility and recruitment method • Preassessment on the web	• Weekly emails—reminders for assessments	• Compliance with the intervention measure was developed by the research team • No reported attrition	Exercise preferences were assessed at the posttest time point
Internet program plus weekly phone coaching sessions, individually or group-wise; MMB[a] program; 8 weeks [33]	Examine the feasibility, acceptability, and preliminary outcomes of MMB for use with pregnant women at risk for depressive relapse (37 participants); open trial—no control group	• Assessed for eligibility (n=48) • Reasons for exclusion • Recruitment methods—flyers and via service providers • Prescreening by phone • Intake interview in person or by phone • Participant enrollment and flow (eg, reasons for declining to participate)	• Participant CONSORT[b] flow diagram (eligibility, enrollment, follow-up, and analysis)	• Session completion and participation in phone coaching calls • Home practice completion • Participant engagement (eg, completion of sessions, practice per week, and time)	Self-reported satisfaction (perceived benefits and challenges) via questionnaire and engagement interview (qualitative) at session completion
Web-based modules: web-based maternity health records, antenatal health education, self-management journals, and infant birth records [34]	Investigate the effectiveness of a web-based antenatal care and education system on pregnancy-related stress, general self-efficacy, and satisfaction with antenatal care (135 participants) quasi-experimental design	• Eligibility—control (n=75) and experimental (n=80) group at pretest • Recruitment methods (convenience sampling) • Assignment methods to experimental or control groups	• Participant CONSORT flow diagram (enrollment, follow-up, and analysis) • Attrition	• Assistance was offered via telephone, email, web conferencing, or face-to-face guidance • Follow-up phone calls were made to the participants • Attrition	N/A[c]

[a] MMB: Mindful Mood Balance.

[b] CONSORT: Consolidated Standards of Reporting Trials.

[c] N/A: not applicable.

Table 3. Group 3: pilot studies (N=7).

Intervention type, format, and duration	Study aims (sample size)	Engagement measures: connect	Engagement measures: attend	Engagement measures: participate	Engagement measures: enact
Brief web-based self-help intervention—5 components considered effective in challenging negative beliefs [35]	Assess positive mood in participating mothers of babies and toddlers (80 participants)	• Eligibility • Recruitment methods—internet, leaflets, and community postnatal groups • Randomization methods	• Only 1 session	• Compliance (missing data)	• Acceptability—an open-response question at the end of the intervention (qualitative) • Implications for policy and practice
Automated self-help internet intervention; 8 lessons—accessible anytime [36]	Assess the efficacy of the intervention to reduce the risk of postnatal depression in participating women (111 participants)	• Assessed for eligibility (n=5071) • Consented (n=2966) • Recruitment methods—web-based search engine directories, (eg, Google advertisements “sponsored links”) • Randomization methods • Initial log-ins to the website	• Participant CONSORT[a] flow diagram (eligibility, consent, allocation, follow-up, and analysis) • Adherence	• Automated email messages • Automated self-help via website • Log-ins, total time spent logged into the website, and the last lesson viewed recorded • Module feedback on the materials viewed (eg, usefulness and understandability) • Attrition	• Includes discussion on experience and engagement and feedback assessment
Minimal contact automated SMS text messaging; *Baby-Text* program [37]	Assess acceptability of an SMS text messaging program to prevent postpartum depression (10 participants [pregnant and postpartum women])	• Eligibility • Recruitment methods—flyers at general public bulletin boards and community agencies; websites and blogs	• Compliance	• Attrition	• Feedback assessment (qualitative) • Acceptability assessment
Intervention—self-guided; 15 steps, each of which takes 45 minutes [38]	Assess feasibility and acceptability; study 1 (n=6): effects of a single teaching and biofeedback session on maternal and fetal biofeedback; study 2 (n=9): effect of consumer satisfaction	• Study 1: eligibility and recruitment methods (flyers at antenatal classes) • Study 2: eligibility and recruitment methods (flyers at antenatal classes)	• Study 1: compliance with baseline and 2 conditions (teaching and practice) • Study 2: compliance to complete 15 steps	• Attrition • Feasibility and acceptability	• Study 1: no postintervention measures • Study 2: postintervention assessment and interview • Qualitative follow-up
8-week web-based prevention intervention; website plus initial phone call; 16 core didactic lessons plus 3 postpartum booster sessions and 5 associated tools [39]	Assess a CBT[b] peer support intervention to prevent postnatal depression in participants (24 participants)	• User-centered-design, recruited via flyers • Assessed for eligibility (n=216) • Completed baseline assessment (n=30) • Enrolled and randomization methods	• Participant CONSORT flow diagram (screened, completed the baseline assessment, and enrolled) • Adherence	• Email notifications • Total log-ins and completion of tools and lessons • Peer support features (likes, comments, nudges, and posts) • Reporting of attrition and site use (log-ins); usability and acceptability	• Usability and satisfaction (Usability, Satisfaction, and Ease of Use questionnaire)

Intervention type, format, and duration	Study aims (sample size)	Engagement measures: connect	Engagement measures: attend	Engagement measures: participate	Engagement measures: enact
Self-guided, web-based intervention to prevent postpartum depression symptoms; *Be a Mom*; 5 modules [40]	Explore the processes underlying therapeutic change for participants in the intervention (194 participants)	• Assessed for eligibility (n=643) • Email invitation to participate • Recruitment methods—in person and web-based • Reasons for exclusion • Baseline assessment (n=241) • Randomization methods (intervention or waitlist control)	• Participant CONSORT flow diagram (eligibility, enrolled, randomized, and follow-up) • Adherence	• Email reminders after 7 days without accessing intervention • Attrition	• Postintervention measures included emotion regulation, psychological flexibility, and self-compassion
Web-based mindfulness and gratitude intervention 4 times a week for 3 weeks [41]	Assess the effect of a novel gratitude and mindfulness-based intervention on prenatal stress, cortisol levels, and well-being in participating women (46 participants)	• Assessed for eligibility (n=362) • Recruitment methods—posters, leaflets, and pregnancy forums • Reasons for exclusion • Randomization methods • SMS text message reminders • No additional contact with the study team during the study period	• Participant CONSORT flow diagram (enrollment, allocation, follow-up, analysis)	• Participant adherence was evaluated as the total frequency of completion of the web-based diary entries • Proxy measure for full intervention use	• Limitations in fidelity evaluation

[a] CONSORT: Consolidated Standards of Reporting Trials.

[b] CBT: cognitive behavioral therapy.

Variables Assessed to Evaluate Engagement

Overview

Participant engagement was evaluated using the CAPE model of engagement [25], which is described in more detail in the following sections. Three-quarters of all the studies included a CONSORT (Consolidated Standards of Reporting Trials) participant flow diagram with similarities in reporting (ie, enrollment, allocation, follow-up, and analysis). In terms of strategies to prompt engagement and promote retention (attendance and participation), various methods were used and reported, including email and text reminders, peer support features, and phone calls to participants.

Connect: Exposure and Enrollment

Connect was operationally defined as the proportion of participants who entered the study and started the intervention relative to those who were aware of the study. Although many studies reported the *exposure* methods for the target audience, for example, advertising via Facebook, Twitter, web-based email lists, community sites (medical and retail), and third-party websites (Ashford et al [13], Krusche et al [54], and Felder et al [33]), there were rarely reporting of the total population size exposed to advertising and other recruitment means. Many studies reported eligible participants who made the initial contact. For example, Barrera et al [36] reported eligible participants (n=5071) as female, pregnant, aged ≥18 years, and interested in the study website for personal use. From this total group, 2966 participants went on to participate by signing an informed consent form. A further 2114 potential participants were excluded, and the reported reasons included website error, current or missing status of major depressive episodes, and incomplete baseline. The final number of randomized participants was 852. Given the substantial drop-off between exposure to recruitment methods and randomization, reporting metrics at each stage of this process can highlight where efforts must be targeted to increase engagement.

All studies in this review reported enrollment rates in the intervention, which we defined as those who commenced the intervention relative to those who expressed interest in the study. Conversion to commencement was based on multiple factors, not just the participants' decision to engage, both dependent and independent of the inclusion or exclusion criteria. Most studies in this review reported reasons for exclusion, ranging from lack of contact or completion of baseline surveys to elevated mental distress scores. Enrollment rates varied from a high rate of 82% (Chan et al [50]; group 1) commencing from the eligible study sample, with the lowest enrollment rate at 12% (Duffecy et al [39]; group 3). Generally, there were higher enrollment rates in the group 1 studies (clustering approximately

60%) than in group 3 (clustering approximately 25%). Only one of the studies reported strategies designed to increase enrollment. Duffecy et al [39] undertook a user-centered design process before the pilot trial to engage women from the target population in the intervention-building process. However, this study also reported the lowest enrollment rate.

Attend: Retention and Continuous Presence

In face-to-face interventions, *attendance* refers to the proportion of the sessions attended by each participant. Ideally, for digital interventions, attendance would be a measure of the amount of intervention completed (eg, the mean number of intervention modules relative to the total number of modules) or similar, such as the number of participants who completed all web-based interventions. This information was not stated in any of the studies included in this review. As a result, we calculated a proxy for intervention *attendance* as study attendance, operationally defined as intervention retention and continuous presence (continued interaction with the intervention), in both the intervention and control arms (where controls were used). All studies reported retention in terms of the rate of those who enrolled versus those who completed the study.

The highest reported study retention (groups 1 and 3) was reported by Ayers et al [35] at 90%. Barrera et al [36] had the lowest intervention retention at 13%. (Figure 3 [13,50-54] and Figure 4 [35,36,39-41]).

In the control arm, Guo et al [53] had the highest participant retention rate of 89%, whereas Barrera et al [36] had the lowest at 13%.

Figure 3. Participant retention in the intervention arm (group 1); 95% CIs determined by test of proportions [13,50-54].

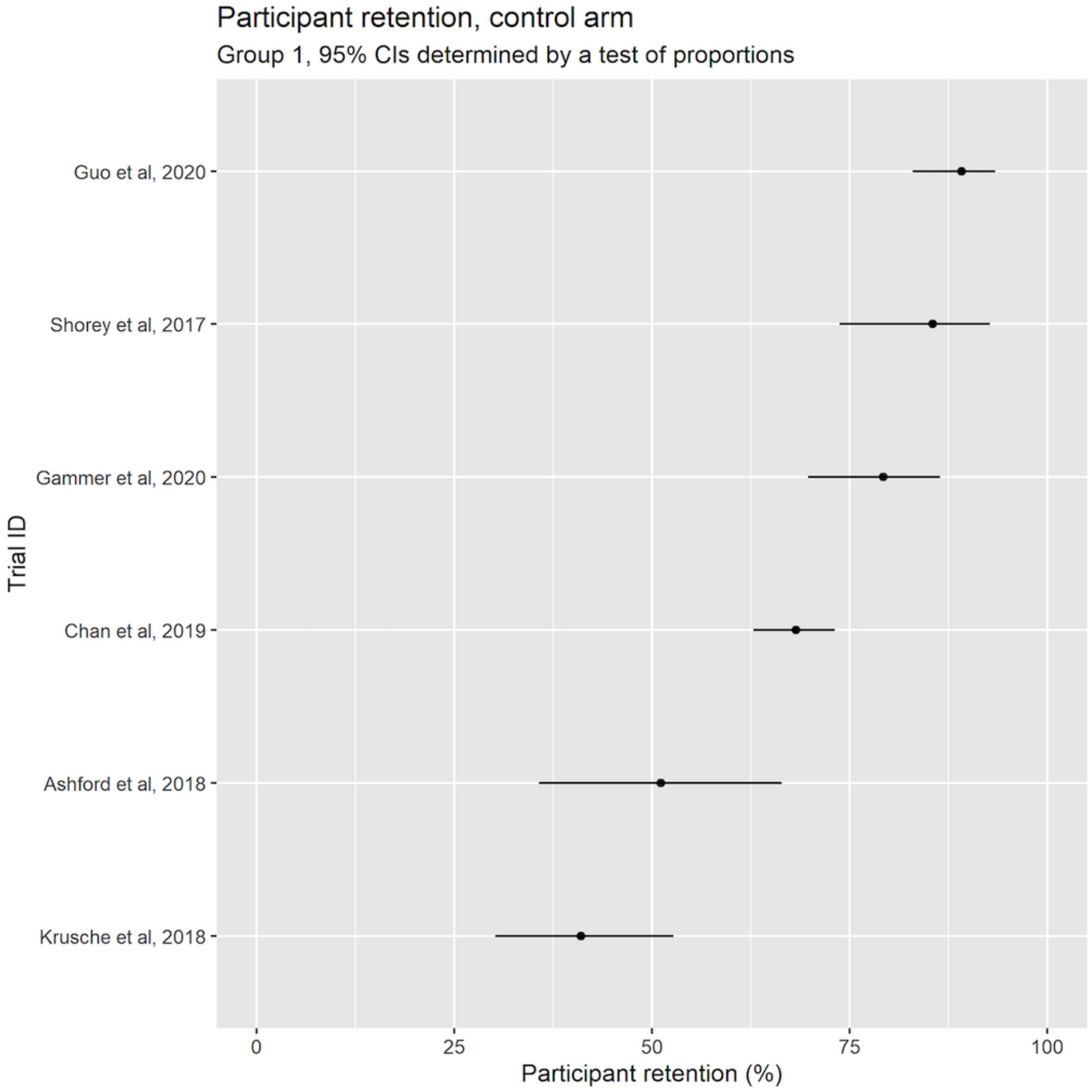
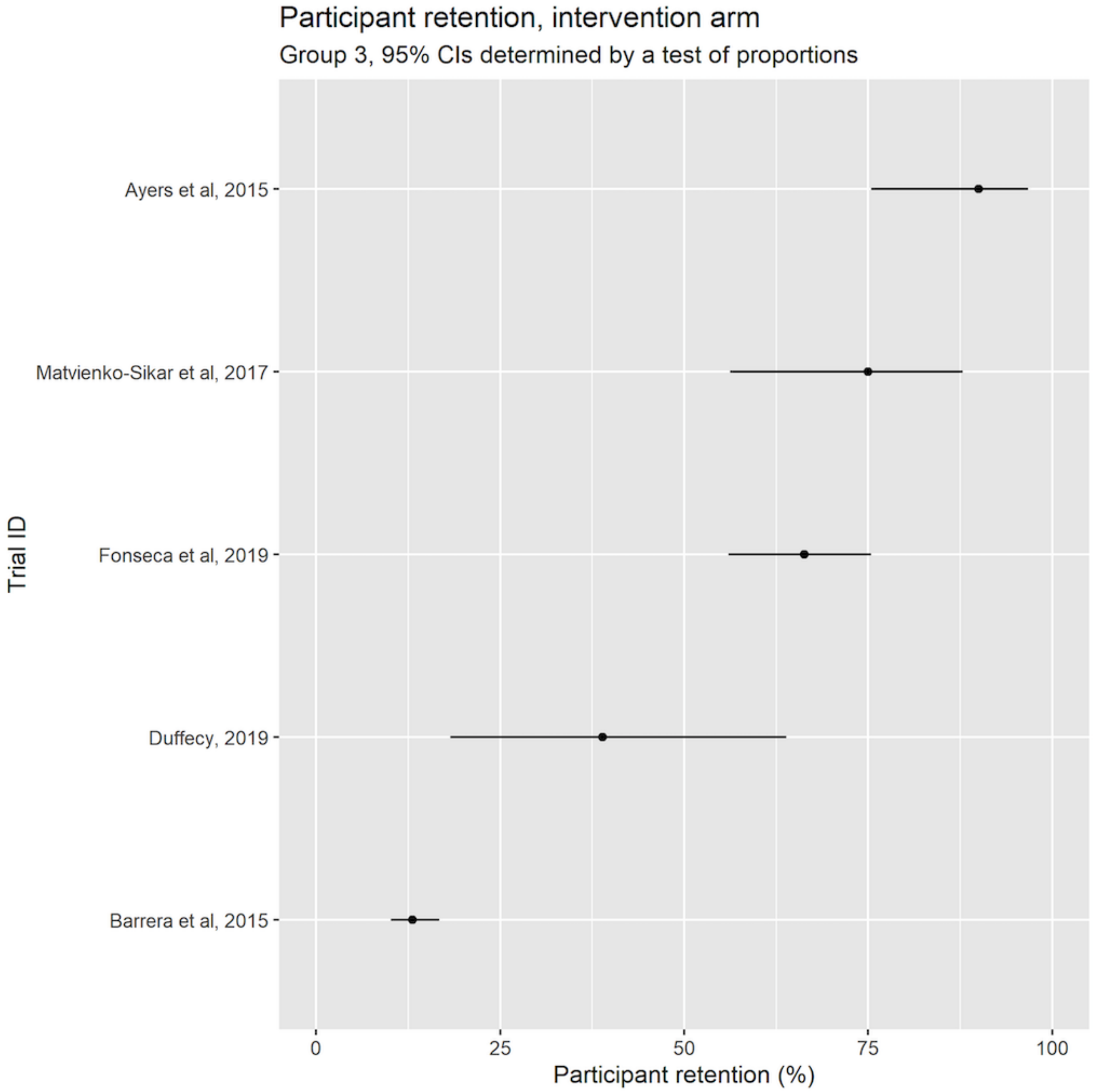


Figure 4. Participant retention in the intervention arm (group 3); 95% CIs determined by test of proportions [35,36,39-41].



Participate: Active Engagement

Participation was operationally defined as the completion of intervention activities; that is, *active engagement* with the intervention material. Follow-up prompts to encourage participation varied across studies; however, there were mostly weekly reminders such as SMS text messages, emails, and phone calls. A broad spectrum of metrics was used to report ongoing participation in each study, including module views, module and home practice completions, engagement with web-based components, use of the app (frequency of log-ins) and other relevant services (eg, antenatal classes), number and duration of support calls, and web-based diary entries. The heterogeneous nature of recording these activities is appropriate for the types of intervention strategies but limits our ability to consistently report and compare across studies.

Enact: Sustained Practice (Leading to Long-term Well-being Effects)

The limited follow-up period restricted our ability to report against measures indicating that participants applied and practiced learning skills [24]. Immediate postintervention follow-up was reported by all studies through a combination of self-report assessments, interviews, program accessibility, acceptability, and satisfaction; however, this did not necessarily include commentary on putting skills into practice. Approximately 31% (5/16) of studies undertook postintervention interviews to unpack outcomes such as usefulness, satisfaction, experience, and engagement. These interviews contributed more robust information to the user experience and provided some insight into the reasons for participation or enactment (or lack thereof) but not necessarily on the enactment itself (ie, use of the intervention skills). For example, internet-based What Am I Worried About (Ashford, 2018) was experienced as not user-friendly enough, too long, and not smartphone-friendly. Parts of the content were experienced as not always relevant or appropriate. The participants felt that the program could be improved by having it in a smartphone app format and by making the content more concise and inclusive of different parenting styles. Guo et al [53] had the highest participation rate and longest follow-up period (12 months post partum), and participants rated the program as highly useful.

Logic Model Development

Through the analysis and reporting of each study, we recorded the types of quantitative and qualitative measures found in the selected studies that could be used to measure engagement. This enabled us to systematically construct a logic model based on our understanding of how interventions are expected to work. This was particularly pertinent for this systematic review as we did not perform a meta-analysis. As indicated, we grouped variables related to the CAPE framework; the logic model includes a range of metrics that could be systematically reported when synthesizing engagement data to visually interpret the underlying theory of change.

The logic model (Figure 1) contains 6 columns detailing the intended pathway from inputs (engagement strategies) to long-term outcomes or impacts. This approach takes a long-standing view of interventions to achieve their intended consequences. To build this model, we incorporated the types of measures undertaken in each study in this systematic review, as well as a broader range of CAPE measures found in the wider literature. The logic model indicates the point in the pathway at which the data should be collected. Mechanisms of action are factors that may facilitate engagement using a strength-based approach.

As part of this systematic review, we aimed to develop a guide for future data collection to enable consistent engagement reporting in web-based (and offline) interventions. Table 4 outlines a range of metrics that could be consistently applied in future data collection and reporting to enhance understanding of engagement and enable comparative intervention assessments.

Table 4. Proposed reportable metrics: engagement.

CAPE[a] model of engagement	Measures	Definitions
Connect	Exposure and enrollment (rates should be reported for each trial arm separately)	• Defined target population (ideally with population size if available) • Methods of recruitment and size or proportion of the population exposed to each recruitment method • Enrollment rate: proportion of participants who start the intervention relative to those who are exposed to the intervention and those who provide consent for the study • Connection rate: proportion of recruited participants electing to enroll relative to those who are eligible
Attend	Intervention retention	• Proportion of participants who complete the intervention relative to those who enroll in the intervention • Mean, SD, and range of the number of modules completed
Participate	Intervention activity	• Active engagement (depending on the nature of the intervention; this may be module completions, exercise completions, proportion of videos watched, and response to emails) • Log-ins (frequency and duration) • Time spent logged into the website or app • Use of recommended resources (eg, downloads of additional resources and clicks to suggested websites)
Enact	Sustained practice	• Follow-up reports (eg, questionnaires about the use and application of learned strategies or skills taught from the DMHI[b]) • Postintervention interviews about using skills in everyday life • Sustained behavior change

[a] CAPE: Connect, Attend, Participate, and Enact.

[b] DMHI: digital mental health intervention.

Discussion

Principal Findings

In this systematic review, we categorized the selected studies according to study type and assessed their individual and pooled characteristics. We applied the CAPE framework [25] to all studies so that we could collectively assess and compare *connection*, *attendance*, *participation*, and *enactment*. Reporting of connection and attendance measures was fairly standardized across studies and frequently reported using a CONSORT diagram; therefore, the number of people who were eligible and expressed interest in participating, proportion of participants recruited and entered in the study, and proportion of participants who were randomized and followed up, including *treatment not started* and attrition rates, was clear. Approximately 75% (12/16) of studies, provided a CONSORT participant flow diagram (indicating aspects of attendance); however, the reporting categories and terminology varied between diagrams. In general, the least frequently reported domain was enactment (ie, real-world use of intervention skills), with only 38% (6/16) of studies clearly recording and reporting results such as satisfaction, usefulness, helpfulness, and perceived benefits of the skills learned in the intervention.

Some studies reported strategies to increase *connection*. For example, Ashford et al [13] included specific details on recruitment methods, such as social media platforms, parenthood websites, and the use of posters and flyers in clinical settings. However, other studies reported minimal details of recruitment methods (eg, Guo et al [53]). No study reported information on the background target population size; that is, the total potential pool of eligible participants. This might be a useful future metric to determine and report as an indicator of the total target population reach and the effective strategies that engage them.

As previously defined, *attendance* is a measure of DMHI completion through modules or similar exercises. As this was not definitively reported, for *attendance*, we calculated a proxy for intervention *attendance* as study attendance, operationally defined as intervention retention and continuous presence, both in the intervention and control arms (RCTs). Guo et al [53] had the highest participant retention and the lowest lost to follow-up participant rate in both the intervention and control groups; the intervention group showed significant improvement in depressive and anxiety behaviors. The women in this study were at a higher risk of presenting with psychological distress at baseline; although they fit our inclusion criteria, they may have had increased motivation to attend. Duffecy et al [39] undertook a user-centered design process before the pilot trial to engage women from the target population in the intervention-building process and consult on aspects such as topics, sites, and usability of potential applications. In theory, this should support *attendance* by reducing attrition and improving retention; however, dropout from baseline assessments to 6 weeks post partum was high (63%) [39]. We recommend that future studies report both *intervention* attendance and *study* attendance as they are distinct metrics.

A key concern in web-based interventions is the lack of *participation* [18]. Logs of access and use of web-based interventions can give researchers insight into people's behavior. As Piotrowska et al [25] suggest, "The CAPE model proposes that despite the immense importance of connecting with parents and encouraging their attendance, it is active participation that has the greatest impact on parenting." Digital interventions provide tools for learning more about participant engagement and how this relates to retention and intervention outcomes, as well as how they might be improved through the use of different ongoing engagement strategies. Crouper et al [18] quantified participant engagement using data such as dosage, exposure, or adherence. In this systematic review, few studies reported clear metrics for participation, with the exceptions of Duffecy et al [39] and Barrera et al [36]. Other suggested metrics for future research include downloading suggested resources or websites, watching suggested videos, completing quizzes and homework, or other metrics that indicate that the participant is continuously engaging with the intervention. Additional features that have potential but continue to be underexplored and underused include chatbots, games, storytelling, rewards, avatars, and personalization [42]. These features could be developed to improve participation in interventions, general app use, and studies.

Enactment is difficult to define but should be represented by measures indicating that participants put what they learned from the program into practice [24]. The limited follow-up of these studies restricted our ability to report these criteria. Only one study, Guo et al [53], followed up for any length of time, and only 31% (5/16) of studies conducted exit or follow-up interviews [13,33,35-37]. These interviews contributed more robust information on the user experience. Studies assessing skill development and use underscore the potential pathways in self-guided internet therapy, such as cognitive behavioral therapy, as mechanisms of positive clinical change [43,44]. Although these studies target clinical groups, they contribute to reinforcing the need to capture behavioral skill adaptations beyond the duration of the intervention.

Understanding the barriers to and enablers of real-world utility and practice is crucial if app developers want pragmatic uptake and efficacy of interventions. Sufficient resourcing may be a factor in longitudinal follow-up; however, to leverage the impact and cost-effectiveness of interventions, studies should factor longer-term assessments in the design process from conceptualization. Nevertheless, easier and low-cost measures of enactment are possible and suggested for future research, including questionnaires on the frequency of using skills taught during the intervention.

Interpreting Results Using a Logic Model

As part of this review, we developed a logic model to facilitate the process of gathering and integrating studies of complex interventions to better inform our interpretations of cumulative results. The logic model included synthesized data capture and engagement methods used in each study. Theoretically, logic models need moderating or mediating factors to understand how the pathway develops. In these studies, there was a common strength-based approach, such as skill development, confidence, satisfaction, and self-efficacy. Overall, the heterogeneous nature of the data collection meant that we were unable to undertake a meta-analysis; however, the range of methods and types of data collection is useful in guiding future web-based interventions targeting this population group and helping decision-makers understand the rationale for how interventions are expected to work and what enablers keep participants engaged to ultimately achieve the intended outcomes.

There is a need for a greater understanding of the individual-level, real-world factors affecting engagement in home and minimal contact practice interventions to ascertain how participants experience interventions and how this relates to their outcomes [19]. Exit and follow-up interviews can provide a deeper understanding of participants' experiences to strengthen real-life sustained engagement in that modality. Experience of an intervention needs to be user-friendly, accessible, and positive, which should be considered in promoting material that is most effective and helpful for users to engage from the outset.

Limitations of This Systematic Review

As the studies in this review were diverse in terms of study design, therapeutic intervention approach and delivery, length of follow-up, and outcome measures, we summarized the engagement data using the CAPE framework but were unable to perform a meta-analysis of the data. Attrition rates were high in many studies, and the number of participants was small, particularly in some pilot studies. We were unable to report this in terms of increasing our understanding of sustained practice as there was limited follow-up in most studies. There are inconsistent reports and terminology regarding engagement behavior. Inconsistencies in language between studies and interchangeability of terms, for example, attrition, withdrawal, dropout, and loss to follow-up, make direct comparison and systematic analysis challenging. Another potential limitation of this review is the lack of inclusion of studies in languages other than English. In addition, the protracted nature of systematic reviews means that the original search was concluded in 2020 and was affected by delays because of the COVID-19 pandemic. Since then, additional studies may have been published and not included in this review but would not necessarily affect our general conclusions or implications for using the logic model or reporting matrix.

Strengths and Future Work

The ability to leverage several frameworks enhanced this systematic review. The SWiM guidelines, part of the Cochrane methods, directed our synthesis and reporting. In addition, the CAPE framework provided an evidence-based approach to reporting on intervention engagement; using this framework, we were able to propose clear metrics for future reporting. It is recommended that future research provide engagement analytics to more clearly delineate between study and intervention compliance, particularly longer-term enactment or sustained practice to reflect pragmatic efficacy. The research team has a strong focus on research translation; therefore, the incorporation of a logic model provides a clear pathway for decision-makers, such as policy makers and commissioners, to interpret and guide the key constructs and evaluation metrics in future digital interventions in this field of research.

There is substantial evidence that psychological programs delivered on the web can be effective in treating and preventing mental health problems; however, the uptake of these programs can be suboptimal, and there remains a lack of evidence on how to increase engagement with evidence-based programs [45]. Poor adherence is a common feature of web-based mental health programs, which affects intervention outcomes [45] and limits real-world efficacy. Eisenstadt et al [42] discussed in their recent systematic review that adherence and retention continue to be challenges to the quality of research, with little or no information about reasons for dropouts given across studies. Further research is needed to unpack the key constructs of experience, including microlevel reporting and qualitative, phenomenological investigation via one-to-one postprogram interviews. Future reporting of DMHI using the CAPE framework could be used to ascertain the cost-benefit of an intervention; that is, if the conversion, recruitment, retention, and participation rates are high, the intervention is likely to be more cost-effective. However, this must be considered alongside the efficacy of the intervention and real-world application. The motivation for engaging in research studies is very different from real-world engagement experiences.

Advances in technology, particularly the internet, have proven to be an effective tool for building individual skills as it is inexpensive and accessible, both geographically and temporally. Despite promising results, internet interventions are still not widely disseminated or well-integrated into health services; successfully doing so will, in part, depend on engagement. As mental health apps have proliferated, choosing among them has become increasingly challenging for not only patients but also clinicians [46]. To address this, we need to understand the barriers and enablers for the delivery and sustainability of internet interventions in practice [17], as well as how we can engage not only participants but also health practitioners to support and disseminate effective interventions. This increased understanding will enable appropriate investment, optimization, and uptake of targeted well-being programs, such as those developed for perinatal women, with the ultimate aim of preventing poor mental health among women and their children.

Conclusions

To invest in accessible, long-term, sustainable health solutions, researchers, policy makers, and clinicians must identify optimal interventions that can be targeted to help specific risk groups or in specific contexts. Advances in technology, particularly the internet, have proven to be an effective tool for building individual skills as it is inexpensive and widely accessible. Pregnancy and the postnatal period can be times of increased psychological distress; therefore, it is an optimal time to intervene with strength-based tools to build affirmative self-efficacy. Although several studies in this field demonstrate efficacy, few robustly explore the construct of engagement, and in particular, there is limited evidence of the long-term enactment of the strategies learned. Our results indicate a disparity in the reporting of short- and long-term participant engagement behaviors, and we recommend the adoption of standardized metrics for reporting DMHI engagement in both research and real-world settings. This systematic review provides a framework for understanding the pathways for enhancing the mental well-being of mothers and their infants. With the world experiencing an endemic escalation in poor mental health across the life course, both in low- and high-income countries [55], it is imperative that we create practical, evidence-based, cost-effective, and scalable solutions to protect current and future generations.

Acknowledgments

This study was funded by the University of Western Australia, Telethon Kids Institute, the 100 Women Grant.

The authors would like to thank the University of Western Australia, Telethon Kids Institute, and the ORIGINS Project, which is funded by the Paul Ramsay Foundation and the Commonwealth Government of Australia through the Channel 7 Telethon Trust.

Data Availability

Data presented in this study are available on reasonable request from the corresponding authors.

Authors' Contributions

JAD, JLO, and ALF-J were involved in the conceptualization of the study. JAD and LYG performed the analysis. JAD wrote the original draft. Review and approval of the manuscript were conducted by JAD, LYG, ALF-J, JLO, and SLP. Funding support was acquired by ALF-J, JLO, and SLP.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and study selection.
[DOCX File , 33 KB - jmir_v24i8e36620_app1.docx]

Multimedia Appendix 2
Detailed study characteristics.
[DOCX File , 52 KB - jmir_v24i8e36620_app2.docx]

References

1. Law KH, Jackson B, Guelfi K, Nguyen T, Dimmock JA. Understanding and alleviating maternal postpartum distress: perspectives from first-time mothers in Australia. Soc Sci Med 2018 May;204:59-66. [doi: 10.1016/j.socscimed.2018.03.022] [Medline: 29579481]
2. Mental Health and Substance Use. World Health Organization. 2021. URL: https://www.who.int/teams/mental-health-and-substance-use/overview [accessed 2022-01-20]
3. Perinatal depression: data from the 2010 Australian National Infant Feeding Survey. Australian Institute of Health and Welfare. 2012. URL: https://www.aihw.gov.au/reports/primary-health-care/perinatal-depression-data-from-the-2010-australia/summary [accessed 2022-01-20]
4. Black MM, Walker SP, Fernald LC, Andersen CT, DiGirolamo AM, Lu C, Lancet Early Childhood Development Series Steering Committee. Early childhood development coming of age: science through the life course. Lancet 2017 Jan 07;389(10064):77-90 [FREE Full text] [doi: 10.1016/S0140-6736(16)31389-7] [Medline: 27717614]
5. Australia's health 2018. Australian Institute of Health and Welfare. 2018 Jun 20. URL: https://www.aihw.gov.au/reports/australias-health/australias-health-2018/contents/table-of-contents [accessed 2022-01-20]
6. Stone SL, Diop H, Declercq E, Cabral HJ, Fox MP, Wise LA. Stressful events during pregnancy and postpartum depressive symptoms. J Womens Health (Larchmt) 2015 May;24(5):384-393 [FREE Full text] [doi: 10.1089/jwh.2014.4857] [Medline: 25751609]
7. Wolford E, Lahti M, Tuovinen S, Lahti J, Lipsanen J, Savolainen K, et al. Maternal depressive symptoms during and after pregnancy are associated with attention-deficit/hyperactivity disorder symptoms in their 3- to 6-year-old children. PLoS One 2017 Dec 21;12(12):e0190248 [FREE Full text] [doi: 10.1371/journal.pone.0190248] [Medline: 29267405]
8. Barker DJ. The fetal and infant origins of adult disease. BMJ 1990 Nov 17;301(6761):1111 [FREE Full text] [doi: 10.1136/bmj.301.6761.1111] [Medline: 2252919]
9. Dubos RJ. Man Adapting. New Haven, CT, USA: Yale University Press; 1965.
10. Lönnberg G, Jonas W, Bränström R, Nissen E, Niemi M. Long-term effects of a mindfulness-based childbirth and parenting program—a randomized controlled trial. Mindfulness 2020 May 13;12(2):476-488. [doi: 10.1007/s12671-020-01403-9]
11. Monshat K, Vella-Brodrick D, Burns J, Herrman H. Mental health promotion in the Internet age: a consultation with Australian young people to inform the design of an online mindfulness training programme. Health Promot Int 2012 Jun;27(2):177-186. [doi: 10.1093/heapro/dar017] [Medline: 21398335]
12. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
13. Ashford MT, Olander EK, Ayers S. Computer- or Web-based interventions for perinatal mental health: a systematic review. J Affect Disord 2016 Jun;197:134-146. [doi: 10.1016/j.jad.2016.02.057] [Medline: 26991368]

14. Drozd F, Haga SM, Brendryen H, Slinning K. An Internet-based intervention (Mamma Mia) for postpartum depression: mapping the development from theory to practice. JMIR Res Protoc 2015 Oct 12;4(4):e120 [FREE Full text] [doi: 10.2196/resprot.4858] [Medline: 26476481]
15. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
16. Davis JA, Gibson LY, Bear NL, Finlay-Jones AL, Ohan JL, Silva DT, et al. Can positive mindsets be protective against stress and isolation experienced during the COVID-19 pandemic? A mixed methods approach to understanding emotional health and wellbeing needs of perinatal women. Int J Environ Res Public Health 2021 Jun 29;18(13):6958 [FREE Full text] [doi: 10.3390/ijerph18136958] [Medline: 34209693]
17. Drozd F, Haga SM, Lisøy C, Slinning K. Evaluation of the implementation of an internet intervention in well-baby clinics: a pilot study. Internet Interv 2018 Sep;13:1-7 [FREE Full text] [doi: 10.1016/j.invent.2018.04.003] [Medline: 30206512]
18. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
19. Parsons CE, Crane C, Parsons LJ, Fjorback LO, Kuyken W. Home practice in mindfulness-based cognitive therapy and mindfulness-based stress reduction: a systematic review and meta-analysis of participants' mindfulness practice and its association with outcomes. Behav Res Ther 2017 Aug;95:29-41 [FREE Full text] [doi: 10.1016/j.brat.2017.05.004] [Medline: 28527330]
20. Vettese LC, Toneatto T, Stea JN, Nguyen L, Wang JJ. Do mindfulness meditation participants do their homework? And does it make a difference? A review of the empirical evidence. J Cogn Psychother 2009 Aug 01;23(3):198-225. [doi: 10.1891/0889-8391.23.3.198]
21. Ingoldsby EM. Review of interventions to improve family engagement and retention in parent and child mental health programs. J Child Fam Stud 2010 Oct 01;19(5):629-645 [FREE Full text] [doi: 10.1007/s10826-009-9350-2] [Medline: 20823946]
22. Panter-Brick C, Burgess A, Eggerman M, McAllister F, Pruett K, Leckman JF. Practitioner review: engaging fathers--recommendations for a game change in parenting interventions based on a systematic review of the global evidence. J Child Psychol Psychiatry 2014 Nov;55(11):1187-1212 [FREE Full text] [doi: 10.1111/jcpp.12280] [Medline: 24980187]
23. Haine-Schlagel R, Walsh NE. A review of parent participation engagement in child and family mental health treatment. Clin Child Fam Psychol Rev 2015 Jun;18(2):133-150 [FREE Full text] [doi: 10.1007/s10567-015-0182-x] [Medline: 25726421]
24. May CD, St George JM, Lane S. From presence to participation: engagement with an SMS program for fathers of children on the autism spectrum. J Child Fam Stud 2020 Nov 07;30(1):29-37. [doi: 10.1007/s10826-020-01845-8]
25. Piotrowska PJ, Tully LA, Lenroot R, Kimonis E, Hawes D, Moul C, et al. Mothers, fathers, and parental systems: a conceptual model of parental engagement in programmes for child mental health-connect, attend, participate, enact (CAPE). Clin Child Fam Psychol Rev 2017 Jun;20(2):146-161 [FREE Full text] [doi: 10.1007/s10567-016-0219-9] [Medline: 27914017]
26. Anderson LM, Petticrew M, Rehfuess E, Armstrong R, Ueffing E, Baker P, et al. Using logic models to capture complexity in systematic reviews. Res Synth Methods 2011 Mar;2(1):33-42. [doi: 10.1002/jrsm.32] [Medline: 26061598]
27. Petticrew M, Roberts H. Systematic Reviews in the Social Sciences: A Practical Guide. Hoboken, NJ, USA: Wiley Online Library; Jan 1, 2006.
28. 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]
29. Parkinson S, Eatough V, Holmes J, Stapley E, Midgley N. Framework analysis: a worked example of a study exploring young people’s experiences of depression. Qual Res Psychol 2016;13(2):109-129. [doi: 10.1080/14780887.2015.1119228]
30. Baxter SK, Blank L, Woods HB, Payne N, Rimmer M, Goyder E. Using logic model methods in systematic review synthesis: describing complex pathways in referral management interventions. BMC Med Res Methodol 2014 May 10;14:62 [FREE Full text] [doi: 10.1186/1471-2288-14-62] [Medline: 24885751]
31. Cochrane Methods Bias. The Cochrane Collection. 2021. URL: https://methods.cochrane.org/bias/ [accessed 2022-01-20]
32. Corno G, Espinoza M, Maria Baños R. A narrative review of positive psychology interventions for women during the perinatal period. J Obstet Gynaecol 2019 Oct;39(7):889-895. [doi: 10.1080/01443615.2019.1581735] [Medline: 31179814]
33. Felder JN, Segal Z, Beck A, Sherwood NE, Goodman SH, Boggs J, et al. An open trial of Web-based mindfulness-based cognitive therapy for perinatal women at risk for depressive relapse. Cogn Behav Pract 2017 Feb;24(1):26-37. [doi: 10.1016/j.cbpra.2016.02.002]
34. Tsai YJ, Hsu YY, Hou TW, Chang CH. Effects of a Web-based antenatal care system on maternal stress and self-efficacy during pregnancy: a study in Taiwan. J Midwifery Womens Health 2018 Mar;63(2):205-213. [doi: 10.1111/jmwh.12685] [Medline: 29533525]
35. Ayers S, Fitzgerald G, Thompson S. Brief online self-help exercises for postnatal women to improve mood: a pilot study. Matern Child Health J 2015 Nov;19(11):2375-2383. [doi: 10.1007/s10995-015-1755-5] [Medline: 26084670]

36. Barrera AZ, Wickham RE, Muñoz RF. Online prevention of postpartum depression for Spanish- and English-speaking pregnant women: a pilot randomized controlled trial. Internet Interv 2015 Sep 01;2(3):257-265 [FREE Full text] [doi: 10.1016/j.invent.2015.06.002] [Medline: 26273567]
37. Barrera AZ, Aguilera A, Inlow N, Servin J. A preliminary study on the acceptability of a brief SMS program for perinatal women. Health Informatics J 2020 Jun;26(2):1079-1087 [FREE Full text] [doi: 10.1177/1460458219866560] [Medline: 31566457]
38. Cornsweet Barber C, Clark M, Williams S, Isler R. Relaxation and mindfulness to manage computerised self-help programme. MIDIRS Midwifery Digest 2013;23:449-454.
39. Duffecy J, Grekin R, Hinkel H, Gallivan N, Nelson G, O'Hara MW. A group-based online intervention to prevent postpartum depression (Sunnyside): feasibility randomized controlled trial. JMIR Ment Health 2019 May 28;6(5):e10778 [FREE Full text] [doi: 10.2196/10778] [Medline: 31140443]
40. Fonseca A, Monteiro F, Alves S, Gorayeb R, Canavarro MC. Be a mom, a Web-based intervention to prevent postpartum depression: the enhancement of self-regulatory skills and its association with postpartum depressive symptoms. Front Psychol 2019 Feb 18;10:265 [FREE Full text] [doi: 10.3389/fpsyg.2019.00265] [Medline: 30873060]
41. Matvienko-Sikar K, Dockray S. Effects of a novel positive psychological intervention on prenatal stress and well-being: a pilot randomised controlled trial. Women Birth 2017 Apr;30(2):e111-e118. [doi: 10.1016/j.wombi.2016.10.003] [Medline: 27810284]
42. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 08;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]
43. Terides MD, Dear BF, Fogliati VJ, Gandy M, Karin E, Jones MP, et al. Increased skills usage statistically mediates symptom reduction in self-guided internet-delivered cognitive-behavioural therapy for depression and anxiety: a randomised controlled trial. Cogn Behav Ther 2018 Jan;47(1):43-61. [doi: 10.1080/16506073.2017.1347195] [Medline: 28724338]
44. Strunk DR, Hollars SN, Adler AD, Goldstein LA, Braun JD. Assessing patients' cognitive therapy skills: initial evaluation of the competencies of cognitive therapy scale. Cognit Ther Res 2014 Oct;38(5):559-569 [FREE Full text] [doi: 10.1007/s10608-014-9617-9] [Medline: 25408560]
45. Batterham PJ, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Gulliver A. A brief intervention to increase uptake and adherence of an online program for depression and anxiety: protocol for the Enhancing Engagement with Psychosocial Interventions (EEPI) randomized controlled trial. Contemp Clin Trials 2019 Mar;78:107-115. [doi: 10.1016/j.cct.2019.01.015] [Medline: 30711664]
46. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digit Health 2019 Jun;1(2):e52-e54 [FREE Full text] [doi: 10.1016/S2589-7500(19)30013-5] [Medline: 33323229]
47. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.rstudio.com/ [accessed 2022-07-04]
48. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Cham, Switzerland: Springer; 2016.
49. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
50. Chan KL, Leung WC, Tiwari A, Or KL, Ip P. Using smartphone-based psychoeducation to reduce postnatal depression among first-time mothers: randomized controlled trial. JMIR Mhealth Uhealth 2019 May 14;7(5):e12794 [FREE Full text] [doi: 10.2196/12794] [Medline: 31094354]
51. Shorey S, Chee CY, Ng ED, Lau Y, Dennis CL, Chan YH. Evaluation of a technology-based peer-support intervention program for preventing postnatal depression (part 1): randomized controlled trial. J Med Internet Res 2019 Aug 29;21(8):e12410 [FREE Full text] [doi: 10.2196/12410] [Medline: 31469084]
52. Gammer I, Hartley-Jones C, Jones FW. A randomized controlled trial of an online, compassion-based intervention for maternal psychological well-being in the first year postpartum. Mindfulness 2020 Jan 17;11(4):928-939. [doi: 10.1007/s12671-020-01306-9]
53. Guo L, Zhang J, Mu L, Ye Z. Preventing postpartum depression with mindful self-compassion intervention: a randomized control study. J Nerv Ment Dis 2020 Feb;208(2):101-107. [doi: 10.1097/NMD.0000000000001096] [Medline: 31868776]
54. Krusche A, Dymond M, Murphy SE, Crane C. Mindfulness for pregnancy: a randomised controlled study of online mindfulness during pregnancy. Midwifery 2018 Oct;65:51-57. [doi: 10.1016/j.midw.2018.07.005] [Medline: 30099285]
55. Mental Health: burden. World Health Organization. 2021 Nov 26. URL: https://www.who.int/health-topics/mental-health#tab=tab_2 [accessed 2022-01-20]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CAPE: Connect, Attend, Participate, and Enact
DMHI: digital mental health intervention

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
SWiM: Synthesis Without Meta-analysis

Edited by T Leung; submitted 20.01.22; peer-reviewed by A Gemmill, A Khaity, K Fadahunsi; comments to author 10.04.22; revised version received 22.05.22; accepted 31.05.22; published 09.08.22.

Please cite as:
Davis JA, Ohan JL, Gibson LY, Prescott SL, Finlay-Jones AL
Understanding Engagement in Digital Mental Health and Well-being Programs for Women in the Perinatal Period: Systematic Review Without Meta-analysis
J Med Internet Res 2022;24(8):e36620
URL: https://www.jmir.org/2022/8/e36620
doi:10.2196/36620
PMID:35943773

Review

mHealth Interventions to Improve Cancer Screening and Early Detection: Scoping Review of Reviews

Désirée Schliemann[1], PhD; Min Min Tan[2,3], PhD; Wilfred Mok Kok Hoe[2,3], PhD; Devi Mohan[3], MD; Nur Aishah Taib[4], PhD; Michael Donnelly[1], PhD; Tin Tin Su[1,2,3], PhD

[1]Centre for Public Health, Queen’s University Belfast, Belfast, United Kingdom

[2]South East Asia Community Observatory (SEACO), Monash University Malaysia, Subang Jaya, Malaysia

[3]Global Public Health, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Subang Jaya, Malaysia

[4]Department of Surgery, Faculty of Medicine, University Malaya Cancer Research Institute, University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author:
Désirée Schliemann, PhD
Centre for Public Health
Queen’s University Belfast
Institute of Clinical Science, Block B
Royal Victoria Hospital
Belfast, BT12 6BA
United Kingdom
Phone: 44 02890245133
Email: d.schliemann@qub.ac.uk

Abstract

Background: Cancer screening provision in resource-constrained settings tends to be opportunistic, and uptake tends to be low, leading to delayed presentation and treatment and poor survival.

Objective: The aim of this study was to identify, review, map, and summarize findings from different types of literature reviews on the use of mobile health (mHealth) technologies to improve the uptake of cancer screening.

Methods: The review methodology was guided by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews). Ovid MEDLINE, PyscINFO, and Embase were searched from inception to May 2021. The eligible criteria included reviews that focused on studies of interventions that used mobile phone devices to promote and deliver cancer screening and described the effectiveness or implementation of mHealth intervention outcomes. Key data fields such as study aims, types of cancer, mHealth formats, and outcomes were extracted, and the data were analyzed to address the objective of the review.

Results: Our initial search identified 1981 titles, of which 12 (0.61%) reviews met the inclusion criteria (systematic reviews: n=6, 50%; scoping reviews: n=4, 33%; rapid reviews: n=1, 8%; narrative reviews: n=1, 8%). Most (57/67, 85%) of the interventions targeted breast and cervical cancer awareness and screening uptake. The most commonly used mHealth technologies for increasing cancer screening uptake were SMS text messages and telephone calls. Overall, mHealth interventions increased knowledge about screening and had high acceptance among participants. The likelihood of achieving improved uptake-related outcomes increased when interventions used >1 mode of communication (telephone reminders, physical invitation letters, and educational pamphlets) together with mHealth.

Conclusions: mHealth interventions increase cancer screening uptake, although multiple modes used in combination seem to be more effective.

(J Med Internet Res 2022;24(8):e36316) doi:10.2196/36316

KEYWORDS

mobile health; mHealth; cancer screening; scoping review of reviews; cancer; cancer detection; oncology; digital health; scoping review; review; mobile phone

Introduction

Background

Globally, cancer is the second leading cause of death; it accounted for approximately 9.6 million deaths in 2018 [1]. Cancer incidence and mortality are predicted to increase to 30.2 million cases and 16.3 million deaths by 2040, respectively, because of aging populations and the adoption of unhealthy lifestyles [2]. Delay between symptom onset and treatment leads to poorer cancer survival [3]. Screening increases the chance of early detection and treatment and, ultimately, survival. In many high-income countries, population-based cancer screening is available for four common cancers and has contributed to reduced breast cancer [4], cervical cancer [5], prostate cancer [6], and colorectal cancer [7] mortality. However, cancer screening in the majority of low- and middle-income countries (LMICs) is opportunistic, and uptake is low compared with cancer screening in high-income countries, leading to delayed presentation, treatment, and survival [8]; for example, the uptake of mammogram screening was 12% to 31% in Brazil [9] and 7% to 25% in Malaysia [10] compared with 66% in Germany [11] and 75% in Spain [12]. Low uptake of cancer screening might indicate poor awareness and knowledge of cancer and cancer screening among the public; for example, Asian Pacific populations with the lowest uptake of colorectal cancer screening, such as India, Malaysia, Indonesia, Pakistan, and Brunei, had correspondingly low levels of awareness and knowledge of colorectal cancer symptoms, risk factors, and screening tests [13]. Poor knowledge about, and negative perceptions toward, mammogram screening are major barriers to mammogram screening uptake in Malaysia [10].

Digital health care, that is, the use of digital technologies for health, is now commonly used in public health care as well as primary health care [14]. According to the World Health Organization Global Observatory for eHealth, *mobile health* (mHealth) is defined as “medical and public health practice supported by the use of mobile devices” such as mobile phones, smartphones, and tablet computers [15]. Worldwide, there are approximately 5.3 billion unique mobile phone users, representing 67.1% of the total population, and smartphones account for approximately 75% of the mobile phones in use [16]. The high penetration rate of mobile phones allows timely data collection as well as transmission and analysis of the data. Thus, mHealth holds great potential for improving health outcomes because of its mobility, instantaneous access, and ease of use. Some of the common mHealth apps offer patient education and behavior change communication, data collection and reporting, population health registries and vital event tracking, and electronic health records, as well as provider training and education [17]. mHealth interventions have a positive impact on clinical outcomes, adherence to treatment and care, health behavior changes, disease management, and primary care attendance rates with regard to various diseases [18]. mHealth has also been used in cancer self-care and self-management among cancer survivors to improve sleep and quality of life; reduce fatigue, stress, and pain; and promote health behaviors such as weight loss [19-22]. The role of mHealth in promoting cancer screening has been explored in different types of reviews. However, it is unknown whether similar findings are observed across the reviews.

Objectives

This scoping review aimed to map and summarize findings from systematic, scoping, narrative, and rapid reviews on the use of mHealth in cancer screening, as well as other screening-related outcomes such as attitudes toward screening and knowledge and awareness of screening. We also included implementation considerations for successful mHealth interventions in improving cancer screening uptake and screening-related outcomes.

Methods

Overview

This scoping review of reviews was conducted based on the framework of Arksey and O’Malley [23] and using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [24]. The protocol of this review has not been preregistered. As the use of mHealth in relation to cancer screening is a relatively nascent field of study, a scoping review is useful in mapping the published literature comprehensively and systematically. The review was guided by the following 5-step framework: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results.

Search Strategy

We first searched Ovid MEDLINE, PyscINFO, and Embase for relevant literature on February 1, 2021, using two categories of key terms: mHealth and early detection of cancer. We then refined the search on May 17, 2021. The key terms were based on Medical Subject Headings indexing as well as free-text terms. We combined key terms from the same category with OR and between categories with AND. The search strategy was developed in Ovid MEDLINE (Multimedia Appendix 1) and adapted for the other databases. We also hand searched the reference lists of selected reviews for relevant reviews. All searches were exported into EndNote (Clarivate), and duplicates were removed.

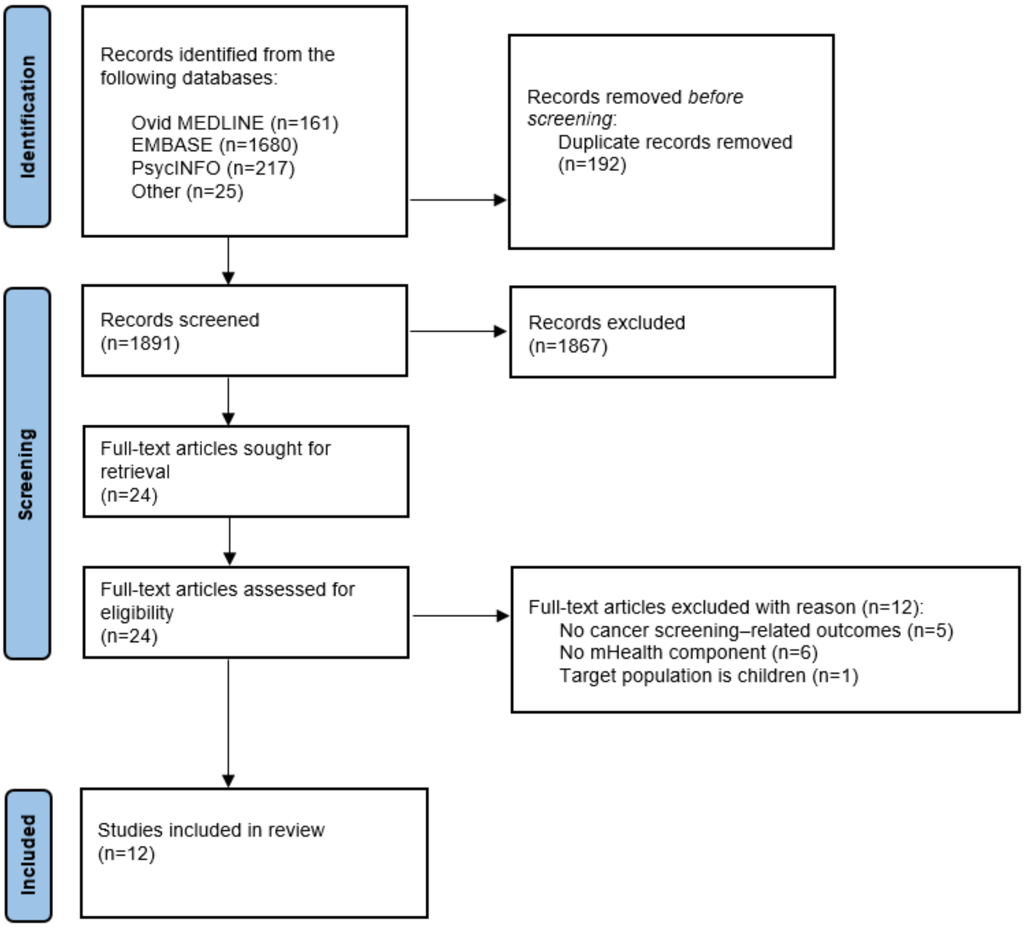
Inclusion Criteria

Papers were included if they satisfied all of the following criteria: (1) a review of any type, (2) the reviewed interventions related to cancer screening (for any cancer type) that were conducted on mobile devices such as mobile phones and tablet computers, (3) described the effectiveness and implementation of mHealth interventions on outcomes related to cancer screening, (4) included adults aged ≥18 years from the general population, and (5) published in English in peer-reviewed journals from inception up to May 2021. We excluded reviews that did not specify the use of mobile technologies but instead reported modes of delivery such as web-based and computer-delivered programs and videos.

Selection of Reviews

We selected the relevant reviews using a 3-stage process: (1) MMT and WMKH conducted the initial screening of titles and abstracts to determine eligibility for inclusion; (2) WMKH retrieved full texts, which were screened by DS, MMT, and WMKH independently for inclusion, with discrepancies resolved through discussion with DS, MMT, and WMKH; and (3) MMT and WMKH extracted relevant data. The screening process is provided in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (Figure 1).

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



Data Extraction and Charting

The following data were extracted by MMT and WMKH from each selected review into an Excel (Microsoft Corporation) spreadsheet:

- Review identifiers (author, year, country, type of review, number of studies, time range, intervention duration, and follow-up duration)
- Study aim
- Types of cancer mentioned in the relevant studies in the review
- Types of mHealth mentioned in the relevant studies in the review
- Details of intervention procedures
- Outcome measures (awareness, knowledge, or attitude; screening uptake; and implementation-related outcomes)
- Key stakeholders in delivering the mHealth intervention, if any

If the aforementioned data were not reported in the selected reviews, we referred to the individual studies included in the selected reviews. For reviews that included studies that focused on mHealth and studies that did not, we only extracted information specifically reported on the studies that included mHealth. Information related to the quality of the reviews was not assessed.

Results

Literature Search

Our initial search identified 2083 citations, resulting in 1981 (95.1%) unique citations after removal of duplicates (Figure 1). The titles and abstracts were assessed based on the inclusion criteria, and of the 1981 unique citations, 24 (1.21%) were included for full-text screening. Of these 24 reviews, 12 (50%) were excluded after the full-text screen: 5 (42%) did not include cancer screening-related outcomes, 6 (50%) did not include mHealth components, and 1 (8%) included children as their target population. Hence, of the 24 reviews included for full-text screening, 12 (50%) were included in this scoping review. Table 1 summarizes the characteristics of the included reviews.

Table 1. Characteristics of included reviews.

Study	Type of review	Aim	Time frame of search strategy	Total number of studies; number of relevant studies[a] by cancer type	Type of mHealth[b] in relevant studies[a]	Key stakeholders delivering mHealth interventions
Bhochhibhoya et al [25], 2020	Scoping review	To identify studies that examined mHealth programs that focused on increasing cervical cancer screening among women to determine if these interventions improved adherence to screening and what factors (barriers and facilitators) were most influential among participants	January 1, 2009, to September, 28, 2019	12; cervical cancer (n=12, 3 of which were qualitative studies)	• Telephone appointment (n=1) • Telephone reminder with tailored counseling versus telephone reminder with print materials (n=1) • SMS text message reminders (n=3) • 15 behavior change messages with transportation e-voucher versus SMS text messages of location and hours of the closest screening clinic (n=1) • Automated SMS text messages or telephone call reminders (n=1) • Automated SMS text messages versus telephone call reminders+manual telephone call+face-to-face interview (n=1) • 3 sequential SMS text message reminders, followed by 3 telephone call attempts (n=1)	• Telephone appointment by midwife (n=1) • Telephone counseling and reminders by research staff (n=1) • Invitation telephone call by clinical secretaries (n=1) • Telephone caller unspecified (n=1)
Uy et al [26], 2017	Systematic review	To assess the effect of SMS text messaging interventions on increasing patient adherence to screening for breast, cervical, colorectal, and lung cancers	January 2000 to January 2017	9; breast (n=5), cervical (n=1), and colorectal (n=3) cancers	• SMS text message reminder (n=5) • SMS text message reminder plus letter (n=4)	• None
Zhang et al [27], 2020	Systematic review	To qualitatively synthesize published articles reporting the impact of mHealth on cervical cancer screening–related health behaviors	Up to October 10, 2019	8 (1 cross-sectional study); cervical cancer (n=7)	• Invitation letter with pamphlet, followed by telephone reminder (n=1) • Educational SMS text messages (n=2) • Educational SMS text message versus SMS text message reminder (n=1) • Educational SMS text messages with transportation e-voucher versus SMS text messages of location and hours of the closest screening clinic (n=1) • Motivational interview over the telephone (n=1) • Training through SMS text message, electronic posters, infographics, podcasts, and video tutorial (n=1)	• Motivational interview through telephone call by nurses (n=1) • Telephone caller unspecified (n=1)

Study	Type of review	Aim	Time frame of search strategy	Total number of studies; number of relevant studies[a] by cancer type	Type of mHealth[b] in relevant studies[a]	Key stakeholders delivering mHealth interventions
Halake and Ogoncho [28], 2017	Scoping review	To establish the extent and nature of the published and gray literature on the use of mHealth-based technologies for cancer prevention, detection, and management in LMICs[c]	1990 to 2014	15; breast cancer (n=2)	• SMS text message invitation and cancer screening information (n=1) • Smartphone app to facilitate BSEd (n=1)	• Not described
Choi et al [29], 2018	Systematic review	To investigate recent research trends related to the use of mobile technology in the prevention and management of skin cancer, focusing on how such technology is evaluated and what impact it has in each phase across the cancer continuum	January 1, 2007, to December 31, 2017	18; skin cancer (n=1)	• Educational SMS text message about skin self-examination (n=1)	• Not described
Houghton et al [30], 2019	Systematic review	To determine how mobile apps are being used for breast cancer prevention among women across the cancer control continuum	Up to February 7, 2019	69; breast cancer (n=4)	• Mammopad, a decision aid mobile app on iPad Mini (n=1) • Mobile app to assist navigator (n=1) • mMammogram mobile app for SMS text message (n=1) • Mobile app for BSE (n=1)	• Mobile apps paired with community health navigators (n=2)
Plackett et al [31], 2020	Scoping review	To map the evidence for social media interventions to improve cancer screening and early diagnosis, including behavior change, and how the interventions facilitate behavior change	2004 to June 2019	23; breast (n=4) and cervical (n=1) cancers	• Facebook (n=3) • Snapchat (n=1)	• Not described
Musa et al [32], 2017	Systematic review and meta-analysis	To review the evidence of the effectiveness of provider recommendations for cervical cancer screening on screening rates in women at risk for cervical cancer	Up to August 2016	28; cervical cancer (n=5)	• Telephone counseling (n=1) • SMS text message or telephone reminder (n=1) • Email, telephone, or multimodal (letter+email+telephone) screening reminder and invitation and education flyer (n=1) • Telephone reminder (n=1) • Invitation letter and information pamphlet, followed by telephone reminder with counseling (n=1)	• Telephone caller unspecified (n=3) • Telephone counseling by health educator (n=1) • Telemarketing company (n=1)

Study	Type of review	Aim	Time frame of search strategy	Total number of studies; number of relevant studies[a] by cancer type	Type of mHealth[b] in relevant studies[a]	Key stakeholders delivering mHealth interventions
Duffy et al [33], 2017	Rapid review	To review the current evidence on effects of interventions to improve cancer screening participation, focusing in particular on effects in underserved populations	Time frame not specified	68; breast (n=9), cervical (n=5), colorectal (n=2), and stomach (n=1) cancers	- Automated telephone and SMS text message reminders or telephone outreach (n=1) - Telephone reminder or motivational telephone call (n=1) - Telephone call to confirm receipt of invitation letter, followed by telephone reminder (n=1) - Telephone reminders (n=8) - SMS text message reminder (n=3) - Tailored telephone counseling (n=2) - Telephone appointment (n=1)	- Colorectal cancer screening navigator (n=1) - Bilingual advocate at a community organization with experience in telephone outreach (n=1) - Local women recruited by Community Links, a community charity (n=1) - Female scheduler and female counselors (n=1) - Female research assistants (n=1) - Telephone counselors (n=2) - Trained GP[e] receptionist (n=1) - Volunteers (n=1) - Researcher (n=1) - Research nurse (n=1) - Telemarketing company (n=1) - Telephone caller unspecified (n=2)
Lott et al [34], 2020	Scoping review	To map the literature on interventions to increase uptake of cervical screening in sub-Saharan Africa and identify opportunities for future intervention development and research	Up to 2019	19; cervical cancer (n=3)	- SMS text message reminders (n=1) - Telephone follow-up and counseling (n=1) - Email (n=1)	- Telephone counselors (n=1)
Déglise et al [35], 2012	Narrative review	To describe the characteristics and outcomes of SMS text messaging interventions for disease prevention in LMICs and provide recommendations for future work	Up to May 2011	17; breast cancer (n=1)	- SMS text message reminder	- Not described
Peiris et al [36], 2014	Systematic review	To critically appraise the role of mHealth in improving health care quality for NCDs[f] in LMICs	Up to May 2014	48; breast cancer (n=1)	- SMS text message reminder	- Not described

[a] Relevant studies are studies that met the inclusion criteria for this review; for example, some reviews included diseases other than cancer. We only reported results from the studies evaluating cancer-related interventions.

[b] mHealth: mobile health.

[c] LMICs: low- and middle-income countries.

[d] BSE: breast self-examination.

[e] GP: general practitioner.

[f] NCD: noncommunicable disease.

Characteristics of Reviews

The included reviews (n=12) were published between 2012 and 2020 (Table 1). Of the 12 reviews, 6 (50%) were systematic reviews [26,27,29,30,32,36], of which 1 (17%) also included a meta-analysis [32]; 4 (33%) were scoping reviews [25,28,31,34]; 1 (8%) was a rapid review [33]; and 1 (8%) was a narrative review [35]. The 12 reviews reported different outcomes of the studies that were relevant to this review of reviews (Table 2): 5 (42%) reported solely the effectiveness of mHealth interventions on cancer screening [26,29,32-34]; 4 (33%) reported outcomes in relation to cancer screening, change in cancer knowledge, and attitudes to screening [25,27,30,31]; 2 (17%) reported outcomes in relation to breast self-examination (BSE) practice [35,36]; and 1 reported outcomes in relation to BSE and cancer screening [28]. Most (7/12, 58%) of the reviews included studies that were conducted mainly in high-income Western countries [25,26,29-33], whereas 42% (5/12) focused on LMICs [27,28,34-36], of which 20% (1/5) focused solely on sub-Saharan Africa [34]. In total, 33% (4/12) of the reviews focused on cervical cancer [25,27,32,34]; 8% (1/12) focused on skin cancer [29]; 8% (1/12) focused on breast cancer [30]; 8% (1/12) examined breast, cervical, lung, and colorectal cancers [26]; 25% (3/12) included any type of cancer [28,31,33]; and 2 reviews focused on disease prevention in general [35,36]. In terms of interventions, 42% (5/12) of the reviews included interventions of various types of mHealth technologies [25,27-29,36], 2 (33%) focused solely on SMS text messages [26,35], 1 (17%) focused on social media interventions [31], 1 (17%) was specifically about mobile apps [30], and 25% (3/12) included any type of communication (mHealth, face-to-face, and other media) [32-34].

Table 2. Summary of screening-related outcomes extracted from each review.

Study	Outcomes		
	Screening outcomes	Screening awareness-, knowledge-, and attitude-related outcomes	Implementation-related outcomes and measures
Bhochhibhoya et al [25], 2020	• Screening uptake: n=5/6[a] (9.1%-17.9% increase between intervention group versus control group; 9.3% increase after the intervention compared with before)[b] • Screening follow-up adherence: n=0/1 • Effective methods: stepwise approach (automated telephone calls and SMS text messages, followed by manual telephone call and face-to-face interview), SMS text messages only, telephone call only, telephone appointment by midwives, telephone reminders combined with other methods such as tailored counseling, and SMS text message with transportation e-voucher	• Knowledge improvement: n=2/2 • Attitude about screening: n=1/2 • Perceived behavior control: n=0/1 • Perceived barriers about screening: n=0/1 • Belief about screening: n=1/1 • Screening intention: n=0/2 • Effective methods: health-specific and spiritually based SMS text messages and personally tailored texts with statistical facts	• Advantages: convenient, time effective, ease of use, and able to receive notification • Concerns: confidentiality of SMS text messages, loss of the mobile phones, clarity of the language used, and receiving negative results through SMS text message • Barriers: inconvenient for older participants, lack of texting proficiency, difficulty in texting, and apprehension that SMS text messages might not be clearly understood • Enabling factors: contact preferences, cell phone ownership, and portability of same number • Enhancing factors: message content (reminder and informative) and short and simple messaging formats
Uy et al [26], 2017	• Screening uptake: n=5/9 (1.2%-9.9% absolute increase) • Effective methods: SMS text message reminder+letter and single SMS text message reminder	—[c]	—
Zhang et al [27], 2020	• Screening uptake: n=3/5 (12.9%-50.9% increase) • Screening follow-up: n=1/1 (91.8%-93.5%; OR[d] 1.37-1.40) • Effective methods: SMS text message with transportation e-voucher, invitation letter with telephone reminder, reminders sent through letter, registered letter, SMS text message or telephone call, and telephone reminders or educational telephone call	• Knowledge improvement: n=1/2 • Perceived benefits of Pap[e] test: n=1/1 • Reduced barriers to undergoing Pap smear: n=1/1 • Attitude about screening: n=0/1 • Effective methods: a combination of SMS text message, electronic posters, infographics, podcasts, and video tutorials	• Interest in receiving screening test results through SMS text message: n=0/1 • Interest in receiving screening test results using nonprivate telephone: n=1/1 (OR 0.31, 95% CI 0.18-0.51) • Interest in receiving appointment reminders through SMS text message: n=1/1 (OR 14.19, 95% CI 1.72-117.13) • Interest in receiving appointment reminders using nonprivate telephone: n=0/1
Halake and Ogoncho [28], 2017	• Screening uptake: n=1/1 (30.7% and 31.6% increase) • BSE[f] practice: n=1/1 • Effective method: BSE smartphone app	—	—
Choi et al [29], 2018	• Screening uptake: n=1/1 (27% absolute increase in skin self-examination) • Effective method: educational SMS text messages with reminders	—	—

Study	Outcomes		
	Screening outcomes	Screening awareness-, knowledge-, and attitude-related outcomes	Implementation-related outcomes and measures
Houghton et al [30], 2019	• Screening uptake: n=3/3 • Effective methods: community health workers (trained or untrained in patient navigation) equipped with smartphone app plus standard risk counselling and mMammogram (SMS text messages plus health navigator)	• Knowledge improvement: n=2/2 (33% increase) • Reduced decisional conflict: n=1/1 • Self-efficacy: n=1/1 • Screening intention: n=0/1 • Screening readiness: n=1/1 • Effective methods: smartphone app plus standard risk counseling, mMammogram (SMS text messages plus health navigator), and smartphone app decision aid (Mammopad)	• Intervention satisfaction (mMammogram): n=1/1 • Effectiveness satisfaction (mMammogram): n=1/1
Plackett et al [31], 2020	• Screening uptake: n=1/1 (12.9% increase) • Effective method: breast cancer screening service Facebook page	• Knowledge improvement: n=2/2 • Screening intention: n=1/1 (82% increase) • Effective methods: Facebook or face-to-face discussions for 2 weeks after 50-minute classroom cervical cancer prevention education lecture (female high school students), receiving breast cancer awareness information through Snapchat, and tailored SMS text message mammography campaign on Facebook during Breast Cancer Awareness Month	• Using Facebook is acceptable for delivering breast cancer screening information: n=1/1
Musa et al [32], 2017	• Screening uptake: n=5/6 (7.8%-31.1% absolute increase) • Reduced screening median time: n=1/1 • Effective methods: direct invitation mail+brochure+telephone counseling by health educators; telephone reminder with educational information and multimodal intervention; invitation letter and information pamphlet, followed by telephone reminder with counseling; telephone reminder with educational information; and multimodal intervention	—	—
Duffy et al [33], 2017	• Screening uptake: n=13/16 (5%-45% absolute increase) • n=3/3, SMS text reminder studies; n=11/13, telephone reminder studies	—	—
Lott et al [34], 2020	• Screening uptake: n=2/3 (8.6% difference in screening uptake between control and intervention groups; 51% increase after the intervention) • Effective methods: SMS text message about cervical cancer and context-specific barriers to screening (and SMS text message plus e-voucher for transportation) and enhanced patient-centered counseling with patient follow-up by telephone (with or without escort to cervical cancer screening)	—	—
Déglise et al [35], 2012	• BSE practice: n=1/1 • Effective method: SMS text message reminder to conduct BSE	—	—

Study	Outcomes		
	Screening outcomes	Screening awareness-, knowledge-, and attitude-related outcomes	Implementation-related outcomes and measures
Peiris et al [36], 2014	• BSE practice: n=1/1 • Effective method: SMS text message reminder to conduct BSE	—	—

[a] Number of studies that reported a positive outcome out of the total number of studies that included the particular outcome.

[b] Percentage of change or odds ratios are included if available.

[c] Not available (ie, not reported).

[d] OR: odds ratio.

[e] Pap: Papanicolaou.

[f] BSE: breast self-examination.

Types of mHealth Interventions

SMS text messages were the most commonly used mHealth technology and were used in 46% (31/67) of the interventions. They were mainly delivered as reminders of cancer screening appointments, alone or in combination with telephone reminders, physical invitation letters, and educational pamphlets. Educational SMS text messages, sent as a one-off or in a series over days or weeks, were also widely used. Their contents included information about cancer risk factors, benefits of screening, location and operating hours of screening clinics, spiritually based health messages, and facts about cancer (eg, incidence, mortality, and screening rates). Educational SMS text messages were used alone or in combination with an e-voucher (to subsidize the cost of transportation to and from the screening facility) [37].

Text messages were most commonly sent as SMS text messages. In later studies, they were also sent through IP-based messaging services such as Telegram and Snapchat and mobile apps specifically designed for the interventions. In almost all (10/12, 83%) reviews, the delivery of SMS text messages was one-way, 8% (1/12) of the reviews reported an intervention that included a specifically designed mobile app (mMammogram) that featured personally tailored messages [38], and 8% (1/12) used social media for communication [31].

Telephone calls were used in 40% (27/67) of the interventions mostly as cancer screening invitations and reminders and to arrange screening appointments. Telephone reminders, automated or live, were used alone or with SMS text message reminders, screening invitation letters, and pamphlets. Participants were contacted through telephone to confirm the receipt of a screening invitation letter. Motivational interviews were conducted over the telephone to increase participants' readiness to attend screening [27]. Knowledge about cancer was provided and barriers to screening addressed through telephone counseling [25,33,34].

A few breast cancer mobile apps were specifically designed for interventions. Mammopad, for example, is a decision aid, a tool that helps women to decide to participate in mammogram screening, that ran on the iPad Mini [30]. Another app was designed to assist community health workers (CHWs) in interviewing participants, reporting data, showing a motivational video, and offering a mammogram appointment for women with an abnormal clinical breast examination (CBE). A BSE-facilitating smartphone app included BSE date reminders and a reminder to encourage mother and daughter to practice BSE together [39].

Other mHealth platforms that were less frequently used were emails and social media. Emails were used to deliver screening invitations, reminders, web-based educational flyers, and cancer- and health-related information. Social media platforms such as Facebook and Snapchat were used as intervention modes to provide information about breast and cervical cancers and screening, promote mammogram screening, and schedule breast screening appointments, as well as a platform for discussions about cervical cancer after a lecture [31].

Almost all (11/12, 92%) reviews described mHealth interventions that included 1 or 2 mHealth technologies. There was only 1 intervention that used a combination of >2 types of mHealth technologies: a training in cervical cancer through SMS text message, electronic posters, infographics, podcasts, and video tutorials [40].

Key Stakeholders in mHealth Interventions

Of the 12 reviews, 5 (42%) included telephone call interventions that were delivered by a broad range of personnel [25,27,32-34]. Telephone reminders or telephone calls to make or confirm screening appointments were delivered by bilingual advocates from a community organization, local women recruited from a community charity, research assistants, general practitioner receptionist, volunteers, research nurses, midwives at antenatal health clinics, clinical secretaries, and telemarketers.

Among the important personnel in mHealth interventions were telephone counselors who called the participants to inquire about their screening intention and ascertain whether they had received the invitation letters, provided information about screening, addressed current or potential barriers to screening uptake through motivational interviews and applied a counseling approach to increase motivation for behavior change, or assisted with appointment scheduling. Telephone counseling was delivered by nurses or hospital-based health counselors.

Health navigator services were mentioned in 8% (1/12) of the reviews [30]. Health navigators used mobile apps to facilitate interviews, report data, show motivational videos, and offer screening appointments. Health navigators or CHWs guided participants in navigating cancer screening information, provided transportation and interpretation services, addressed technical problems related to mobile app use, and reminded participants to complete cancer screening.

Cancer Screening Uptake

All (12/12, 100%) reviews included in this review reported mainly improved cancer screening uptake or self-examination practice (for breast or skin cancer; Table 2). The increase in screening between the intervention and control groups (from relevant studies) ranged from 1.2% to 50.9%.

Overall, the reviews concluded that interventions that included >1 communication mode seemed more effective than those that included a single telephone call or SMS text message reminder. A 3-step sequential approach (an automated reminder telephone call and SMS text message, followed by manual telephone calls and face-to-face interviews) conducted at Portuguese primary health care units resulted in 51% of the women in the intervention group attending cervical cancer screening compared with 34% of the women in the control group who received only written invitation letters [25,41]. In another study, women in northern Tanzania who received transportation e-vouchers to cover return transportation to the nearest screening facility as well as a series of 15 behavior change messages delivered through SMS text message were more likely to attend cervical cancer screening (uptake: 18%; OR 4.7, 95% CI 2.9-7.4) compared with those who received only the same SMS text message (uptake: 12.9%; OR 3.0, 95% CI 1.5-6.2) and those who received 3 SMS text messages with the location and hours of the nearest screening clinic (uptake: 4.3%) [27,34,37]. Participants from Iran who received a Health Belief Model–based training in cervical cancer through SMS text messages, electronic posters, infographics, podcasts, and a video tutorial were more likely to complete a Papanicolaou (Pap) test (47.9%) than the participants in the control group (5.8%) [27,40].

A once-a-month SMS text message reminder over 6 months combined with a BSE training through a lecture, video, and demonstration of the technique on a breast model led to a 32% increase in BSE practice [35,36,42]. An Android operating system–based smartphone app that included a BSE date alarm, a reminder to encourage mother and daughter to practice BSE together, a *mother motivation function* that allows the user to call her mother using a notification function to practice BSE together, and educational videos increased the percentage of Korean women practicing BSE from 62.2% to 71.1% [39].

Of the 12 reviews, 1 (8%) included interventions that incorporated navigation to health services [30], which was found to be effective in increasing screening uptake. All (3/3, 100%) of the interventions that included health navigation services were effective in increasing screening uptake. Korean American immigrant women who received a series of 8 to 21 SMS text messages about breast cancer through a specially designed mobile app (mMammogram) and were provided with health navigation services had a significantly higher percentage of completed mammograms after 6 months than women who received printed brochures only (75% vs 30%; $P<.001$) [30,38]. CHWs in Bangladesh who used mobile apps to facilitate CBE, such as showing a motivational video and offering an appointment, detected 3 times more women with abnormal CBEs than CHWs without smartphone support (3.1% without navigation training and 3.2% with navigation training vs 1% without smartphone) [30,43]. CHWs who used mobile apps and were trained in navigation had the highest percentage of participants with an abnormal CBE who attended further clinical assessment compared with those who used mobile apps only or without smartphone support. In a study in the United States, participants who failed to complete a fecal occult blood test were much more likely to complete a second fecal occult blood test than those in usual care if they had been contacted through telephone call by colorectal screening navigators (82.2% vs 37.3% among those who received standard care; $P<.001$) [33,44].

There were a number (46/67, 69%) of studies that used only 1 mode of mHealth communication, and the findings related to screening uptake after the intervention compared with before the intervention were mixed; for example, in an email intervention study, whether an email message was loss-framed (focused on risk), gain-framed (focused on health and well-being improvement), or neutrally framed (provided only facts) had no effect on cervical cancer screening uptake [34,45]. An exception was a study conducted in western Sweden where there was telephone contact through midwives to offer an appointment for a Pap test, which increased the uptake of Pap tests compared with the usual annual invitations without telephone contact (uptake at 3-month follow-up: 13% vs 3.9%; risk ratio 3.37, 95% CI 2.83-4.01) [25,46]. Another exception was the use of Facebook to share breast cancer information and schedule breast screening appointments, which increased breast cancer screening attendance by an average of 12.9% [31,47].

A brief invitation SMS text message was as effective as a detailed informative SMS text message: there was no significant difference in screening uptake between Lebanese women who received an SMS text message mammogram invitation and those who received the same SMS text message and an additional informative SMS text message about the benefits of mammogram screening [28,48].

Screening Awareness, Knowledge, Intention, and Attitude

Of the 12 reviews, 4 (33%) [25,27,30,31] included studies specifically on knowledge, awareness, intention, or attitude in relation to cervical cancer screening (2/4, 50%), breast cancer screening (1/4, 25%), or both (1/4, 25%), and almost all of the individual interventions (7/8, 88%) reported improvements in knowledge, whereas few studies reported an improvement in screening intention (1/4, 25%; Table 2). Interventions that were successful in increasing screening uptake were also successful in increasing knowledge and awareness about screening for both cervical and breast cancer.

The CervixCheck intervention was designed for African American women and consisted of a series of 22 health-specific, spiritually based, cervical cancer–related SMS text messages (eg, on the importance of keeping the body healthy and attending screening) that were sent over 16 days. It resulted in a significant increase in knowledge about cervical cancer and the Pap test (mean difference=0.619; *P*=.001) [25,49]. A 1-week personally tailored SMS text message intervention significantly increased Korean American women's knowledge of cervical cancer screening guidelines (mean difference=0.31-0.71; *P*=.006) [25,50]. Participants who went through the Health Belief Model–based cervical cancer training scored significantly higher in perceived benefits of a Pap test and lower in barriers to obtaining a Pap test, in addition to a higher uptake of Pap tests [27,40]. Female high school students who participated in Facebook or face-to-face discussions for 2 weeks after a 50-minute classroom cervical cancer prevention education lecture that included knowledge about Pap testing increased their knowledge about cervical cancer compared with those in the control group (β=2.942; *P*<.001) [31,51]. Compared with a telephone reminder and invitation intervention, an educational telephone call that provided a brief explanation on cervical cancer, its risks, and colpocytological examination increased knowledge about colpocytological examination but not attitude toward it [27,52].

Korean women who used the mMammogram app and were provided with health navigation services had increased knowledge of breast cancer screening compared with the control group (group difference=mean 16.93, SD 4.77; *P*=.001) [30,38]. Users of Mammopad, a decision aid for mammogram screening, reduced decisional conflict and increased self-efficacy in relation to mammography, although there was no significant change in screening intention [30,53]. Saudi Arabian women who received breast cancer awareness information through Snapchat had better breast cancer awareness and knowledge, including knowledge about breast cancer screening (*P*=.01), than those in the control group who did not receive any awareness information [31,54]. Among women who were surveyed in the tailored SMS text message mammography campaign on Facebook during Breast Cancer Awareness Month, 82% intended to get a mammography in the next year [31,55].

Implementation Outcomes and Measures

Of the 12 reviews, 4 (33%) [25,27,30,31] included studies that examined outcomes related to the implementation of mHealth in cancer screening uptake interventions (Table 2). Of these 4 reviews, 3 (75%) reported a high acceptance of such interventions [25,30,31]. In a 1-week personally tailored SMS text message intervention, 83% of the participants expressed satisfaction with the intervention, and 97% reported that they would recommend the program to their friends [25,50]. In the CervixCheck intervention, 83% of the participants reported being either "satisfied" or "very satisfied," and 85% found the SMS text messages either "useful" or "very useful" [25,49]. The mMammogram intervention participants were satisfied with the intervention (*P*=.003) and agreed that it was effective (*P*<.001) [30,38]. In a tailored SMS text message mammography campaign on Facebook during Breast Cancer Awareness Month, 25% of the women surveyed agreed that they used Facebook to find breast cancer screening information, and 43% agreed with seeing more mammogram information on Facebook [31,55].

Some of the concerns of the participants regarding mHealth interventions included confidentiality of SMS text messages, loss of mobile phones, clarity of the language used, and receiving negative results through SMS text messages. Participants were interested in receiving SMS text message reminders for appointments; however, there was reluctance to receive screening results through SMS text messages in case someone else accessed their mobile phones and saw the results (OR 0.31, 95% CI 0.18-0.51), although they reported no issue with making an appointment.

The barriers to using mHealth in reaching out to people to encourage cancer screening included inconvenience for older participants, lack of texting proficiency, difficulty in texting, and apprehension that SMS text messages might not be clearly understood [56]. Including a reminder and keeping the SMS text messages informative, short, and simple was suggested to increase screening uptake [57].

Discussion

Principal Findings

This scoping review of reviews suggests that mHealth interventions can be effective in increasing cancer screening uptake and practice, as well as improving other screening-related outcomes such as knowledge and awareness about screening. The results are consistent across different types of reviews. The most commonly used mHealth technologies used were SMS text messages and telephone calls. Interventions that included >1 mode of communication, such as telephone calls and SMS text message reminders combined or together with invitation letters, health education, or navigation services, seemed to be more effective than interventions that included only 1 mode of communication. A few (4/12, 33.3%) of the reviews reported implementation measures, and 75% (3/4) suggested that mHealth interventions were well accepted by participants.

The effectiveness of interventions that used >1 mode of communication has been demonstrated in cancer screening uptake in LMICs [58]; for example, in Malaysia, mass media campaigns that used different channels of health promotion successfully increased symptom awareness of breast cancer [59] and colorectal cancer [60].

A very effective intervention was a combination of educational SMS text messages and e-vouchers to subsidize the transportation to attend screening [37], which is especially relevant in rural areas in LMICs. In many LMICs, public transport and e-hailing services are mainly available in cities, and the majority of health care facilities that offer cancer screening are located in town areas; for example, in Malaysia, travel distance to the nearest mammogram screening facility ranged between 2 km and 340 km with a median of 22 (IQR 12-42) km [61]. Longer travel distance to cancer services is associated with lower likelihood of cancer screening uptake [62] and presentation of more advanced stages of breast cancer [63] and colorectal cancer [64]. Interventions that increase knowledge might not translate into higher screening uptake if underlying structural barriers to screening, such as lack of transportation, are not addressed [25,27]. The use of e-vouchers has been described as a form of an "enablement" intervention that reduces “barriers to increase capability or opportunity” [65].

Approximately half (31/67, 46%) of the interventions included in the reviews included SMS text messaging, which uses a cellular network and is preinstalled on every mobile phone, unlike internet-based instant messaging apps. Almost 100% of SMS text messages are read, and 90% of them are read within 30 minutes of receipt compared with emails (approximately 18% are read) [66], which might explain the ineffectiveness of emails in improving cancer screening uptake and related outcomes. Worldwide, IP-based chat apps are gaining popularity: WhatsApp, Facebook Messenger, and WeChat have 2 billion users, 1.3 billion users, and 1.2 billion users, respectively [67]. Chat apps, especially those with high open rates, such as SMS text messaging [68], enable more efficient communication by allowing users to send longer messages; share pictures, videos, or audio messages; and chat in real time. However, because SMS text messaging is operator-based, it is more useful in rural areas where there is poor mobile internet coverage. In addition, SMS text messaging is simple to use and does not require additional apps, which might be more user friendly for those who are less tech savvy; for example, older adults.

In addition to SMS text messaging and chat apps, social networking sites, with their large numbers of users, hold great potential in mHealth interventions. As of July 2021, popular social networking sites such as Facebook and Instagram had 2.85 billion users and 1.39 billion users, respectively, and the numbers are increasing rapidly [67]. However, in the only review that examined social media solely [31], the studies included were mostly about low-level engagement (number of impressions, reach, likes, comments, and sharing of tweets and posts), and the review highlighted the lack of studies (1/4, 25%) that examined high-level engagement with social media interventions, such as uptake of screening [31]. This is likely because of the difficulty in linking screening uptake and social media data because social media posts are not designed for such analysis. The fast pace of social media means that social media contents could be outdated quickly or get inundated by other information, which reduces their reach to the target population and long-term sustainability. Running multiple campaigns on multiple social media platforms also means that it is difficult to pinpoint which campaign or platform has the greatest impact on behavior change. In addition, there are age differences in social media use; many individuals in the targeted age groups for cancer screening might not be reached through social media. In a survey of American adults, >80% of those aged 18 to 49 years and 73% of those aged 50 to 64 years used social media sites, whereas only 49% of those aged ≥65 years reported so [69].

mHealth interventions will only work if there is access to mobile phones and mobile internet. Globally, although the penetration of mobile phones and mobile internet is high, there is an unequal access to mobile technology and internet between urban and rural areas and between sexes. All urban areas are covered by a mobile broadband network; however, in some LMICs, 19% of the rural population are covered by only a second generation network, and 17% of the rural population have no mobile coverage at all [70]. The rural-urban gap is especially prominent in LMICs, where urban access to a mobile broadband network is 2.3 times as high as rural access [70]. In LMICs, women’s mobile phone ownership and internet use is significantly lower than that of men’s, and the gap ranges from 50% in South Asia and 20% in sub-Saharan Africa to 12% in the Middle East and North Africa [71].

The gap in mobile phone ownership and internet use has important public health implications. mHealth interventions to increase cancer screening uptake might be less effective in rural areas, where screening uptake is already low [72]. Many (57/67, 85%) of the mHealth interventions targeted cervical and breast cancers, the 2 most common cancers among women. Thus, the rural-urban gap in mobile access means that women from rural areas are at a greater risk of inequitable access to information and interventions on cancer screening.

Given the rapid development of mHealth technologies, there is a need for researchers to incorporate them effectively into interventions. However, the speed of research does not advance at the speed of mobile technology, and researchers have little control over app development [30]. Most smartphone apps address tertiary cancer prevention [30], such as support for patients with cancer in health information management [73], medication adherence [74], weight management [75], and mental health improvement among cancer survivors [76], and there is a lack of smartphone apps for secondary cancer prevention. Many of the apps developed for research are not available for download and have not been widely adopted after the studies were concluded.

mHealth holds great potential to reach out to many people in low-cost settings, and it is also safe in times of the COVID-19 pandemic where social contact has to be minimized. However, it might not be acceptable to pass on personal information through certain mHealth technology; for example, there were participants who mentioned that although it was acceptable to receive SMS text message reminders about their screening appointment, they would not want to be informed about their screening results through SMS text messages. The gap could be filled by CHWs, who could act as the link between mHealth technologies and participants by informing the latter personally through telephone calls of their screening results. A recent review found that CHWs play a critical role, particularly during pandemics, in community engagement [77]. CHWs are usually members from the same communities as the intervention participants and are knowledgeable about the resources available within the communities. They may be able to reach out to vulnerable populations and encourage uptake of cancer screening [78] and mobile technologies [79]. However, despite proven effectiveness of CHWs in cancer screening and early diagnosis interventions [80], there were very few (1/12, 8%) reviews that included interventions that combined mHealth and CHWs.

Limitations

Given the heterogeneity of reporting and differences in the details reported in each review, it was challenging to summarize the evidence from the reviews concisely. In addition, some reviews did not exclusively examine mHealth and cancer screening; they included other types of interventions and preventive measures. Furthermore, unlike in a systematic review, the quality of the selected articles was not assessed.

Future Research and Recommendations

Future interventions should consider combining at least two modes of mHealth communication, for example, SMS text messages and telephone calls, and screening interventions are likely to achieve better attendance when participants receive at least one reminder. In addition, future interventions should consider incorporating instant messaging apps such as WhatsApp, Facebook Messenger, and WeChat, in addition to SMS text messaging, because the number of users is increasing exponentially, and more educational information using videos, audio messages, or graphics could be shared. Social media platforms, especially Facebook, should be incorporated for health promotion, sharing of educational information, and appointment making. When social media platforms are used, there is a need to take into account their popularity and acceptability within the country where the interventions are conducted. In addition, engagement with different social media apps varies among age groups. Other incentives such as transport vouchers may be included when interventions are conducted among those with poor access to screening facilities. Facilitators to improving access to, and engagement with, mHealth among older adults have been described, including support from the government and family, addressing digital problems in deprived areas, and increasing accessibility to mobile phones or tablet computers [81]. CHWs and navigation services may be provided along with mobile technologies to support participants' needs, promote and facilitate the use of mHealth, and pass on information such as screening results.

Conclusions

mHealth interventions have the potential to increase cancer screening uptake and other cancer screening–related outcomes such as knowledge about screening and intention to screen. Combining >1 mode of communication may have a better impact on cancer screening uptake.

Acknowledgments

This work was supported by the UK Research and Innovation Medical Research Council (537084059) and the Newton Fund Impact Scheme 2020-2021 (MIGHT/CEO/NUOF/4-2020 (06) for the study "Improving early detection and diagnosis of breast cancer among multi-ethnic rural communities in Malaysia – the implementation of the community education and navigation programme (CENP)." MD and DS were also supported by the UK Research and Innovation Medical Research Council and Global Challenges Research Fund (MR/S014349/1) during the conduct and write-up of this study.

Authors' Contributions

DS and MMT developed the review protocol, MMT and WMKH conducted the screening independently, DS and MMT wrote the manuscript, and TTS and MD planned the study and led in revising the manuscript. DS, WMKH, DM, and NAT revised the manuscript. All authors provided feedback and helped shape the research, analysis, and manuscript. DS and MMT contributed to the study equally and are joint first authors of the manuscript. TTS and MD contributed equally to the study and are joint last authors of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 19 KB - jmir_v24i8e36316_app1.docx]

References

1. Cancer. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/cancer [accessed 2022-02-16]
2. 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]
3. Clegg-Lamptey JA, Vanderpuye V, Dedey F. Late presentation of breast cancer in lower- and middle-income countries. Curr Breast Cancer Rep 2019 Jun 20;11(3):143-151. [doi: 10.1007/s12609-019-00312-8]
4. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2013 Jun 04(6):CD001877 [FREE Full text] [doi: 10.1002/14651858.CD001877.pub5] [Medline: 23737396]
5. Jansen EE, Zielonke N, Gini A, Anttila A, Segnan N, Vokó Z, et al. Response to the letter commenting on 'Effect of organised cervical cancer screening on cervical cancer mortality in Europe: a systematic review'. Eur J Cancer 2020 Oct;138:232-233. [doi: 10.1016/j.ejca.2020.07.012] [Medline: 32811688]
6. Hugosson J, Carlsson S, Aus G, Bergdahl S, Khatami A, Lodding P, et al. Mortality results from the Göteborg randomised population-based prostate-cancer screening trial. Lancet Oncol 2010 Aug;11(8):725-732. [doi: 10.1016/s1470-2045(10)70146-7]

7. Navarro M, Nicolas A, Ferrandez A, Lanas A. Colorectal cancer population screening programs worldwide in 2016: an update. World J Gastroenterol 2017 May 28;23(20):3632-3642 [FREE Full text] [doi: 10.3748/wjg.v23.i20.3632] [Medline: 28611516]
8. Sankaranarayanan R, Swaminathan R, Brenner H, Chen K, Chia KS, Chen JG, et al. Cancer survival in Africa, Asia, and Central America: a population-based study. Lancet Oncol 2010 Feb;11(2):165-173. [doi: 10.1016/s1470-2045(09)70335-3]
9. Freitas-Junior R, Rodrigues DC, Corrêa RD, Peixoto JE, de Oliveira HV, Rahal RM. Contribution of the Unified Health Care System to mammography screening in Brazil, 2013. Radiol Bras 2016 Oct;49(5):305-310 [FREE Full text] [doi: 10.1590/0100-3984.2014.0129] [Medline: 27818544]
10. Aidalina M, Syed Mohamed AS. The uptake of Mammogram screening in Malaysia and its associated factors: a systematic review. Med J Malaysia 2018 Aug;73(4):202-211 [FREE Full text] [Medline: 30121682]
11. Albert US, Kalder M, Schulte H, Klusendick M, Diener J, Schulz-Zehden B, et al. [The population-based mammography screening programme in Germany: uptake and first experiences of women in 10 federal states]. Gesundheitswesen 2012 Feb;74(2):61-70. [doi: 10.1055/s-0030-1268441] [Medline: 21229475]
12. Carrasco-Garrido P, Hernandez-Barrera V, Lopez de Andres A, Jimenez-Trujillo I, Gallardo Pino C, Jimenez-Garcıa R. Awareness and uptake of colorectal, breast, cervical and prostate cancer screening tests in Spain. Eur J Public Health 2014 Apr 27;24(2):264-270. [doi: 10.1093/eurpub/ckt089] [Medline: 23813710]
13. Koo JH, Leong RW, Ching J, Yeoh K, Wu D, Murdani A, Asia Pacific Working Group in Colorectal Cancer. Knowledge of, attitudes toward, and barriers to participation of colorectal cancer screening tests in the Asia-Pacific region: a multicenter study. Gastrointest Endosc 2012 Jul;76(1):126-135. [doi: 10.1016/j.gie.2012.03.168] [Medline: 22726471]
14. WHO Guideline Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019.
15. MHealth New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
16. Digital around the world. DataReportal. URL: https://datareportal.com/global-digital-overview [accessed 2022-05-05]
17. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug 06;1(2):160-171. [doi: 10.9745/ghsp-d-13-00031]
18. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
19. Buneviciene I, Mekary RA, Smith TR, Onnela J, Bunevicius A. Can mHealth interventions improve quality of life of cancer patients? A systematic review and meta-analysis. Crit Rev Oncol Hematol 2021 Jan;157:103123 [FREE Full text] [doi: 10.1016/j.critrevonc.2020.103123] [Medline: 33190065]
20. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-tested mobile apps for breast cancer care: systematic review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
21. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep 10;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
22. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb 11;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
23. 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]
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 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
25. Bhochhibhoya S, Dobbs PD, Maness SB. Interventions using mHealth strategies to improve screening rates of cervical cancer: a scoping review. Prev Med 2021 Feb;143:106387. [doi: 10.1016/j.ypmed.2020.106387] [Medline: 33383069]
26. Uy C, Lopez J, Trinh-Shevrin C, Kwon SC, Sherman SE, Liang PS. Text messaging interventions on cancer screening rates: a systematic review. J Med Internet Res 2017 Aug 24;19(8):e296 [FREE Full text] [doi: 10.2196/jmir.7893] [Medline: 28838885]
27. Zhang D, Advani S, Waller J, Cupertino A, Hurtado-de-Mendoza A, Chicaiza A, et al. Mobile technologies and cervical cancer screening in low- and middle-income countries: a systematic review. JCO Glob Oncol 2020 Apr;6:617-627 [FREE Full text] [doi: 10.1200/JGO.19.00201] [Medline: 32302236]
28. Halake DG, Ogoncho I. The role of mobile health in supporting cancer prevention, detection, treatment and palliative care in low and middle income countries: a scoping review. Public Health Res 2017;7(6):123-135. [doi: 10.5923/j.phr.20170706.01]
29. Choi J, Cho Y, Woo H. mHealth approaches in managing skin cancer: systematic review of evidence-based research using integrative mapping. JMIR Mhealth Uhealth 2018 Aug 02;6(8):e164 [FREE Full text] [doi: 10.2196/mhealth.8554] [Medline: 30072362]

30. Houghton LC, Howland RE, McDonald JA. Mobilizing breast cancer prevention research through smartphone apps: a systematic review of the literature. Front Public Health 2019 Nov 6;7:298 [FREE Full text] [doi: 10.3389/fpubh.2019.00298] [Medline: 31781525]
31. Plackett R, Kaushal A, Kassianos AP, Cross A, Lewins D, Sheringham J, et al. Use of social media to promote cancer screening and early diagnosis: scoping review. J Med Internet Res 2020 Nov 09;22(11):e21582 [FREE Full text] [doi: 10.2196/21582] [Medline: 33164907]
32. Musa J, Achenbach CJ, O'Dwyer LC, Evans CT, McHugh M, Hou L, et al. Effect of cervical cancer education and provider recommendation for screening on screening rates: a systematic review and meta-analysis. PLoS ONE 2017 Sep 5;12(9):e0183924. [doi: 10.1371/journal.pone.0183924]
33. Duffy SW, Myles JP, Maroni R, Mohammad A. Rapid review of evaluation of interventions to improve participation in cancer screening services. J Med Screen 2017 Sep 17;24(3):127-145 [FREE Full text] [doi: 10.1177/0969141316664757] [Medline: 27754937]
34. Lott BE, Trejo MJ, Baum C, McClelland DJ, Adsul P, Madhivanan P, et al. Interventions to increase uptake of cervical screening in sub-Saharan Africa: a scoping review using the integrated behavioral model. BMC Public Health 2020 May 11;20(1):654 [FREE Full text] [doi: 10.1186/s12889-020-08777-4] [Medline: 32393218]
35. Déglise C, Suggs LS, Odermatt P. Short message service (SMS) applications for disease prevention in developing countries. J Med Internet Res 2012 Jan 12;14(1):e3 [FREE Full text] [doi: 10.2196/jmir.1823] [Medline: 22262730]
36. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
37. Erwin E, Aronson KJ, Day A, Ginsburg O, Macheku G, Feksi A, et al. SMS behaviour change communication and eVoucher interventions to increase uptake of cervical cancer screening in the Kilimanjaro and Arusha regions of Tanzania: a randomised, double-blind, controlled trial of effectiveness. BMJ Innov 2019 Jan 22;5(1):28-34 [FREE Full text] [doi: 10.1136/bmjinnov-2018-000276] [Medline: 31645991]
38. Lee H, Ghebre R, Le C, Jang YJ, Sharratt M, Yee D. Mobile phone multilevel and multimedia messaging intervention for breast cancer screening: pilot randomized controlled trial. JMIR Mhealth Uhealth 2017 Nov 07;5(11):e154 [FREE Full text] [doi: 10.2196/mhealth.7091] [Medline: 29113961]
39. Heo J, Chun M, Lee KY, Oh Y, Noh OK, Park RW. Effects of a smartphone application on breast self-examination: a feasibility study. Healthc Inform Res 2013 Dec;19(4):250-260 [FREE Full text] [doi: 10.4258/hir.2013.19.4.250] [Medline: 24523989]
40. Khademolhosseini F, Noroozi A, Tahmasebi R. The effect of health belief model-based education through telegram instant messaging services on pap smear performance. Asian Pac J Cancer Prev 2017 Aug 27;18(8):2221-2226 [FREE Full text] [doi: 10.22034/APJCP.2017.18.8.2221] [Medline: 28843259]
41. Firmino-Machado J, Varela S, Mendes R, Moreira A, Lunet N, SCAN-Cervical Cancer collaborators. A 3-step intervention to improve adherence to cervical cancer screening: the SCAN randomized controlled trial. Prev Med 2019 Jun;123:250-261. [doi: 10.1016/j.ypmed.2019.03.025] [Medline: 30936001]
42. Khokhar A. Short text messages (SMS) as a reminder system for making working women from Delhi Breast Aware. Asian Pac J Cancer Prev 2009;10(2):319-322 [FREE Full text] [Medline: 19537904]
43. Ginsburg OM, Chowdhury M, Wu W, Chowdhury MT, Pal BC, Hasan R, et al. An mHealth model to increase clinic attendance for breast symptoms in rural Bangladesh: can bridging the digital divide help close the cancer divide? Oncologist 2014 Feb;19(2):177-185 [FREE Full text] [doi: 10.1634/theoncologist.2013-0314] [Medline: 24396050]
44. Baker DW, Brown T, Buchanan DR, Weil J, Balsley K, Ranalli L, et al. Comparative effectiveness of a multifaceted intervention to improve adherence to annual colorectal cancer screening in community health centers: a randomized clinical trial. JAMA Intern Med 2014 Aug 01;174(8):1235-1241. [doi: 10.1001/jamainternmed.2014.2352] [Medline: 24934845]
45. Adonis L, Paramanund J, Basu D, Luiz J. Framing preventive care messaging and cervical cancer screening in a health-insured population in South Africa: implications for population-based communication? J Health Psychol 2017 Sep 17;22(11):1365-1375. [doi: 10.1177/1359105316628735] [Medline: 26888327]
46. Broberg G, Gyrd-Hansen D, Miao Jonasson J, Ryd M, Holtenman M, Milsom I, et al. Increasing participation in cervical cancer screening: offering a HPV self-test to long-term non-attendees as part of RACOMIP, a Swedish randomized controlled trial. Int J Cancer 2014 May 01;134(9):2223-2230 [FREE Full text] [doi: 10.1002/ijc.28545] [Medline: 24127304]
47. Using Facebook to improve attendance at screening appointments. Good Things Foundation. URL: https://static1.squarespace.com/static/5a744e79bff20048181ee226/t/5f7ec8861c77e72d9096b3c3/1602144395078/Evaluation_+Stoke+Screening+v4.pdf [accessed 2022-05-05]
48. Lakkis NA, Atfeh AM, El-Zein YR, Mahmassani DM, Hamadeh GN. The effect of two types of sms-texts on the uptake of screening mammogram: a randomized controlled trial. Prev Med 2011 Oct;53(4-5):325-327. [doi: 10.1016/j.ypmed.2011.08.013] [Medline: 21871480]
49. Le D, Holt CL. CervixCheck: a spiritually-based text messaging intervention to promote cervical cancer awareness and pap test screening intention among African-American women. J Health Commun 2018 Oct 09;23(9):842-853 [FREE Full text] [doi: 10.1080/10810730.2018.1528317] [Medline: 30300091]

50. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. J Med Internet Res 2014 Aug 27;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
51. Lai C, Wu W, Tsai S, Cheng S, Lin K, Liang S. The effectiveness of a Facebook-assisted teaching method on knowledge and attitudes about cervical cancer prevention and HPV vaccination intention among female adolescent students in Taiwan. Health Educ Behav 2015 Jun 15;42(3):352-360. [doi: 10.1177/1090198114558591] [Medline: 25512072]
52. Lima TM, Nicolau AI, Carvalho FH, Vasconcelos CT, Aquino PD, Pinheiro AK. Telephone interventions for adherence to colpocytological examination. Rev Lat Am Enfermagem 2017 Feb 06;25:e2844 [FREE Full text] [doi: 10.1590/1518-8345.1683.2844] [Medline: 28177055]
53. Eden KB, Scariati P, Klein K, Watson L, Remiker M, Hribar M, et al. Mammography decision aid reduces decisional conflict for women in their forties considering screening. J Womens Health (Larchmt) 2015 Dec;24(12):1013-1020 [FREE Full text] [doi: 10.1089/jwh.2015.5256] [Medline: 26360918]
54. Alanzi TM, Alobrah A, Alhumaidi R, Aloraifi S. Evaluation of the SnapChat mobile social networking application for breast cancer awareness among Saudi students in the Dammam Region of the Kingdom of Saudi Arabia. Breast Cancer Target Ther 2018 Jul;Volume 10:113-119. [doi: 10.2147/bctt.s166135]
55. Klippert H, Schaper A. Using Facebook to communicate mammography messages to rural audiences. Public Health Nurs 2019 Mar 29;36(2):164-171. [doi: 10.1111/phn.12556] [Medline: 30370687]
56. Romocki LS, Des Marais A, Cofie L, Anderson C, Curington T, Smith JS. Use of short messaging service to improve follow-up for abnormal pap test results in minority and medically underserved women in north Carolina: questionnaire on attitudes and acceptability. JMIR Form Res 2019 Aug 06;3(3):e12675 [FREE Full text] [doi: 10.2196/12675] [Medline: 31389334]
57. Soto M, Martinez-Gutierrez J, Momany M, Capurro D, Ciampi Spode F, Cea E, et al. Preferences of underserved Chilean women on a mobile technology intervention for cervical cancer screening: qualitative study. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e196 [FREE Full text] [doi: 10.2196/mhealth.9494] [Medline: 30459141]
58. Schliemann D, Ramanathan K, Matovu N, O'Neill C, Kee F, Su TT, et al. The implementation of colorectal cancer screening interventions in low-and middle-income countries: a scoping review. BMC Cancer 2021 Oct 19;21(1):1125 [FREE Full text] [doi: 10.1186/s12885-021-08809-1] [Medline: 34666704]
59. Schliemann D, Htay MN, Dahlui M, Paramasivam D, Cardwell CR, Ibrahim Tamin NS, et al. Impact of a mass media campaign on breast cancer symptoms awareness and screening uptake in Malaysia: findings from a quasi-experimental study. BMJ Open 2020 Aug 20;10(8):e036503 [FREE Full text] [doi: 10.1136/bmjopen-2019-036503] [Medline: 32819988]
60. Schliemann D, Paramasivam D, Dahlui M, Cardwell CR, Somasundaram S, Ibrahim Tamin NS, et al. Change in public awareness of colorectal cancer symptoms following the Be Cancer Alert Campaign in the multi-ethnic population of Malaysia. BMC Cancer 2020 Mar 25;20(1):252 [FREE Full text] [doi: 10.1186/s12885-020-06742-3] [Medline: 32213173]
61. Mahmud A, Aljunid SM. Availability and accessibility of subsidized mammogram screening program in peninsular Malaysia: a preliminary study using travel impedance approach. PLoS One 2018;13(2):e0191764 [FREE Full text] [doi: 10.1371/journal.pone.0191764] [Medline: 29389972]
62. Brustrom JE, Hunter D. Going the distance: how far will women travel to undergo free mammography? Mil Med 2001 Apr;166(4):347-349. [Medline: 11315478]
63. Huang B, Dignan M, Han D, Johnson O. Does distance matter? Distance to mammography facilities and stage at diagnosis of breast cancer in Kentucky. J Rural Health 2009;25(4):366-371. [doi: 10.1111/j.1748-0361.2009.00245.x] [Medline: 19780916]
64. Massarweh NN, Chiang Y, Xing Y, Chang GJ, Haynes AB, You YN, et al. Association between travel distance and metastatic disease at diagnosis among patients with colon cancer. J Clin Oncol 2014 Mar 20;32(9):942-948. [doi: 10.1200/jco.2013.52.3845]
65. 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]
66. 2019 Mobile usage report: how consumers are really texting. EZTexting. URL: https://cdn.brandfolder.io/YAYWC2JN/at/5j668ngq5jk54hg425jm769f/2019_Mobile_Usage_Report.pdf [accessed 2022-06-05]
67. Most popular messaging apps. Statista. URL: https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/ [accessed 2022-04-04]
68. Church K, Oliveira RD. What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS. In: Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services. 2013 Presented at: MobileHCI '13: 15th International Conference on Human-Computer Interaction with Mobile Devices and Services; Aug 27 - 30, 2013; Munich Germany. [doi: 10.1145/2493190.2493225]
69. Social media use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2022-04-15]
70. Measuring digital development: facts and figures 2020. ITU. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx [accessed 2022-04-15]

71. Mariscal J, Mayne G, Aneja U. Bridging the gender digital gap. Economics 2019;13. [doi: 10.5018/economics-ejournal.ja.2019-9]
72. Walji LT, Murchie P, Lip G, Speirs V, Iversen L. Exploring the influence of rural residence on uptake of organized cancer screening - a systematic review of international literature. Cancer Epidemiol 2021 Oct;74:101995. [doi: 10.1016/j.canep.2021.101995] [Medline: 34416545]
73. Klasnja P, Hartzler A, Powell C, Pratt W. Supporting cancer patients' unanchored health information management with mobile technology. AMIA Annu Symp Proc 2011;2011:732-741 [FREE Full text] [Medline: 22195130]
74. Chalela P, Munoz E, Inupakutika D, Kaghyan S, Akopian D, Kaklamani V, et al. Corrigendum to "Improving adherence to endocrine hormonal therapy among breast cancer patients: study protocol for a randomized controlled trial" [Contemp. Clin. Trials Commun. 12 (2018) 109-115]. Contemp Clin Trials Commun 2019 Jun;14:100349 [FREE Full text] [doi: 10.1016/j.conctc.2019.100349] [Medline: 31049460]
75. Stubbins R, He T, Yu X, Puppala M, Ezeana CF, Chen S, et al. A behavior-modification, clinical-grade mobile application to improve breast cancer survivors' accountability and health outcomes. JCO Clin Cancer Inform 2018 Dec(2):1-11. [doi: 10.1200/cci.18.00054]
76. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]
77. Bhaumik S, Moola S, Tyagi J, Nambiar D, Kakoti M. Community health workers for pandemic response: a rapid evidence synthesis. BMJ Glob Health 2020 Jun 10;5(6):e002769 [FREE Full text] [doi: 10.1136/bmjgh-2020-002769] [Medline: 32522738]
78. Morère JF, Eisinger F, Touboul C, Lhomel C, Couraud S, Viguier J. Decline in cancer screening in vulnerable populations? Results of the EDIFICE surveys. Curr Oncol Rep 2018 Mar 05;20(Suppl 1):17. [doi: 10.1007/s11912-017-0649-7] [Medline: 29508084]
79. Liu P, Astudillo K, Velez D, Kelley L, Cobbs-Lomax D, Spatz ES. Use of mobile health applications in low-income populations. Circ Cardiovasc Qual Outcomes 2020 Sep;13(9):e007031. [doi: 10.1161/circoutcomes.120.007031]
80. O'Donovan J, Newcomb A, MacRae MC, Vieira D, Onyilofor C, Ginsburg O. Community health workers and early detection of breast cancer in low-income and middle-income countries: a systematic scoping review of the literature. BMJ Glob Health 2020 May 13;5(5):e002466 [FREE Full text] [doi: 10.1136/bmjgh-2020-002466] [Medline: 32409331]
81. Abbaspur-Behbahani S, Monaghesh E, Hajizadeh A, Fehresti S. Application of mobile health to support the elderly during the COVID-19 outbreak: a systematic review. Health Policy Technol 2022 Mar;11(1):100595 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100595] [Medline: 35018280]

Abbreviations

BSE: breast self-examination
CBE: clinical breast examination
CHW: community health worker
LMICs: low- and middle-income countries
mHealth: mobile health
Pap: Papanicolaou
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by A Mavragani; submitted 13.01.22; peer-reviewed by A Spini, HL Tam; comments to author 07.03.22; revised version received 14.06.22; accepted 16.06.22; published 15.08.22.

Please cite as:
Schliemann D, Tan MM, Hoe WMK, Mohan D, Taib NA, Donnelly M, Su TT
mHealth Interventions to Improve Cancer Screening and Early Detection: Scoping Review of Reviews
J Med Internet Res 2022;24(8):e36316
URL: https://www.jmir.org/2022/8/e36316
doi:10.2196/36316
PMID:35969450

Review

Digital Twins for Managing Health Care Systems: Rapid Literature Review

Safa Elkefi[1], MSc; Onur Asan[1], PhD

[1] School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

Corresponding Author:
Onur Asan, PhD
School of Systems and Enterprises
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ, 07030
United States
Phone: 1 4145264330
Email: oasan@stevens.edu

Abstract

Background: Although most digital twin (DT) applications for health care have emerged in precision medicine, DTs can potentially support the overall health care process. DTs (twinned systems, processes, and products) can be used to optimize flows, improve performance, improve health outcomes, and improve the experiences of patients, doctors, and other stakeholders with minimal risk.

Objective: This paper aims to review applications of DT systems, products, and processes as well as analyze the potential of these applications for improving health care management and the challenges associated with this emerging technology.

Methods: We performed a rapid review of the literature and reported available studies on DTs and their applications in health care management. We searched 5 databases for studies published between January 2002 and January 2022 and included peer-reviewed studies written in English. We excluded studies reporting DT usage to support health care practice (organ transplant, precision medicine, etc). Studies were analyzed based on their contribution toward DT technology to improve user experience in health care from human factors and systems engineering perspectives, accounting for the type of impact (product, process, or performance/system level). Challenges related to the adoption of DTs were also summarized.

Results: The DT-related studies aimed at managing health care systems have been growing over time from 0 studies in 2002 to 17 in 2022, with 7 published in 2021 (N=17 studies). The findings reported on applications categorized by DT type (system: n=8; process: n=5; product: n=4) and their contributions or functions. We identified 4 main functions of DTs in health care management including safety management (n=3), information management (n=2), health management and well-being promotion (n=3), and operational control (n=9). DTs used in health care systems management have the potential to avoid unintended or unexpected harm to people during the provision of health care processes. They also can help identify crisis-related threats to a system and control the impacts. In addition, DTs ensure privacy, security, and real-time information access to all stakeholders. Furthermore, they are beneficial in empowering self-care abilities by enabling health management practices and providing high system efficiency levels by ensuring that health care facilities run smoothly and offer high-quality care to every patient.

Conclusions: The use of DTs for health care systems management is an emerging topic. This can be seen in the limited literature supporting this technology. However, DTs are increasingly being used to ensure patient safety and well-being in an organized system. Thus, further studies aiming to address the challenges of health care systems challenges and improve their performance should investigate the potential of DT technology. In addition, such technologies should embed human factors and ergonomics principles to ensure better design and more successful impact on patient and doctor experiences.

(J Med Internet Res 2022;24(8):e37641) doi:10.2196/37641

KEYWORDS

health care; digital twins; safety; information management; supply chain management; operational control; well-being promotion; human factors; technology; health informatics; literature synthesis; scheduling and optimization; digital health

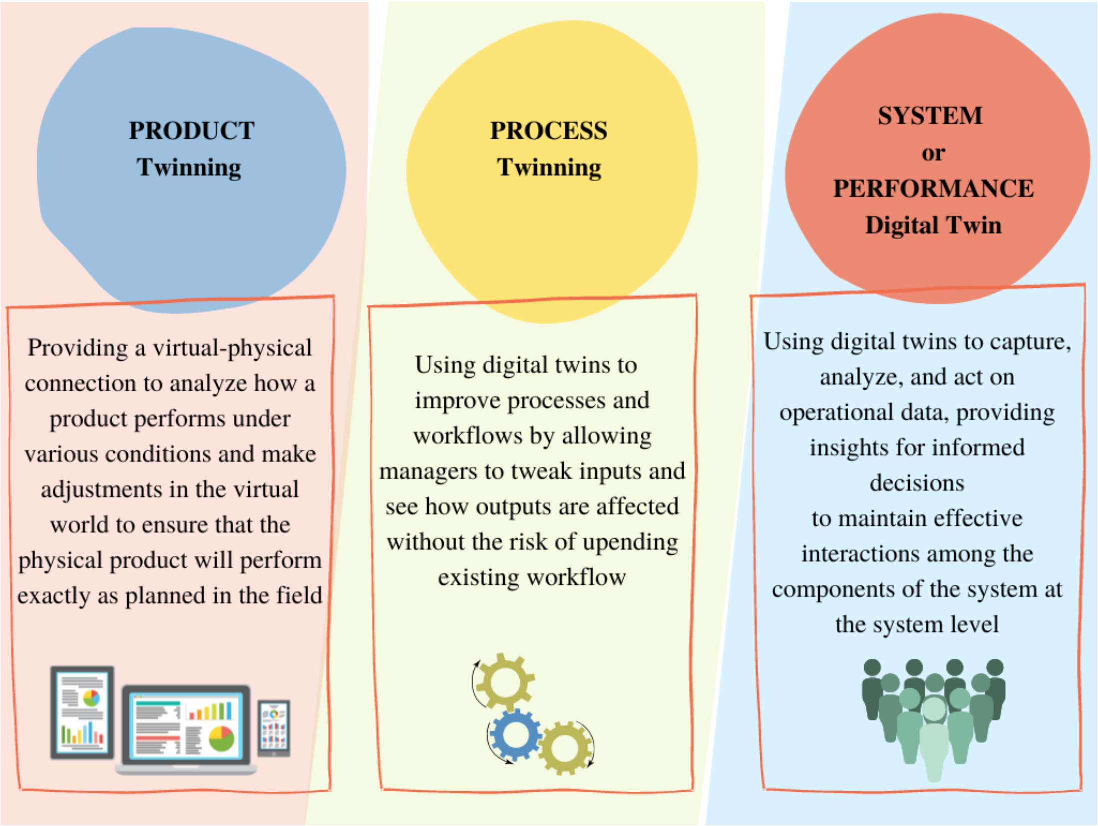
Introduction

One of the fastest growing sectors of the global economy is the health care industry [1,2]. For a complex system like a hospital, many problems and obstacles arise owing to the variability resulting from the incongruity between demand, and capacity and resource availability. In addition to the operational management of resources, having an almost instantaneous and reliable vision of the available resources would permit a more adaptive management of the resources as the demand varies. As the demand changes, managing staff schedules, patient flow, bed sizes, and room usage would be a challenge [1]. Technology-based strategies can be promising contributors to improving the efficiency of health care delivery. The increasing adoption of various health information technologies has created new channels for management [3] and communication [4] that revolutionize health care systems.

Meanwhile, a revolution toward an intelligent industry or "Industry 4.0" combining advanced technologies emerged in 2011 [5]. This revolution affected all sectors, including health care. One of the supporting concepts in implementing Industry 4.0 is the digital twin (DT) [6]. A DT is a virtual representation of a physical asset replicated virtually through data connection [7,8], making it possible to link the system with its virtual copies in a bidirectional way [9]. Digital technologies provide many opportunities to revolutionize health care. For example, real-time data can be provided by Internet of Things solutions, and large data flows are managed and secured by robust digital infrastructures. In addition, flows and decision-making support are improved by command centers, artificial intelligence, and machine learning [10-12]. However, it is through the creation of DTs that much of this can be brought together [8]. The medical DT concept is considered more beneficial for personalized medicine to help health care professionals realize more effective interventions by digitally replicating the human body, allowing prevention, early detection, and targeted treatments of many diseases [13,14]. This paradigm is not limited to medical practice improvement; it also offers a solution to the issues related to health care systems and supports their strategic management. A DT can help design, optimize, and test products; design and operate production systems; manage and control supply chains; diagnose problems; and provide after-market services [15]. There are 3 types of DTs, as illustrated in Figure 1.

In this context, this rapid review aims to highlight what DTs have accomplished in correlation with health care management support. We intend to cover the interventions that used DTs (products, processes, systems) to improve the management of medical services. We report the DT type and its role in the system (function). This classification of DTs is adopted from Siemens, classifying DTs into 3 types, with 1 related to processes (eg, production), 1 related to product design, and 1 related to system performance (eg, performance) [16]. The combination and integration of the 3 DTs as they evolve together is known as the digital thread [16].

Figure 1. Types of digital twins.



Methods

Study Design

We performed a rapid review of studies involving DT technology to improve health care services management. Rapid reviews are a form of evidence synthesis that can provide timelier decision-making information than standard systematic reviews. They are suitable for emerging research topics where systematic reviews are unpractical [17]. Rapid reviews typically do not include an exhaustive set of studies, do not involve formal analyses of the study quality, and report findings from prior studies via narrative synthesis by simplifying the evidence synthesis process [18].

Our protocol was registered on January 28, 2022, with the Open Science Framework [19]. The primary goal of this review was to identify the opportunities that DTs have offered to support the improvement of the health care system. We summarized the literature on existing applications of DTs and the challenges associated with their design, use, and implementation. Publications spanning the last 20 years were considered, from January 1, 2002, to January 25, 2022. We started in January 2002 because the concept of DTs was publicly first introduced in 2002 by Grieves [20]. Grieves proposed the DT as the conceptual model underlying product lifecycle management [20].

Search Strategy

We searched PubMed, Web of Science, IEEE Xplore, Scopus, and ScienceDirect using “digital twin” and “health” as the keywords. The studies included journal and conference articles that covered only the health care applications of DTs (no industrial, manufacturing, or energy-related initiatives). We excluded the following types of studies: studies published in a language other than English; reviews, short communications, and briefs not reporting the impact of DTs through empirical studies and approaches; papers that are not peer-reviewed; studies that present an initiative to support a medical practice (precision medicine, organ transplant, etc).

Studies are discussed based on the contribution of DTs to improving user experience in health care from human factors and systems engineering perspectives, accounting for the type of impact (product, process, or performance/system level).

Results

Of all the sources found, 72 papers were screened comprehensively, and 17 papers were included in this review. The screening process is summarized in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram shown in Figure 2.

Even though very few articles matched our scope, the research trend is evolving. The selected papers covered different application areas, including process development (n=5), a system improvement initiative (n=8), and developing/designing and testing a product (n=4). We adhered to conventions for narrative reviews by combining our results with interpretations and discussion in the Results and Discussion sections.

We identified 4 main functions that DTs perform in managing health care systems. We summarize the functions and their adopted definitions in Table 1.

Figure 2. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram of the article selection process.

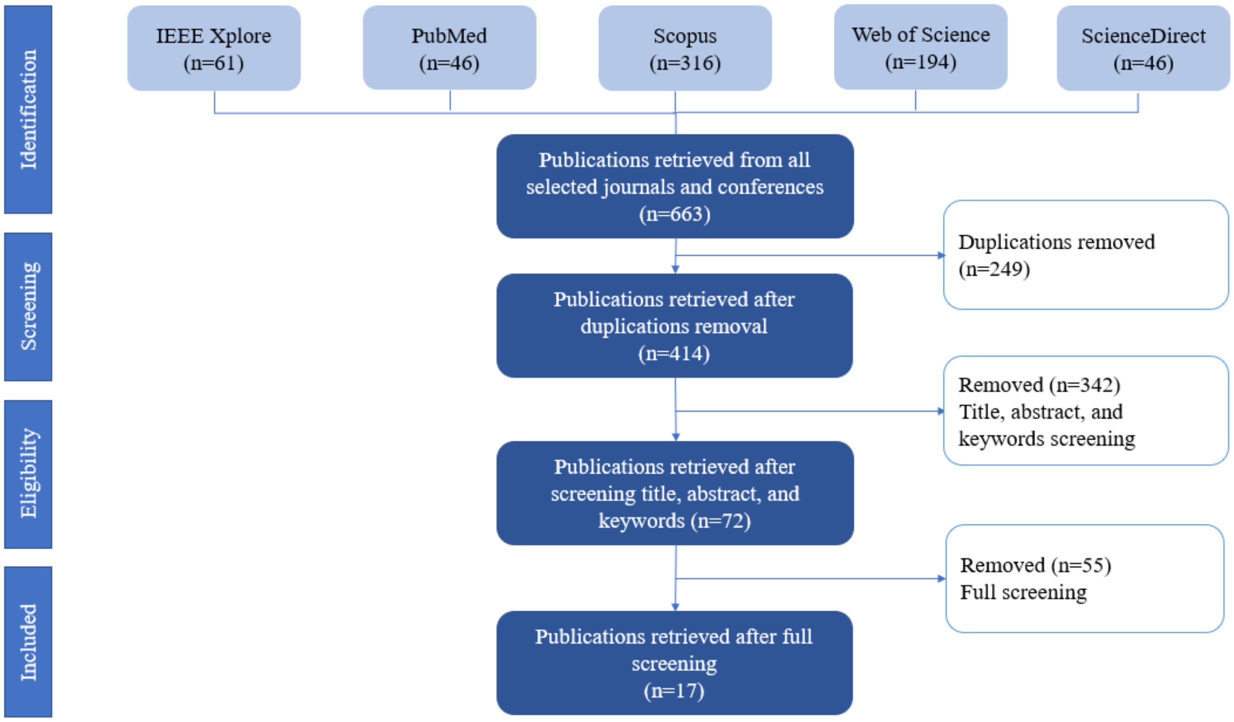


Table 1. Functions of digital twins identified in this study.

Function	Definition
Safety management	Avoiding unintended or unexpected harm to people during the provision of health care processes; identifying crisis-related threats to a system and helping control the impacts
Operational control	Ensuring high system efficiency levels by making sure that health care facilities run smoothly and offer high-quality care to every patient
Information management	Ensuring privacy, security, and real-time information access to stakeholders
Health management and well-being promotion	Empowering self-care abilities by enabling health management practices

These technologies have the potential to avoid unintended or unexpected harm to people during the provision of health care processes and identify crisis-related threats to a system and help control the impacts, which we define as safety management. They are also used to ensure privacy, security, and real-time information access to all stakeholders, which can be considered information management. Another function identified is well-being promotion and health management, where DTs empower the self-care abilities of patients by enabling health management practices. The last function is operational control and management. This technology has the potential to ensure a high system efficiency level by ensuring that health care facilities run smoothly and offer high-quality care to every patient.

Table 2 summarizes the functions of DTs, along with the key findings of the studies considered in this review.

Our findings are presented per the function of DTs (safety management, information management, health management and well-being promotion, and operational control). The operational control findings are presented in 5 subparts (process control of anomalies, scheduling of interventions, resource allocation, operation optimization, and strategy optimization). After presenting the state of the art of DT use, we present the challenges associated with the use and implementation of DTs, followed by the limitations of our study.

Table 2. Functions and key findings.

Function and type of DT[a]	Key findings
Safety management	
System (Jovanović et al, 2021) [21]	The DT used allows management of the infection peak and development of precisely targeted vaccination strategies that allow targeting based on individuals' number of social contacts.
System (Talukder, 2021) [22]	A DT architecture is proposed to ensure safety within an ecosystem disrupted by COVID-19. It mitigates the system challenges and increases patient safety in post–COVID-19 health care delivery.
System (Alrashed et al, 2022) [23]	A DT was used to simulate the different possible strategies and scenarios to predict the spread of the COVID-19 virus and minimize the impacts while ensuring continuity in providing services to citizens.
Information management	
Product (Lutze, 2019) [24]	DTs embedded in wearable devices were used to gather personal information to make group or system decisions.
System (Pang et al, 2021) [25]	A city DT was developed based on federated learning principles to serve as a local central server of information. It allows centralizing information, sharing knowledge, sharing local strategies, and sharing responses to adopted plans in real time among multiple cities.
Health management and well-being promotion	
Product (Díaz et al, 2021) [26]	The DT developed (DTCoach) serves as not only an accompanying educator but also as a mentor that can be used on portable devices like smartphones. It enables web-based pose estimation and performance measurement by providing a person-centered digital coaching experience with a platform that serves as a coach, an accompanying educator, and a mentor who can help make the necessary adjustments based on the individuals' capabilities.
Product (Liu et al, 2019) [27]	The suggested cloud-based system, ClouDTH, uses personal data from digitally twinned wearable medical devices to achieve interaction and convergence between physical and virtual medical spaces to facilitate personal health management for elderly patients.
Product (Tröbinger et al, 2021) [28]	The DT product developed serves as an alternative to telemedicine solutions. It presents a new approach to the remote doctor visit. The dual doctor-patient twin paradigm involves 2 robotic systems (patient GARMI, doctor MUCKI). Control, interaction, and knowledge transfer are enhanced by artificial intelligence, visual motion, and facial expression analysis in the DT. Thus, it enables a transparent remote doctor visit and better-informed and robot-assisted telerehabilitation with bidirectional telepresence control.
Operational control	
Process (Nonnemann et al, 2019) [29]	DTs of the processes in an ICU[b] station are integrated into a system (Health@Hand) to allow remote monitoring; it detects faults and anomalies immediately and will enable interventions at an early stage.
Process (Chase et al, 2021) [30]	DTs of the processes are used to optimize the interventions in a medical ICU. They aim to optimize patient care by clinical staff at the enterprise level for more productivity and quality improvement.
System **(**Karakra et al, 2018) [31]	The hospital's DT proposed developing a predictive decision support model that employs real-time services data drawn from these systems and devices. This model enables assessing the efficiency of existing health care delivery systems and evaluating the impact of changes in services without disrupting the daily activities of the hospital. It allows foreseeing the effectiveness of changes in the models before they are applied in reality.
System (Augusto et al, 2018) [32]	The DT of an emergency unit is developed to optimize the pathway of patient care in the unit. The system accounts for various arrival processes to account for massive arrivals in case of a crisis and determine the best available leverages to optimize the operations of the system.
Process (Mylrea et al, 2021) [33]	BioSecure DT monitors every step in the supply chain process to ensure good productivity and cybersecurity by applying Cyber-Informed Engineering.
Process (Karakra et al, 2019) [34]	It predicts the near future and monitors the processes in real time through the HospiT'Win (DT) system. It allows detection of unexpected situations before problems occur in real life (delay, change in schedule, etc). It will enable the tracking of data flow from the real world to the virtual world.

Function and type of DT[a]	Key findings
System (Karakra et al, 2020) [35]	Using discrete event simulation and DTs through a system called HospiT'Win allows tracking the pathways of patients inside the health care organization to manage growing demand and decrease waiting times and delays. The solution enhances resilience to sustain critical operations under expected and unexpected conditions. It conveys key information to decision-makers in real time.
System (Rodríguez-Aguilar et al, 2020) [36]	The DT of the health care system is developed to better respond to contingencies and ensure optimal allocation of available resources in a DHPES[c].
Process (Croatti et al, 2020) [37]	A trauma DT is used to digitalize and support the process of severe trauma management, considering it as a physical asset that is mirrored by 2 DTs.

[a] DT: digital twin.

[b] ICU: intensive care unit.

[c] DHPES: Digital Health Public Emergency System.

Discussion

State of the Art of DT Usage

Health care has evolved away from focusing solely on illness toward primary health care and health promotion, considering health care as a complex ecosystem [38]. The most significant contributions of digital twinning in health care have been precision medicine efforts that provide patients with targeted treatment and diagnosis [39]. However, its use to develop novel customized health care management approaches started in 2018 and is still an evolving concept [40].

Safety Management

The goal of the patient safety movement is to reduce adverse outcomes or injuries resulting from health care processes. It is imperative that these adverse outcomes are avoided, prevented, or minimized [41]. With the improvements in safety standards and policies, more attention is accorded to analyzing safety issues and the sources of these issues [42]. Errors and inefficiencies in inpatient care are frequently the results of conflicting, incomplete, or suboptimal systems in which patients participate and interfere [43]. The report published by the Institute of Medicine at the beginning of the 21st century resulted in the increased and rapid adoption of health information technology in health care settings, especially for patient safety purposes [44]. For example, the wrong site, wrong side, wrong procedure, and wrong individual (WSWP) errors have been mitigated to some degree by electronic health records. However, these errors continue to be quite significant [45]. DT-assisted safety management systems can be implemented within the Safety 4.0 framework to manage complex safety procedures with minimum human error [46]. In addition to providing operators with contextual information about the surroundings, DTs can guide them through safety tasks [47].

Our study found different DT applications (n=3) that contribute to patient and system safety management through system twinning and product twinning. For instance, an opportunity that DT developers seized was the COVID-19 safety crisis within the health care ecosystem [21]. In 2020, COVID-19 disrupted the health care system and caused an ecosystem crisis that harmed public safety, including resource shortages, misinformation, and medical errors. Virtual interventions and technology initiatives became the preferred mode of service. For example, DTs could accurately contribute to vaccination strategy development. Vaccinations can be targeted based on the number of social contacts of each individual, and infections can be restricted to isolated hotspots and delayed by precisely targeted vaccination, inherent immunity, and public health measures that reduce the infection peak. Thus, DT technology supports decision-making to control the spread of the virus [21]. Apart from this, twinning systems contributed toward predicting the COVID-19 spread. Alrashed et al [23] used a DT system to simulate different strategies and scenarios to minimize the impacts of the virus spread and prevent it while continuing to provide necessary services to citizens with no interruptions to ensure their safety with minimal risk. Safety risk for patients was not only caused by the virus itself but also by the inability of the systems to respond to the new challenges. These challenges will continue to impact the system even after the COVID-19 crisis. Talukder [22] suggested a system architecture that ensures safety within an ecosystem disrupted by COVID-19. It mitigates system-related challenges and increases patient safety in post–COVID-19 health care delivery [22]. In conclusion, DTs (systems and products) can ensure safety management in health care systems by identifying potential threats, redesigning the systems to mitigate hazards, and improving the safety strategies implemented. This leads to the right care at the right price and time for everyone and everywhere at any point of care in a safe manner.

Although the applications of DT technology in patient safety were inspired by COVID-19, it is essential to investigate its potential in other settings and crisis situations. In addition, this technology helps address safety issues in health care without interrupting day-to-day work; it can also be used to address other medical safety issues, such as surgical errors, workplace safety issues, and medical bias in diagnosis.

Information Management

The American Health Information Management Association describes health information management as the process of collecting, analyzing, and securing digital and traditional medical records that are vital to providing quality patient care [48]. Health care organizations seek to analyze patients' information efficiently and quickly, both internally and externally [49]. However, they face many challenges such as privacy, exchange restrictions, and extensive data. Increasing amounts of patient data are forcing health care institutions to replace traditional approaches that cannot cope with increases in data. The United States is taking steps to boost health care information and communication access by leveraging advances in information and communication technologies [50]. Health information management systems have grown rapidly in recent years and are being used to derive important health trends and provide timely preventive care [51]. DT technology can revolutionize clinical research with the changes that it can bring to the basis of health care systems and medical practices. By leveraging this technology, users can better ask questions, get better answers, and gain data-driven actionable insights without compromising the health of real-life subjects. In fact, using a DT, people can gather, aggregate, and represent individual information about their health and well-being [52]. As more data are collected, more DTs will be enabled, leading to more discoveries and better treatment, thus allowing the assembly of more data with less cost, and especially eliminating the risks and consent issues associated with actual human subjects [53]. In our review, we found applications related to information management DTs (n=2).

According to Lutze [24], eHealth systems can manage knowledge by implementing DTs that are based on artificial intelligence. He proposed a DT that collects daily activity data from smart assistance systems linked to wearable sensors for elderly people and extracts behavioral knowledge for information management. The technology suggested accounts for systems, processes, and group changes to provide unbiased conclusions based on learned, trained data. It has a human-centered design, as it allows the self-determination and autonomy of patients to share or refuse the usage of their data with providers and clinical staff. It also establishes solid robustness by automatically tracing the use of all knowledge sources and verifying conclusions drawn about patients after system changes [24]. Auditability is established by tracing the use of all knowledge sources and recording and verifying conclusions drawn about patients and users after system changes. Robustness is supported by automatically checking the containment of a patient within the designated user group of the system and verifying the continued validity of the assessed acceptance conditions after system changes. Moreover, human oversight is facilitated in all critical situations [24].

Another example of an information management DT was suggested by Pang et al [25]. During pandemics, sharing information among different cities and countries in real time through a shared learning model (federated learning) remains critical while ensuring enhanced privacy protection. Pang et al used DT technology in a novel collaborative paradigm that allows DTs in multiple cities to share the local strategy and status quickly without violating any privacy rules to help manage the COVID-19 pandemic [25]. These 2 examples show that DTs allow information management and knowledge extraction by encoding, storing, retrieving, and sharing data in a secure, smart, and real-time environment.

Health Management and Well-being Promotion

Health and well-being goals are challenging to achieve for many individuals. In response to this challenge, a growing number of technologies are being developed to improve people's diet, physical activity, sleep, and mental health. By promoting behavior change and controlling health care costs through modern digital health interventions, people can maintain better health and a healthier lifestyle. Today, a variety of sensors are miniaturized and widely used to track basic physiological indicators on the move to help with better health and well-being management. Moreover, because smartphones are extremely easy to access, mobile health apps are currently considered the most beneficial platform for promoting healthy lifestyles and changing behavior [54]. In modern medicine, personal health management services are viewed as electronic, remote, and digitally enabled care that helps individuals manage their own care and reduce the need for in-clinic visits that are typically expensive and time-consuming [55].

In our review, we found that DTs were used for health management and well-being promotion (n=3). One such DT was introduced by Díaz et al [26] in 2021. Their DT application was called DTCoach. It is a user-centered smart coach that serves as a mentor and an accompanying educator to the users. It helps the users make the necessary adjustments in their posture and performance based on measurements taken that characterize their individual capabilities [26].

Another example was ClouDTH, suggested by Liu et al [27]. This cloud-based health care system uses personal data from digitally twinned wearable medical devices to achieve a convergent interaction between the medical and physical spaces, and their virtual twins. Then, it facilitates self-management of health for elderly patients [27]. Patients' needs depend on many factors, and age is one of them. Elderly patients have higher demands for many medical services. Therefore, DTs are used extensively in geriatric care to support health management promotion of elderly patients. Furthermore, another example that we cover in our review was introduced by Tröbinger [28], which is a new DT approach serving as an alternative to telemedicine. It consists of a transparent remote doctor visit and a better-informed and robot-assisted telerehabilitation initiative that allows bidirectional telepresence control [28]. In summary, DTs are used to accompany patients and give them control over their health by promoting well-being and lifestyle activities and supervising them to maximize their performance in a safe environment.

Operational Control

Operating a health care facility on a day-to-day basis impacts patient experiences and organizational goals [56]. Thus, operations management helps in understanding and optimizing the business processes inside medical departments to reduce and alleviate the effects of overcrowding, waiting times, delays, and other problems that facilities are facing [57]. DTs have the potential to contribute to the effective operation of health care units. Our review found that most DT initiatives to improve health services have health care operations management focusing on revolutionizing clinical processes and enhancing medical care. They replicate hospitals or treatment facilities and help improve their performance in a safe manner with less risk. Applications are numerous and range from predicting resource shortage to managing patient flow. Using DT technology, an institution can execute a digital stress test to observe how the technology would fare under extreme conditions like crises. By creating a virtual twin of a hospital, stakeholders can review the operational strategy, capacity, staffing, and care model on the DT to determine what actions to take and mitigate future challenges. Our review identified extensive efforts (9 out of 17studies) to support operational control and account for health care system challenges. The efforts consisted of digitally twinned systems and processes for performance control.

Detecting anomalies in processes is essential to prevent hazards and predict the corrective actions that need to be implemented. With some interventions, such as the solution suggested by Croatti et al [37], DTs can be used to support physical processes through their digital representations and monitor their changes. In this study, by mirroring the real system by building an agent-based smart DT, they aimed to digitalize and support the process of severe trauma management [37]. This DT represents the operative phase of trauma management and starts when the trauma is marked as severe in the previous phase. The fact that this DT starts before the patient's arrival to the unit is very important for this case study [37]. In this way, the trauma team is prealerted about the incoming patient and starts collecting and receiving information directly from the accident site. Its internal state changes when the patient is delivered to the emergency department, where the trauma team starts taking care of the patient. A very preliminary version of a system prototype has been developed according to the designed conceptual model [37].

Another example is Health@Hand, suggested by Nonnemann et al [29]. They twinned the processes of an intensive care unit (ICU) station and integrated them into a digital system (Health@Hand) to allow remote monitoring of the processes. With this intervention, hospital managers can detect anomalies and faults immediately and intervene in an early stage. Moreover, while improving the productivity and efficiency of processes, some digital interventions forget to address the problem of cybersecurity, which may harm the systems. As a solution to this problem, Mylrea et al [33] propose BioSecure. It is a process twinning that allows managers to monitor every step in the supply chain process to ensure good productivity and cybersecurity to secure the system and data by applying Cyber-Informed Engineering.

One of the challenges in health care is providing an optimized scheduling strategy that can effectively use the hospital's resources and prevent delays, errors, and long lengths of stay. An application in the same settings (ICU and process twins) was developed by Chase et al [30] that aimed to optimize patient care by clinician staff at the enterprise level to improve the productivity of the staff and the quality of care delivered to patients. In addition, using resources effectively has always been a challenging decision for managers in all industries. In health care, resource allocation needs to be regulated by providing efficient services on time. Rodríguez-Aguilar et al [36] suggested a DT for a hospital that supports resource allocation (financial and human) called the Digital Health Public Emergency System (DHPES). DHPES provides efficient health services [36]. The DT design seeks to generate virtual instances that emulate the real operation of the provision of highly specialized public services, including the supply of medications, supplies, devices, and equipment as well as the management of human resources and financial resources in the event of a contingency [36].

Another DT initiative called HospiT'Win was developed to manage the patients' pathways inside the hospitals [34,35]. This DT can help hospitals equilibrate demand and supply and control the growing workload while reducing waiting times, lengths of stay, and delays. It also provides key information to decision-makers in real time, controlling the real-time flow of data. As high demand can disturb health care systems, such DT systems might be useful in times of crisis. Hospit'Win performed well during the COVID-19 pandemic by managing the demand [34,35].

Some operational strategies need to be implemented first to evaluate their efficiency. In hospitals, evaluating interventions would disrupt the daily services and activities of the units. This is where DTs could be most useful. Karakra et al [31] proposed a decision support system that employs real-time services data drawn from real systems and devices to enable evaluation of the impact of changes in services without disrupting the daily activities of the hospital. This idea allows foreseeing the effectiveness of changes in the models before they are applied in reality [31].

Providing patients safe and high-quality care is a demanding process. A DT framework for a system is proposed by Augusto et al [32] to optimize a patient’s care pathway in health care units. The system accounts for various arrival processes and simulates different scenarios to determine the best available leverages to optimize the operations of the system even under high demand and variability to account for uncertainties [32]. The framework has been conceived and tested in close collaboration with health care professionals to be as close to the real system as possible. Furthermore, the framework is intended to be used regularly by the head of the emergency unit [32]. Data collection was performed using the hospital information system for the following parameters: patient arrivals, the total length of stay; type of patient including less critical (fast and normal track), moderately critical, and life-threatening emergencies; and number of requested paraclinical examinations per patient. On the other hand, processing times were recorded by interviewing doctors, nurses, caregivers, and interns because the related data in the hospital information system were not reliable enough [32]. The model was validated and shown to reduce the waiting time and length of stay in different scenarios.

Future of DTs: Challenges Associated With Use and Implementation

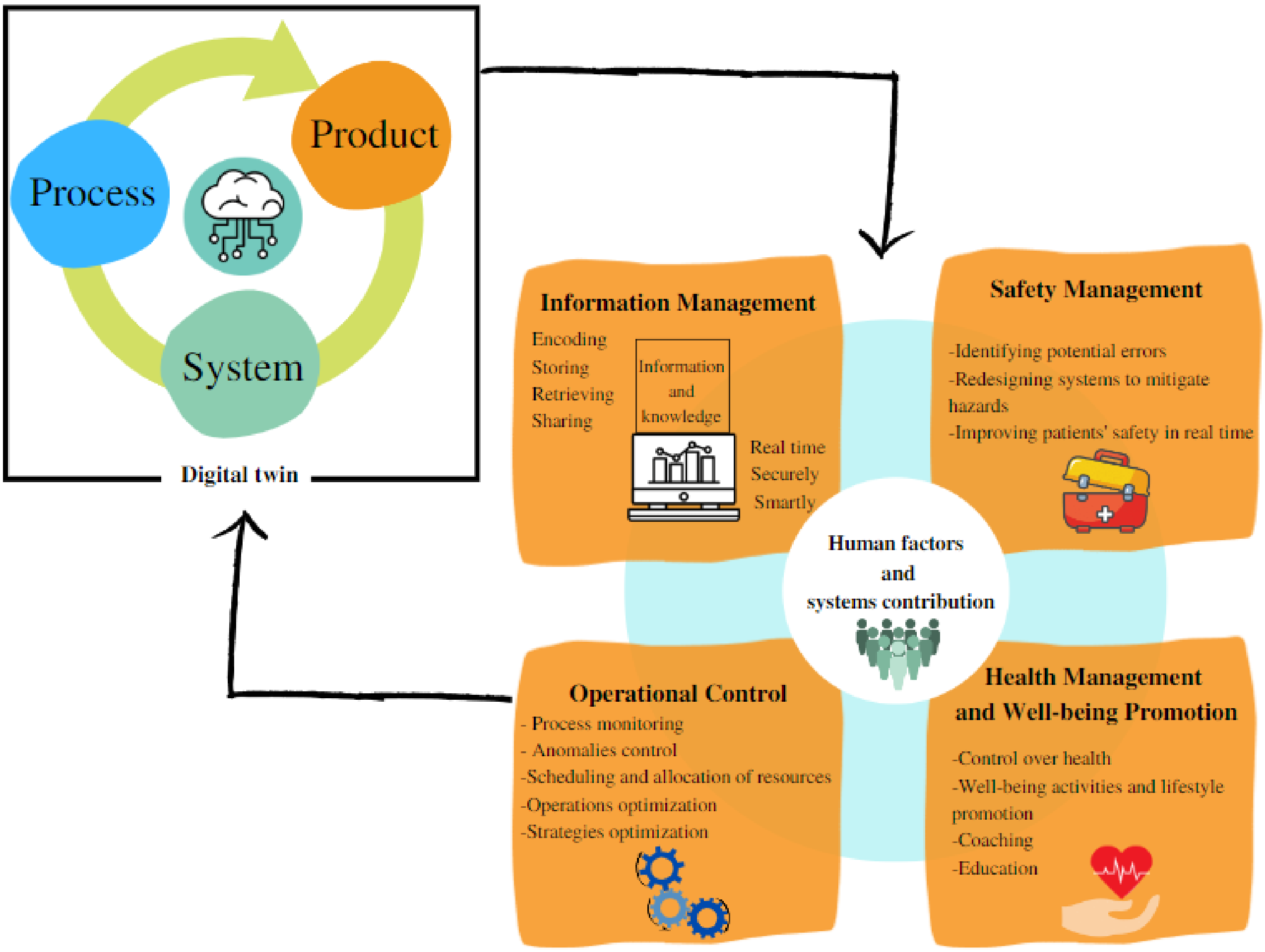
A DT system would provide patients with a safe and secure monitoring option; medical staff would have safe and secure monitoring methods, and authorities would be able to track extreme crisis scenarios in real time accurately. However, digital twinning is facing many challenges that are hindering its growth. The first obstacle is the infrastructure of data flow. For example, to prevent false positives in the digitally twinned sensors, we need a good understanding of the variability in the personal parameters and characteristics of the users [58]. Another concern to be addressed before DTs can go mainstream is data security and privacy. The data used by DTs are confidential and sensitive, and interconnected devices are an easy target for cyberattacks that can harm health care systems. Therefore, governments and policy makers need to consolidate the law regulation factor to have more protected data-sharing procedures. Moreover, as data form the core of DTs, quality control protocols need to be embedded in the real physical systems to merge the data with the simulated systems (twins) and ensure good performance of the DT models. Additionally, ethical concerns like ownership of the data extracted with DTs are still not addressed. Finally, it is essential to explore factors that affect DT implementation and adoption.

DTs also have the potential to offer new important pathways for various care processes in health care. For instance, strategies to improve communication and patient-centered care can be implemented digitally to evaluate their effectiveness before adopting them in real life to avoid repetitive trials that may disturb patients' pathways. In addition, some patients are hard to deal with because of their critical medical situations. Digital twinning of the care processes for these patients can give more visibility to health care professionals to understand the best possible care strategy for these patients. Furthermore, usability studies are sometimes costly in terms of facilities, equipment, and time. DTs can facilitate remote usability testing across diverse populations, accounting for their lower literacy or health literacy and individuals with cognitive or physical disabilities. They can also help testers gain time, reduce effort, and earn money while providing real-time decision support by solving recruitment problems for surveys, interviews, and clinical trials.

As illustrated in Figure 3, we suggest a framework that highlights the possible contributions of DTs from human factors and systems perspectives. Irrespective of whether the technology designed is by twinning a system, process, or tool, a good DT design can potentially improve safety management, improve operational control of the health care system, allow better information management, and promote the health and well-being of patients.

Finally, this study also has several limitations. The included studies largely reported postintervention data, so we could not determine the preintervention-to-postintervention change or ascertain whether the intervention groups were matched at baseline for key characteristics and outcome measure scores. In addition, we may have missed some articles in our screening because the research was limited to the following databases: PubMed, Web of Science, IEEE Xplore, Scopus, and ScienceDirect.

Figure 3. Framework of the impact classification of digital twins from human factors and systems perspectives.



Conclusions

DTs are replications of systems, products, or processes that bridge reality using data and expand the same to virtual models. In medical services, DTs are primarily used in personalized medicine; however, they also have the potential to be used at the system level. These applications vary from safety to information management, health and well-being promotion, and operations control. This rapid review shows that digital twinning for health care system management is still an emerging field with considerable potential that was also used during the COVID-19 pandemic. Therefore, interdisciplinary teams from various disciplines, including human factors and ergonomics, human-computer interaction, data science, and digital health, should further investigate the potential of this technology and address the challenges that may influence the design and adoption of this technology in the health care system.

Authors' Contributions

Both SE and OA conceptualized the study and participated in the literature review, finding interpretation, and manuscript writing. Both the authors have approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Chraibi A, El Cadi AA, Kharraja S, Artiba A. Using discrete event simulation to evaluate operating theater layout. IFAC-PapersOnLine 2016;49(12):95-98 [FREE Full text] [doi: 10.1016/j.ifacol.2016.07.556]
2. Sarabu VK. Health economics - an over view - a way forward. ResearchGate. 2020. URL: https://www.researchgate.net/publication/346954593_HEALTH_ECONOMICS_-AN_OVER_VIEW_-_A_WAY_FORWARD [accessed 2022-05-10]
3. Livinus VMF, Adhikara MA, Kusumapradja R. Hospital management information system usefulness in the health services industry at Indonesia: mandatory or voluntary? JMMR (Jurnal Medicoeticolegal dan Manajemen Rumah Sakit) 2021 Apr;10(1):1-12. [doi: 10.18196/jmmr.v10i1.10296]
4. ElKefi S, Asan O. How technology impacts communication between cancer patients and their health care providers: A systematic literature review. Int J Med Inform 2021 May;149:104430 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104430] [Medline: 33684711]
5. Liao Y, Deschamps F, Loures EDFR, Ramos LFP. Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal. Int J Prod Res 2017 Mar 28;55(12):3609-3629. [doi: 10.1080/00207543.2017.1308576]
6. Hsu Y, Chiu J, Liu J. Digital twins for industry 4.0 and beyond. 2019 Dec Presented at: IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); December 15-18, 2019; Macao, China p. 526-530. [doi: 10.1109/IEEM44572.2019.8978614]
7. Jeske SJ. Digital twins in healthcare [dissertation]. Paris Lodron University of Salzburg and Aalborg University in Copenhagen. 2020. URL: https://projekter.aau.dk/projekter/files/360456256/Jeske_MasterThesis.pdf [accessed 2022-05-10]
8. Grieves M, Vickers J. Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen FJ, Flumerfelt S, Alves A, editors. Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. Cham, Switzerland: Springer; Aug 17, 2016:85-113.
9. Cimino C, Negri E, Fumagalli L. Review of digital twin applications in manufacturing. Comput Ind 2019 Dec;113:103130. [doi: 10.1016/j.compind.2019.103130]
10. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. Am J Med 2019 Jul;132(7):795-801 [FREE Full text] [doi: 10.1016/j.amjmed.2019.01.017] [Medline: 30710543]
11. Uppal M, Gupta D, Juneja S, Dhiman G, Kautish S. Cloud-Based Fault Prediction Using IoT in Office Automation for Improvisation of Health of Employees. J Healthc Eng 2021 Nov;2021:8106467 [FREE Full text] [doi: 10.1155/2021/8106467] [Medline: 34777738]
12. Lopez-Arevalo I, Gonzalez-Compean JL, Hinojosa-Tijerina M, Martinez-Rendon C, Montella R, Martinez-Rodriguez JL. A WoT-Based Method for Creating Digital Sentinel Twins of IoT Devices. Sensors (Basel) 2021 Aug 17;21(16):5531 [FREE Full text] [doi: 10.3390/s21165531] [Medline: 34450973]
13. Corral-Acero J, Margara F, Marciniak M, Rodero C, Loncaric F, Feng Y, et al. The 'Digital Twin' to enable the vision of precision cardiology. Eur Heart J 2020 Dec 21;41(48):4556-4564 [FREE Full text] [doi: 10.1093/eurheartj/ehaa159] [Medline: 32128588]
14. Dillenseger A, Weidemann ML, Trentzsch K, Inojosa H, Haase R, Schriefer D, et al. Digital Biomarkers in Multiple Sclerosis. Brain Sci 2021 Nov 16;11(11):1519 [FREE Full text] [doi: 10.3390/brainsci11111519] [Medline: 34827518]
15. Al Faruque MA, Muthirayan D, Yu SY, Khargonekar PP. Cognitive digital twin for manufacturing systems. 2021 Presented at: 2021 Design, Automation & Test in Europe Conference & Exhibition (DATE); February 1-5, 2021; Grenoble, France p. 440-445. [doi: 10.23919/date51398.2021.9474166]
16. Digital twin. Siemens. URL: https://www.plm.automation.siemens.com/global/en/our-story/glossary/digital-twin/24465 [accessed 2022-05-01]

17. Tricco A, Khalil H, Holly C, Feyissa G, Godfrey C, Evans C, et al. Rapid reviews and the methodological rigor of evidence synthesis: a JBI position statement. JBI Evid Synth 2022 Apr;20(4):944-949. [doi: 10.11124/JBIES-21-00371] [Medline: 35124684]
18. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
19. Elkefi S, Asan O. Digital twins for better healthcare management: rapid literature review. OSF. 2022 Jan. URL: https://osf.io/5hupx/ [accessed 2022-04-25]
20. Grieves MW. Virtually intelligent product systems: digital and physical twins. In: Flumerfelt S, Schwartz KG, Mavris D, Briceno S, editors. Complex Systems Engineering: Theory and Practice. Reston, VA: American Institute of Aeronautics and Astronautics; Aug 2019:175-200.
21. Jovanović R, Davidović M, Lazović I, Jovanović M, Jovašević-Stojanović M. Modelling voluntary general population vaccination strategies during COVID-19 outbreak: influence of disease prevalence. Int J Environ Res Public Health 2021 Jun;18(12):6217 [FREE Full text] [doi: 10.3390/ijerph18126217] [Medline: 34201285]
22. Talukder AK. Bridging the inferential gaps in healthcare. 2021 Dec Presented at: BDA 2021: International Conference on Big Data Analytics; December 15-18, 2021; Virtual p. 31-43. [doi: 10.1007/978-3-030-93620-4_3]
23. Alrashed S, Min-Allah N, Ali I, Mehmood R. COVID-19 outbreak and the role of digital twin. Multimed Tools Appl 2022 Jan;81(19):26857-26871 [FREE Full text] [doi: 10.1007/s11042-021-11664-8] [Medline: 35002471]
24. Lutze R. Digital twins in eHealth: prospects and challenges focussing on information management. 2019 Presented at: 2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC); June 17-19, 2019; Valbonne Sophia-Antipolis, France p. 1-9. [doi: 10.1109/ice.2019.8792622]
25. Pang J, Huang Y, Xie Z, Li J, Cai Z. Collaborative city digital twin for the COVID-19 pandemic: A federated learning solution. Tsinghua Sci Technol 2021 Oct;26(5):759-771. [doi: 10.26599/tst.2021.9010026]
26. Diaz RG, Laamarti F, El Saddik A. DTCoach: Your Digital Twin Coach on the Edge During COVID-19 and Beyond. IEEE Instrum Meas Mag 2021 Sep;24(6):22-28. [doi: 10.1109/mim.2021.9513635]
27. Liu Y, Zhang L, Yang Y, Zhou L, Ren L, Wang F, et al. A novel cloud-based framework for the elderly healthcare services using digital twin. IEEE Access 2019 Apr;7:49088-49101. [doi: 10.1109/access.2019.2909828]
28. Tröbinger M, Costinescu A, Xing H, Elsner J, Hu T, Naceri A, et al. A dual doctor-patient twin paradigm for transparent remote examination, diagnosis, and rehabilitation. 2021 Presented at: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); September 27-October 1, 2021; Prague, Czech Republic p. 2933-2940. [doi: 10.1109/IROS51168.2021.9636626]
29. Nonnemann L, Haescher M, Aehnelt M, Bieber G, Diener H, Urban B. Health@Hand a visual interface for eHealth monitoring. 2019 Presented at: 2019 IEEE Symposium on Computers and Communications (ISCC); June 29-July 3, 2019; Barcelona, Spain p. 1093-1096. [doi: 10.1109/iscc47284.2019.8969647]
30. Chase JG, Zhou C, Knopp JL, Shaw GM, Näswall K, Wong JHK, et al. Digital Twins in Critical Care: What, When, How, Where, Why? IFAC-PapersOnLine 2021;54(15):310-315. [doi: 10.1016/j.ifacol.2021.10.274]
31. Karakra A, Fontanili F, Lamine E, Lamothe J, Taweel A. Pervasive computing integrated discrete event simulation for a hospital digital twin. 2018 Presented at: IEEE/ACS 15th international conference on computer systems and Applications (AICCSA); October 28-November 1, 2018; Aqaba, Jordan p. 1-6. [doi: 10.1109/aiccsa.2018.8612796]
32. Augusto V, Murgier M, Viallon A. A modelling and simulation framework for intelligent control of emergency units in the case of major crisis. 2018 Presented at: 2018 Winter Simulation Conference (WSC); December 9-12, 2018; Gothenburg, Sweden p. 2495-2506. [doi: 10.1109/wsc.2018.8632438]
33. Mylrea M, Fracchia C, Grimes H, Austad W, Shannon G, Reid B, et al. BioSecure digital twin: manufacturing innovation and cybersecurity resilience. In: Lawless WF, Llinas J, Sofge DA, Mittu R, editors. Engineering Artificially Intelligent Systems. Cham, Switzerland: Springer; 2021:53-72.
34. Karakra A, Fontanili F, Lamine E, Lamothe J. HospiT'Win: a predictive simulation-based digital twin for patients pathways in hospital. 2019 Presented at: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); May 19-22, 2019; Chicago, IL p. 1-4. [doi: 10.1109/BHI.2019.8834534]
35. Karakra A, Lamine E, Fontanili F, Lamothe J. HospiT’Win: a digital twin framework for patients’ pathways real-time monitoring and hospital organizational resilience capacity enhancement. 2020 Presented at: International Workshop on Innovative Simulation for Healthcare (IWISH 2020); September 16-18, 2020; Virtual p. 62-71. [doi: 10.46354/i3m.2020.iwish.012]
36. Rodríguez-Aguilar R, Marmolejo-Saucedo JA. Conceptual framework of Digital Health Public Emergency System: digital twins and multiparadigm simulation. EAI Endorsed Trans Pervasive Health Technol 2020 Jan;6(21):e3. [doi: 10.4108/eai.13-7-2018.164261]
37. Croatti A, Gabellini M, Montagna S, Ricci A. On the integration of agents and digital twins in healthcare. J Med Syst 2020 Aug;44(9):161 [FREE Full text] [doi: 10.1007/s10916-020-01623-5] [Medline: 32748066]
38. Edington DW, Schultz AB, Pitts JS, Camilleri A. The future of health promotion in the 21st century: a focus on the working population. Am J Lifestyle Med 2016 Sep;10(4):242-252 [FREE Full text] [doi: 10.1177/1559827615605789] [Medline: 30202279]

39. Rivera LF, Jiménez M, Angara P, Villegas NM, Tamura G, Müller HA. Towards continuous monitoring in personalized healthcare through digital twins. In: Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering. 2019 Presented at: 29th Annual International Conference on Computer Science and Software Engineering; November 4-6, 2019; Toronto, ON p. 329-335.
40. Ahmadi-Assalemi G, Al-Khateeb H, Maple C, Epiphaniou G, Alhaboby ZA, Alkaabi S, et al. Digital twins for precision healthcare. In: Jahankhani H, Kendzierskyj S, Chelvachandran N, Ibarra J, editors. Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity. Advanced Sciences and Technologies for Security Applications. Switzerland: Springer, Cham; 2020:133-158.
41. Alotaibi YK, Federico F. The impact of health information technology on patient safety. Saudi Med J 2017 Dec;38(12):1173-1180 [FREE Full text] [doi: 10.15537/smj.2017.12.20631] [Medline: 29209664]
42. Vincent C, Amalberti R. Safety in healthcare is a moving target. BMJ Qual Saf 2015 Sep;24(9):539-540 [FREE Full text] [doi: 10.1136/bmjqs-2015-004403] [Medline: 26150547]
43. Carayon P, Schoofs HA, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15(Suppl 1):i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
44. Institute of Medicine. In: Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington, DC: National Academic Press; 2000.
45. Tanner C, Gans D, White J, Nath R, Pohl J. Electronic health records and patient safety. Appl Clin Inform 2017 Dec;6(1):136-147. [doi: 10.4338/aci-2014-11-ra-0099]
46. Wang B. Safety intelligence as an essential perspective for safety management in the era of Safety 4.0: from a theoretical to a practical framework. Process Saf Environ Prot 2021 Apr;148:189-199. [doi: 10.1016/j.psep.2020.10.008]
47. Gattullo M, Scurati GW, Evangelista A, Ferrise F, Fiorentino M, Uva AE. Informing the use of visual assets in industrial augmented reality. 2019 Sep Presented at: International Conference of the Italian Association of Design Methods and Tools for Industrial Engineering; September 9-10, 2019; Modena, Italy p. 106-117. [doi: 10.1007/978-3-030-31154-4_10]
48. Health information is human information. AHiMA. URL: https://www.ahima.org/certification-careers/certifications-overview/career-tools/career-pages/health-information-101/ [accessed 2022-05-10]
49. Alshawi S, Missi F, Eldabi T. Healthcare information management: the integration of patients' data. Logist Inf Manag 2003 Jun;16(3/4):286-295. [doi: 10.1108/09576050310483772]
50. Lustria MLA, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]
51. Chen PT, Lin CL, Wu WN. Big data management in healthcare: adoption challenges and implications. Int J Inf Manage 2020 Aug;53:102078. [doi: 10.1016/j.ijinfomgt.2020.102078]
52. Exploring the possibilities offered by digital twins in medical technology. SIEMENS Healthineers. 2018. URL: https://corporate.webassets.siemens-healthineers.com/1800000005899262/fcb74e87168b/Exploring-the-possibilities-offered-by-digital-twins-in-medical-technology_1800000005899262.pdf [accessed 2022-05-10]
53. Ghazanfari A. How digital-twin technology could revolutionise the healthcare industry. ComputerWeekly.com. 2022 Jan. URL: https://www.computerweekly.com/opinion/How-Digital-Twin-Technology-Could-Revolutionise-the-Healthcare-Industry [accessed 2022-05-02]
54. Poole ES. HCI and mobile health interventions: how human-computer interaction can contribute to successful mobile health interventions. Transl Behav Med 2013 Dec;3(4):402-405 [FREE Full text] [doi: 10.1007/s13142-013-0214-3] [Medline: 24294328]
55. Sellak H, Grobler M. mHealth4U: designing for health and wellbeing self-management. 2020 Presented at: 2020 35th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW); September 21, 2020; Melbourne, Australia p. 41-46. [doi: 10.1145/3417113.3422179]
56. Henderson MD. Operations management in health care. J Health Care Finance 1995;21(3):44-47. [Medline: 7600237]
57. Soremekun O, Terwiesch C, Pines J. Emergency medicine: an operations management view. Acad Emerg Med 2011 Dec;18(12):1262-1268 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01226.x] [Medline: 22168189]
58. Patrone C, Lattuada M, Galli G, Revetria R. The role of internet of things and digital twin in healthcare digitalization process. 2018 Presented at: The World Congress on Engineering and Computer Science; October 23-25, 2018; San Fransisco, CA p. 30-37. [doi: 10.1007/978-981-15-6848-0_3]

Abbreviations

DHPES: Digital Health Public Emergency System
DT: digital twin
ICU: intensive care unit

Edited by R Kukafka; submitted 01.03.22; peer-reviewed by X Xu, M Zahmatkeshan, L Suppan; comments to author 09.05.22; revised version received 30.05.22; accepted 28.07.22; published 16.08.22.

Please cite as:
Elkefi S, Asan O
Digital Twins for Managing Health Care Systems: Rapid Literature Review
J Med Internet Res 2022;24(8):e37641
URL: https://www.jmir.org/2022/8/e37641
doi:10.2196/37641
PMID:35972776

Review

Effects of eHealth Interventions on Quality of Life and Psychological Outcomes in Cardiac Surgery Patients: Systematic Review and Meta-analysis

Ruping Ni[1,2*], BS; Maobai Liu[1,2*], MM; Shunmin Huang[1,2], MS; Jing Yang[1,2], MS

[1]Department of Pharmacy, Fujian Medical University Union Hospital, Fuzhou, China
[2]College of Pharmacy, Fujian Medical University, Fuzhou, China
[*]these authors contributed equally

Corresponding Author:
Jing Yang, MS
Department of Pharmacy
Fujian Medical University Union Hospital
29 Xinquan Road
Gulou District
Fuzhou, 350001
China
Phone: 86 13706987692
Email: jiangyang@fjmu.edu.cn

Abstract

Background: Patients undergoing heart surgery may experience a range of physiological changes, and the postoperative recovery time is long. Patients and their families often have concerns about quality of life (QoL) after discharge. eHealth interventions may improve patient participation, ensure positive and effective health management, improve the quality of at-home care and the patient's quality of life, and reduce rates of depression.

Objective: The purpose of this study was to evaluate the effects of eHealth interventions on the physiology, psychology, and compliance of adult patients after cardiac surgery to provide a theoretical basis for clinical practice.

Methods: We conducted systematic searches of the following 4 electronic databases: PubMed, Embase, CINAHL, and the Cochrane Central Register of Controlled Trials. Mean (SD) values were used to calculate the pooled effect sizes for all consecutive data, including QoL, anxiety, and depression. Where the same results were obtained using different instruments, we chose the standardized mean difference with a 95% CI to represent the combined effect size; otherwise, the mean difference (MD) with a 95% CI was used. Odds ratios were used to calculate the combined effect size for all dichotomous data. The Cohen Q test for chi-square distribution and an inconsistency index (I^2) were used to test for heterogeneity among the studies. We chose a fixed-effects model to estimate the effect size if there was no significant heterogeneity in the data ($I^2 \leq 50\%$); otherwise, a random-effects model was used. The quality of the included studies was assessed using the Cochrane risk-of-bias tool for randomized trials (RoB 2).

Results: The search identified 3632 papers, of which 19 met the inclusion criteria. In terms of physical outcomes, the score of the control group was lower than that of the intervention group (MD 0.15, 95% CI 0.03-0.27, I^2=0%, P=.02). There was no significant difference in the mental outcomes between the intervention and control groups (MD 0.10, 95% CI –0.03 to 0.24, I^2=46.4%, P=.14). The control group's score was lower than that of the intervention group for the depression outcomes (MD –0.53, 95% CI –0.89 to –0.17, I^2=57.1%, P=.004). Compliance outcomes improved in most intervention groups. The results of the sensitivity analysis were robust. Nearly half of the included studies (9/19, 47%) had a moderate to high risk of bias. The quality of the evidence was medium to low.

Conclusions: eHealth improved the physical component of quality of life and depression after cardiac surgery; however, there was no statistical difference in the mental component of quality of life. The effectiveness of eHealth on patient compliance has been debated. Further high-quality studies on digital health are required.

Trial Registration: PROSPERO CRD42022327305; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=327305

(J Med Internet Res 2022;24(8):e40090) doi:10.2196/40090

KEYWORDS

eHealth; eHealth intervention; cardiac surgery; depression; anxiety; quality of life; meta-analysis; heart disease; surgery; heart surgery; post-operative; postoperative; mental health; home care; digital health intervention; digital health; outcomes; psychological; physiological; physiology; psychology; compliance

Introduction

Quality of Life and Cardiac Surgery

More than 1.5 million patients worldwide undergo heart surgery annually, and this number continues to grow [1]. Patients experience a series of psychophysiological changes before and after surgery. Preoperative anxiety and depression trigger the psychological response system, which in turn activates the endocrine and autonomic nervous systems, affecting postoperative outcomes, length of hospital stay, and quality of life [2,3]. Moreover, psychological changes related to chronic stress, such as anxiety and depression, can affect not only quality of life but also physiological parameters such as respiratory rate, heart rate, blood pressure, inflammatory markers, and brain activity, which may be detrimental to postoperative recovery [4-6]. The recovery period after cardiac surgery is relatively long, and most of the recovery processes, such as the healing of surgical wounds and the recovery of cardiac function, take place at home or in other facilities outside the hospital [7]. After cardiac surgery, patients and their families often have concerns regarding quality of life after discharge [8,9], since they will be solely responsible for at-home care [10,11]. Many problems can arise due to a lack of self-care knowledge and skills, and these problems increase with inadequate follow-up for patient education, counseling, and postoperative care [12].

eHealth Interventions

In recent years, both health professionals and patients have been increasingly involved in eHealth [13], which includes mobile health, mobile and wireless technologies, health information technology, telemedicine, and personalized medicine, to improve clinical care, such as public health, health administration, and health-related education [14]. eHealth is often designed to support the achievement of health goals. With the increasing social demand for electronic technology, the use of mobile devices has the great potential to transform conventional health care and implement patient-centered initiatives [15-17]. Increased patient engagement is a key factor in eHealth and has the potential to motivate users and enable them to become more proactive and effective in managing their own health, ultimately improving quality of care [18]. The quality of health care has improved significantly with the use of telemedicine [19]. In addition, electronic medical interventions are already widely used in perioperative nursing [20].

Approximately 70% of patients consult the internet for information soon after learning about their upcoming surgery [21,22]. Studies [23,24] have reported that eHealth interventions for cardiac rehabilitation can also improve patients' quality of life. These interventions provide continuous education regarding patient care and treatment, and offer counseling and support to at-home care providers while allowing access to vital information for patients, their families, and health care providers [25].

Many studies have evaluated the potential benefits of eHealth interventions in patients who have undergone cardiac surgery. However, to date, there has been no systematic evaluation of the effectiveness of these eHealth interventions compared to conventional care in terms of physiology, psychology, and compliance of adult patients after cardiac surgery. Therefore, we conducted this systematic review to assess the impact of eHealth interventions after cardiac surgery on quality of life, psychology, and compliance.

Methods

Design

This study was conducted and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Multimedia Appendix 1) [26]. The systematic review protocol was registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42022327305).

Search Strategy and Data Sources

The PubMed, Embase, CINAHL, and the Cochrane Central Register of Controlled Trials databases were searched from inception to April 2022. The search strategy consisted of 2 components: clinical situation (adult, cardiac surgery) and intervention type (health management using mobile phones, wearables, personal digital assistants, and other wireless devices). Relevant search items and combinations of Medical Subject Headings were used to identify trials related to eHealth and cardiac surgery. Searches were not limited to a specific geographic region, language, or period, but any literature without its full text was excluded. We exclusively included randomized controlled trials. The exact search terms used in each of the databases and the corresponding number of results are provided (Multimedia Appendix 2). EndNote 20 (Clarivate) was used for database management.

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: (1) patients aged 18 years or older at the time of heart surgery and studies that did not specify the type of heart surgery, (2) studies that evaluated any type of eHealth intervention, and (3) randomized controlled clinical studies.

Exclusion criteria were as follows: (1) studies where the full text could not be obtained, (2) insufficient clinical data that were reported in the form of meeting abstracts and did not provide detailed treatment methods or report the relevant results, and (3) duplicate studies.

Document Screening and Data Extraction

Two researchers (RN and ML) independently performed the literature screening and data extraction according to the literature inclusion and exclusion criteria. Decisions on inclusion or exclusion were made by the researchers after a joint discussion of the results. Disagreements were resolved by a third party.

One researcher extracted the data using a literature data extraction table, and a second researcher confirmed the accuracy and authenticity of the data. The extracted content included study information (research topic, author, publication date, and region), baseline characteristics of the study participants (sample size and age), specific details of the intervention, follow-up time, and other outcome indicators (quality of life, anxiety and depression, cardiovascular events, treatment, and medication compliance).

Data Analyses

Mean (SD) values were used to calculate the pooled effect sizes for all consecutive data, including quality of life and depression. When measuring the same outcome using different instruments, we chose the standardized mean difference with a 95% CI to represent the combined effect size; otherwise, we used the mean difference (MD) with a 95% CI to represent the combined effect size. Odds ratios were used to calculate the combined effect size for all dichotomous data. The Cohen Q test for chi-square distribution and an inconsistency index (I^2) were used to test for heterogeneity among the studies. We selected a fixed-effects model to estimate the effect size if there was no significant heterogeneity in the data ($I^2 \leq 50\%$). Otherwise, a random effects model was used. A sensitivity analysis was performed using the leave-one-out method. All meta-analyses were performed using the Stata software (version 15.1; StataCorp).

Quality Assessment

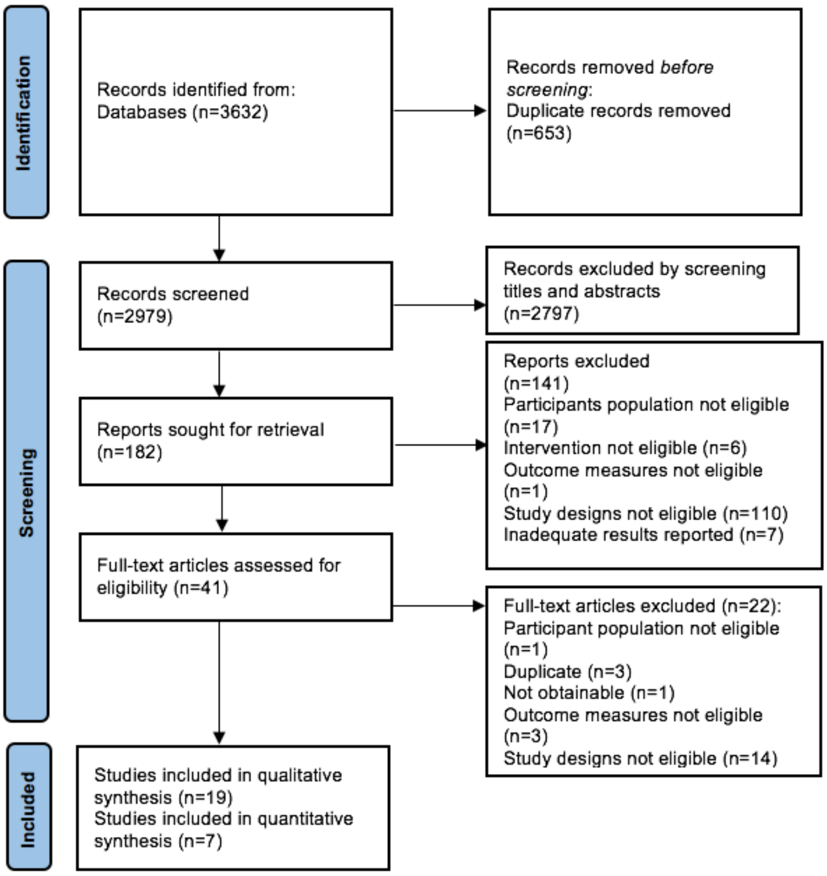
Before analyzing the extracted data, 2 researchers independently assessed the quality of the included studies. A discussion with a third reviewer was conducted until a consensus was reached and disagreements were resolved. The quality of each study was assessed according to the guidelines provided by the Cochrane risk-of-bias tool for randomized trials, version 2.0 (RoB 2) [27]. The overall quality of evidence for each outcome was assessed using the GRADE (Grading of Recommendations, Assessment, Development and Evaluations) approach [28].

Results

Identification of Studies

The PRISMA flowchart in Figure 1 summarizes the search results and selection process for all studies included in our synthesis. A total of 3632 articles were retrieved through a systematic literature search. After removing duplicate studies, the remaining 2979 records were screened. After reading 41 eligible full-text articles, 22 were excluded, and 19 were selected [29-47]. A summary of the study characteristics and participant demographics are presented in Multimedia Appendix 3 [29-47].

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



Meta-analysis

Primary Outcome

Quality of Life

The fixed-effects analysis model was used to analyze the physical and mental outcomes of quality of life. In terms of physical outcomes, the scores of the control group were lower than those of the intervention group (MD 0.15, 95% CI 0.03-0.27, I^2=0%, P=.02) (Figure 2). However, there was no significant difference in the mental outcomes between the intervention and control groups (MD 0.10, 95% CI –0.03 to 0.24, I^2=46.4%, P=.14) (Figure 3).

Figure 2. Forest plot of the effect of physical outcomes of quality of life after cardiac surgery. SMD: standardized mean difference.

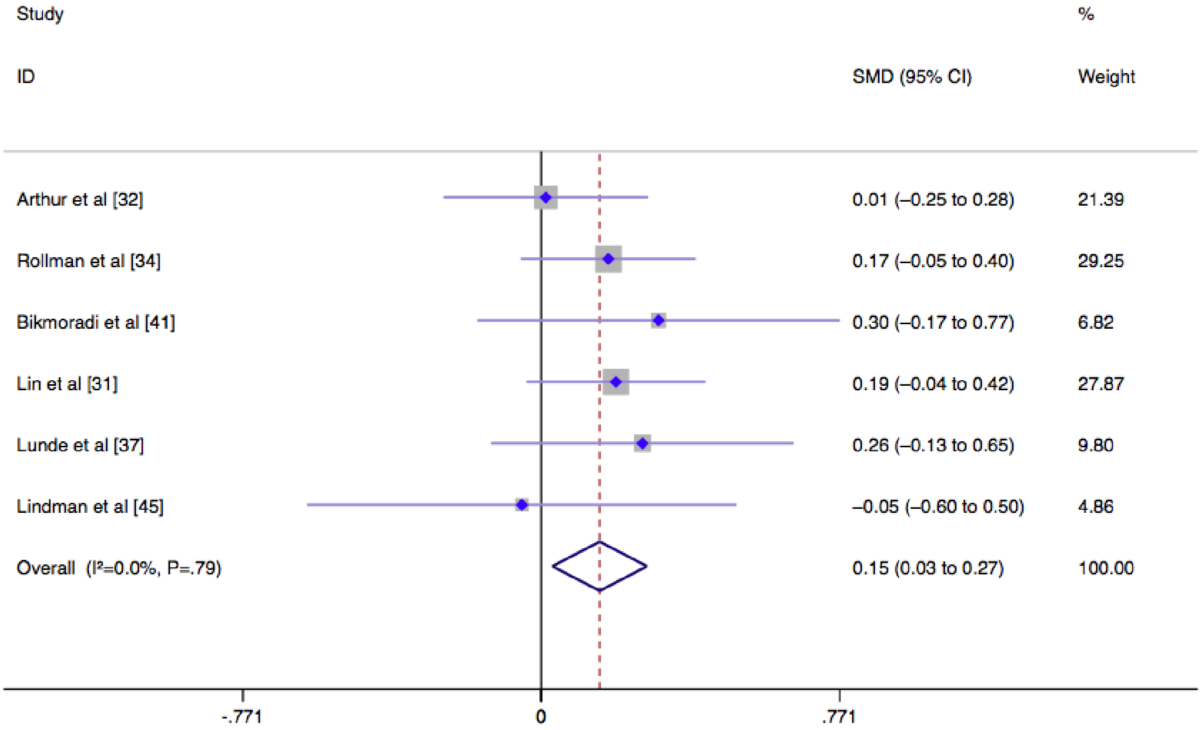
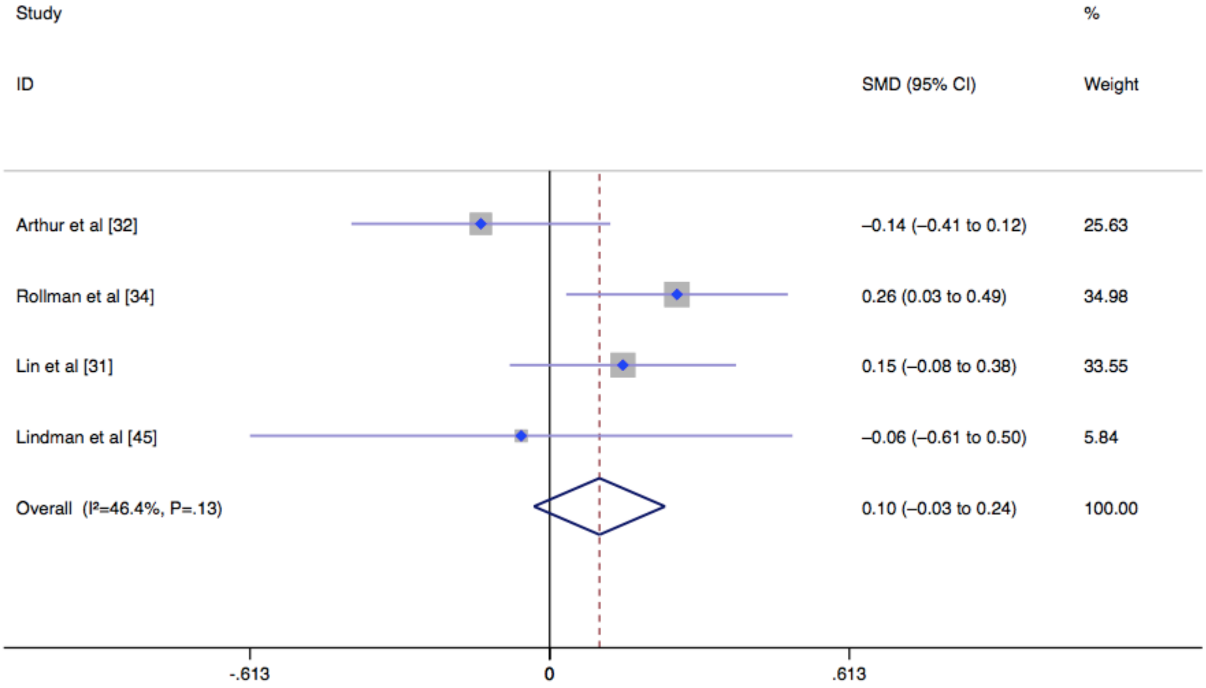


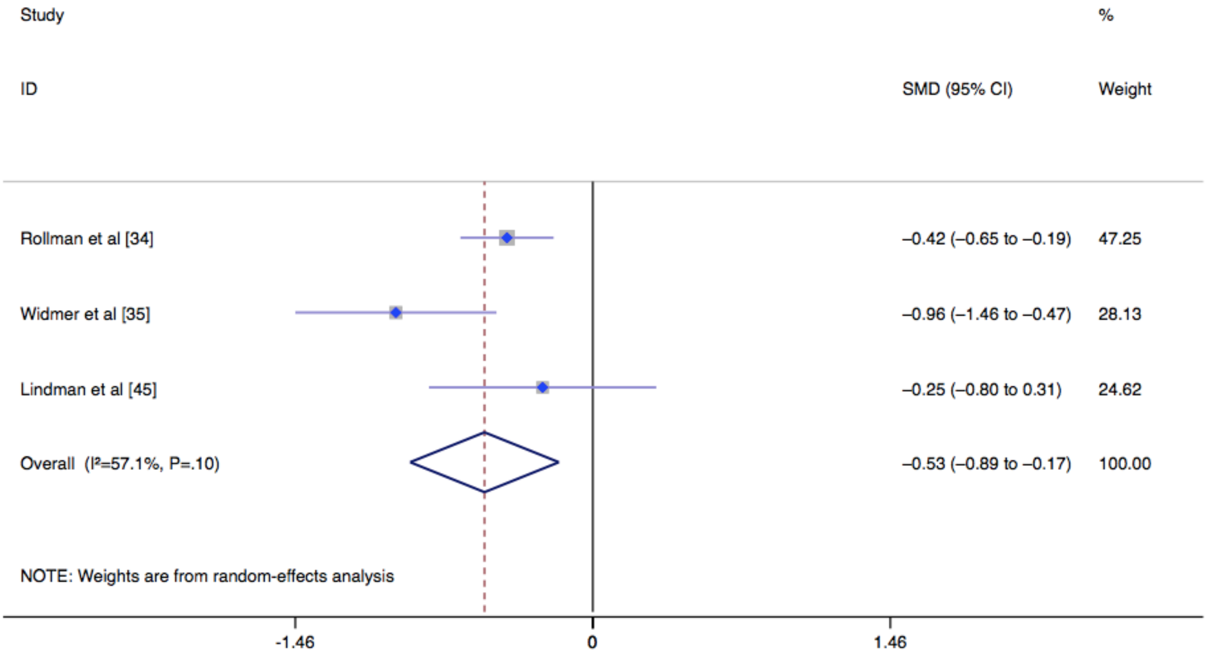
Figure 3. Forest plot of the effect of mental outcomes of quality of life after cardiac surgery. SMD: standardized mean difference.



Depression

To evaluate depression outcomes, we used a random-effects analysis model. The score of the control group was lower than that of the intervention group (MD –0.53, 95% CI –0.89 to –0.17, I^2=57.1%, *P*=.004) (Figure 4). Owing to sparse data, there was no subgroup analysis of the main outcome indicators based on a follow-up period of 3 months.

Figure 4. Forest plot of the effect of depression after cardiac surgery. SMD: standardized mean difference.



Other Outcomes

Three studies [29,30,42] reported no significant difference in the improvement of anxiety between the intervention and control groups. The occurrence of mortality was reported in 4 studies [31,32,36,39], of which 1 study [31] showed a statistically significant reduction in mortality in the intervention group, while the other 3 studies reported different conclusions. A total of 4 studies [30,35,39,42] reported no statistical significance for readmission between the intervention and control groups. Among the 3 studies [31,39,43] that reported on compliance, 2 studies [31,43] showed better compliance in the experimental group compared to the control group. However, 1 study [39] showed no statistical difference in compliance between the two groups. Two studies [35,37] indicated that none of the 4 lipid indexes had statistical significance. We generated 2 forest plots to show the effects of eHealth on other outcomes (Multimedia Appendices 4 and 5). Most of the studies had no significant differences in their results, apart from those for compliance, bleeding events, secondary prophylactic medication, patient satisfaction, and time in the therapeutic range (Multimedia Appendices 4 and 5).

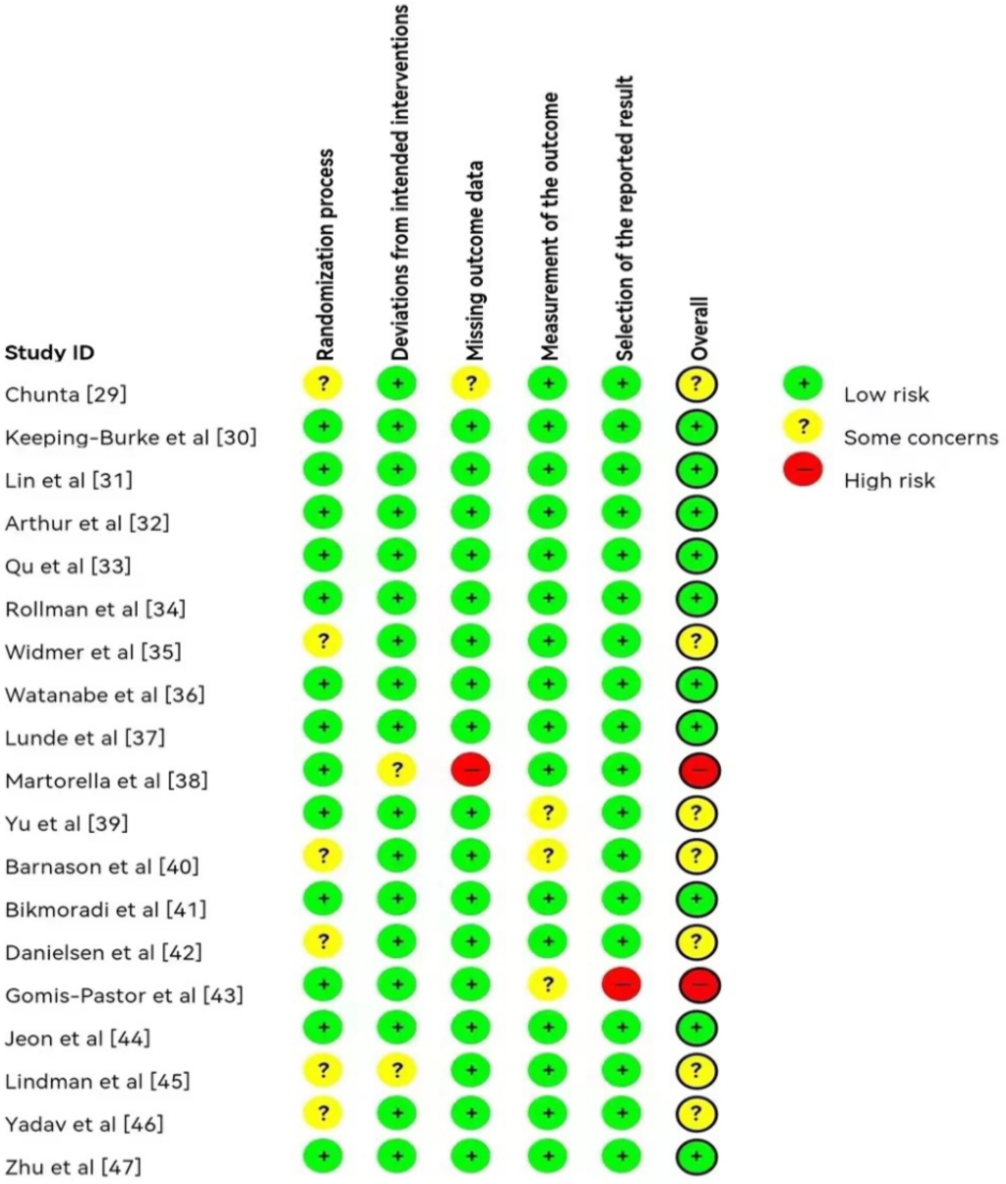
Sensitivity Analysis

A sensitivity analysis of quality of life and depressive outcomes was performed using the leave-one-out method, as shown in Multimedia Appendix 6 [31,32,34,35,37,41,45], and the results were consistent.

Quality Assessment

RoB 2 [27] was used for quality evaluation. Overall, the included studies had a low to moderate risk of bias, as shown in Figure 5. Most articles did not clearly report the randomization process. The overall quality of evidence for each outcome was assessed using the GRADE approach [28]. The quality rates of each outcome are shown in Multimedia Appendix 7. In summary, although the quality of some outcomes was moderate, the overall quality was low.

Figure 5. A risk-of-bias map using the Cochrane systematic evaluation method to assess the quality of the included randomized controlled trials.



Study ID	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall
Chunta [29]	?	+	?	+	+	?
Keeping-Burke et al [30]	+	+	+	+	+	+
Lin et al [31]	+	+	+	+	+	+
Arthur et al [32]	+	+	+	+	+	+
Qu et al [33]	+	+	+	+	+	+
Rollman et al [34]	+	+	+	+	+	+
Widmer et al [35]	?	+	+	+	+	?
Watanabe et al [36]	+	+	+	+	+	+
Lunde et al [37]	+	+	+	+	+	+
Martorella et al [38]	+	?	−	+	+	−
Yu et al [39]	+	+	+	?	+	?
Barnason et al [40]	?	+	+	?	+	?
Bikmoradi et al [41]	+	+	+	+	+	+
Danielsen et al [42]	?	+	+	+	+	?
Gomis-Pastor et al [43]	+	+	+	?	−	−
Jeon et al [44]	+	+	+	+	+	+
Lindman et al [45]	?	?	+	+	+	?
Yadav et al [46]	?	+	+	+	+	?
Zhu et al [47]	+	+	+	+	+	+

+ Low risk
? Some concerns
− High risk

Discussion

Principal Findings

In this systematic review, we assessed the impact of eHealth interventions on cardiac surgery recovery based on the results of 19 studies. These studies reported at least an equal (n=6) [29,30,33,38,39,42] or positive effect (n=13) [31,32,34-37,40,41,43-47] of the eHealth intervention compared to conventional care. According to the results of the meta-analysis, when compared with the control group, the eHealth intervention group showed an improvement in both the physical component of quality of life and the depressive status of patients after cardiac surgery. The mental component of quality of life was not significantly different in the two groups. This may be related to the shorter follow-up period of the included studies. Lin et al [31] showed that an effect on quality of life was not observed until the follow-up after 18 months, which was not long enough for most of the studies we included. van der Meij et al [48] showed that eHealth interventions have similar effects on different types of postoperative outcome measurements. The results for physical and psychological indicators were comparable. Therefore, we only conducted a meta-analysis on quality of life and depression. Study results on patients with different types of heart surgery, eHealth interventions, and measured outcomes varied widely. Due to the lack of reported data and heterogeneity, analysis of the other results using statistical methods was not performed. The economics of eHealth interventions have also not been studied yet. In terms of medication adherence, 2 of the 3 studies reported improved medication adherence after eHealth intervention [31,43]. Fewer than 10 studies were included in the quantitative analysis for each outcome; therefore, publication bias analysis was not performed. However, the findings should be interpreted with caution, as the overall quality of the body of evidence was low to moderate because of the risk of bias in the included studies.

Patients undergoing different cardiac surgeries have different postoperative needs. The studies in this review included eHealth interventions for medication education, consultation, follow-up, postoperative exercise, and rehabilitation. eHealth interventions were also used specifically for postoperative pain [38], anticoagulant management [44,47], and secondary drug prevention [33,35,39]. Martorella et al [38] revealed that patients in the experimental group did not experience less intense pain but reported significantly less pain interference when breathing or coughing (*P*=.04). However, the experimental group consumed more opioid medication (mean 31.2, SD 23.2 mg) than the control group (mean 18.8, SD 15.3 mg; *P*=.001). Two studies [44,47] showed that the use of eHealth improved efficacy in maintaining the therapeutic range of prothrombin time. Another study [44] showed improvement in self-management knowledge, self-efficacy, and improved behavior of patients undergoing cardiac valve replacement, as well as reduced adverse events for bleeding thrombosis [47] through eHealth intervention. Qu et al [33] showed that eHealth interventions have limited ability to increase prescription rates for statins or other drugs. Widmer et al [35] showed that eHealth interventions can improve the secondary prevention of cardiovascular diseases. Yu et al [39] showed that the intervention group had no significant impact on mortality, major adverse cerebrovascular events, and cardiovascular rehospitalization, which may be related to low patient participation.

Limitations

First, due to sparse data, there was no subgroup analysis of the main outcomes according to follow-up time, nor was there a comparative analysis of the pros and cons of different types of electronic interventions and different types of cardiac surgery on postoperative effects. Moreover, due to the limited number of included studies and the lack of publication bias analysis, the number of measured depression outcomes was small, and there was a possibility of deviation. Finally, allocation hiding was not explicitly reported in most of the included studies. The quality of the study outcomes was relatively low, and more high-quality randomized controlled trials should be included in the future.

Comparison With Prior Work

According to our literature review, there have been many studies on the clinical application of eHealth interventions. However, to our knowledge, there is no systematic study on the impact of electronic interventions on patients after cardiac surgery. This is the first published systematic evaluation of the effects of using eHealth on patients who have undergone cardiac surgery. We ensured the use of robust methodology to conduct this review by following the PRISMA guidelines [26].

Among the published systematic evaluations, studies on the application of electronic interventions included patients with cancer, respiratory diseases, and arthritis. In terms of quality of life, 3 studies [49-51] reported that electronic interventions were ineffective, but 7 [52-58] reported improvement in quality of life. Two articles [51,57] reported that electronic intervention was ineffective in relieving anxiety, and another [59] showed mixed views. Electronic intervention was reported to be ineffective for depression in 2 studies [51,60], whereas 3 articles [54,55,59] provided mixed conclusions. There were mixed results regarding the effect of pain relief, with some studies [50,51] indicating no effect on pain relief, and others [55,61] reporting the opposite. Seven studies reported on patient compliance: 2 studies [51,62] showed no statistical significance in improving patient compliance, 3 studies [63-65] showed a positive impact, and according to the remaining 2 studies [66,67], the impact was uncertain. Another study [68] indicated that electronic intervention could effectively improve maximum aerobic capacity and alter cardiovascular risk factors. In addition, we found that the effectiveness of electronic interventions may be related to the disease type. eHealth interventions showed positive effects on the outcome of some patients [52-55], but studies reporting on patients with cancer [49] and patients with arthritis [50,51] reported negative results.

There are many different types of cardiac surgery, such as valve replacement, bypass, and heart transplantation. The various results relating to different disease types might indicate that the effect of eHealth intervention may vary according to patient type. More high-quality studies are needed to verify these findings.

Conclusions

Based on this systematic review, the eHealth intervention group showed improvement in both the physical component of quality of life and depressive status after cardiac surgery, but the positive effects of the intervention were small. Moreover, the mental component of quality of life was not significantly different between the two groups. The overall quality of the evidence was low to medium. The compliance outcomes improved in most intervention groups. In the future, higher-quality randomized controlled studies of eHealth interventions are needed to provide more evidence for clinical practice.

Acknowledgments

We acknowledge the grants received from the Fujian Undergraduate Education and Teaching Reform Research Major Project (grant FBJG20210119) and the Education and Teaching Reform Research Project of Fujian Medical University (grant J21007).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.

[DOCX File , 27 KB - jmir_v24i8e40090_app1.docx]

Multimedia Appendix 2
Search terms.
[DOCX File , 27 KB - jmir_v24i8e40090_app2.docx]

Multimedia Appendix 3
Baseline characteristics of the 19 studies selected for the meta-analysis.
[DOCX File , 28 KB - jmir_v24i8e40090_app3.docx]

Multimedia Appendix 4
Forest plot of the effect of eHealth on other cardiac postoperative outcomes (continuous data).
[DOCX File , 126 KB - jmir_v24i8e40090_app4.docx]

Multimedia Appendix 5
Forest plot of the effect of eHealth on other cardiac postoperative outcomes (dichotomous data).
[DOCX File , 116 KB - jmir_v24i8e40090_app5.docx]

Multimedia Appendix 6
Influence analysis.
[DOCX File , 23 KB - jmir_v24i8e40090_app6.docx]

Multimedia Appendix 7
The overall quality of the evidence for each outcome.
[DOCX File , 23 KB - jmir_v24i8e40090_app7.docx]

References

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics—2016 update. Circulation 2016 Jan 26;133(4):e38-e360. [doi: 10.1161/cir.0000000000000350]
2. Poole L, Ronaldson A, Kidd T, Leigh E, Jahangiri M, Steptoe A. Pre-surgical depression and anxiety and recovery following coronary artery bypass graft surgery. J Behav Med 2017 Apr;40(2):249-258 [FREE Full text] [doi: 10.1007/s10865-016-9775-1] [Medline: 27552993]
3. Phadke SSD, Parkar H, Yardi S. Effect of music intervention on immediate post operative coronary artery bypass graft surgery (CABG) patients. Ind Jour of Physioth and Occupat Therapy - An Inter Jour 2014 Dec;8(4):106 [FREE Full text] [doi: 10.5958/0973-5674.2014.00021.5] [Medline: 26835471]
4. Rosengren A, Hawken S, Ôunpuu S, Sliwa K, Zubaid M, Almahmeed WA, et al. Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52 countries (the INTERHEART study): case-control study. The Lancet 2004 Sep;364(9438):953-962. [doi: 10.1016/S0140-6736(04)17019-0] [Medline: 15364186]
5. Richardson S, Shaffer JA, Falzon L, Krupka D, Davidson KW, Edmondson D. Meta-analysis of perceived stress and its association with incident coronary heart disease. Am J Cardiol 2012 Dec 15;110(12):1711-1716 [FREE Full text] [doi: 10.1016/j.amjcard.2012.08.004] [Medline: 22975465]
6. Custodis F, Gertz K, Balkaya M, Prinz V, Mathar I, Stamm C, et al. Heart rate contributes to the vascular effects of chronic mental stress: effects on endothelial function and ischemic brain injury in mice. Stroke 2011 Jun;42(6):1742-1749. [doi: 10.1161/STROKEAHA.110.598607] [Medline: 21527760]
7. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, American Heart Association Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation 2012 Jan 03;125(1):e2-e220 [FREE Full text] [doi: 10.1161/CIR.0b013e31823ac046] [Medline: 22179539]
8. Bikmoradi A, Masmouei B, Ghomeisi M, Roshanaei G. Impact of tele-nursing on adherence to treatment plan in discharged patients after coronary artery bypass graft surgery: A quasi-experimental study in Iran. Int J Med Inform 2016 Feb;86:43-48. [doi: 10.1016/j.ijmedinf.2015.12.001] [Medline: 26725694]
9. Bikmoradi A, Zafari A, Oshvandi K, Mazdeh M, Roshanaei G. Effect of progressive muscle relaxation on severity of pain in patients with multiple sclerosis: a randomized controlled trial. Hayat 2014;20(1):26-37 [FREE Full text]
10. Bikmoradi A, Seifi Z, Poorolajal J, Araghchian M, Safiaryan R, Oshvandi K. Effect of inhalation aromatherapy with lavender essential oil on stress and vital signs in patients undergoing coronary artery bypass surgery: A single-blinded randomized clinical trial. Complement Ther Med 2015 Jun;23(3):331-338. [doi: 10.1016/j.ctim.2014.12.001] [Medline: 26051567]

11. Seifi Z, Beikmoradi A, Oshvandi K, Poorolajal J, Araghchian M, Safiaryan R. The effect of lavender essential oil on anxiety level in patients undergoing coronary artery bypass graft surgery: A double-blinded randomized clinical trial. Iran J Nurs Midwifery Res 2014 Nov;19(6):574-580 [FREE Full text] [Medline: 25558253]
12. Khodaveisi M, Pazargadi M, Yaghmaei F, Bikmoradi A. Identifying challenges for effective evaluation in nursing education: A qualitative study. J Res Med Sci 2012 Jul;17(7):710-717 [FREE Full text] [Medline: 23798935]
13. Digital health. US Food and Drug Administration. 2018. URL: http://www.fda.gov/medicaldevices/digitalhealth/ [accessed 2020-04-09]
14. van Dyk L. A review of telehealth service implementation frameworks. Int J Environ Res Public Health 2014 Jan 23;11(2):1279-1298 [FREE Full text] [doi: 10.3390/ijerph110201279] [Medline: 24464237]
15. Vallespin B, Cornet J, Kotzeva A. Ensuring evidence-based safe and effective mHealth applications. Stud Health Technol Inform 2016;222:248-261. [Medline: 27198107]
16. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. JCO 2016 Feb 20;34(6):557-565. [doi: 10.1200/jco.2015.63.0830]
17. Park LG, Howie-Esquivel J, Dracup K. A quantitative systematic review of the efficacy of mobile phone interventions to improve medication adherence. J Adv Nurs 2014 Sep 01;70(9):1932-1953. [doi: 10.1111/jan.12400] [Medline: 24689978]
18. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015 Jan;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
19. Raikhelkar J, Raikhelkar JK. The impact of telemedicine in cardiac critical care. Crit Care Clin 2015 Apr;31(2):305-317. [doi: 10.1016/j.ccc.2014.12.008] [Medline: 25814456]
20. Cook DJ, Manning DM, Holland DE, Prinsen SK, Rudzik SD, Roger VL, et al. Patient engagement and reported outcomes in surgical recovery: effectiveness of an e-health platform. J Am Coll Surg 2013 Oct;217(4):648-655. [doi: 10.1016/j.jamcollsurg.2013.05.003] [Medline: 23891066]
21. Di Giammarco GG, Di Mascio T, Di Mauro M, Tarquinio A, Tarquinio A, Vittorini P. SmartHeart CABG Edu: first prototype and preliminary evaluation. In: Methodologies and Intelligent Systems for Technology Enhanced Learning. Cham: Springer; 2013:57-66.
22. Dale JG, Midthus E, Dale B. Using information and communication technology in the recovery after a coronary artery bypass graft surgery: patients' attitudes. J Multidiscip Healthc 2018 Aug 27;11(5):417-423 [FREE Full text] [doi: 10.2147/JMDH.S175195] [Medline: 30214223]
23. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Effect of comprehensive cardiac telerehabilitation on one-year cardiovascular rehospitalization rate, medical costs and quality of life: A cost-effectiveness analysis. Eur J Prev Cardiol 2016 May 19;23(7):674-682. [doi: 10.1177/2047487315602257] [Medline: 26289723]
24. Chauvet-Gelinier J, Bonin B. Stress, anxiety and depression in heart disease patients: A major challenge for cardiac rehabilitation. Ann Phys Rehabil Med 2017 Jan;60(1):6-12 [FREE Full text] [doi: 10.1016/j.rehab.2016.09.002] [Medline: 27771272]
25. Mohammad-Alizadeh-Charandabi S, Malakoti J, Sohrabi F, Shokranian N. The effect of telephone support on postpartum depression: a randomized controlled trial. J Caring Sci 2013 Jun;2(2):147-155 [FREE Full text] [doi: 10.5681/jcs.2013.018] [Medline: 25276721]
26. 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]
27. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Schünemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. The GRADE Working Group. 2013 Oct. URL: https://gdt.gradepro.org/app/handbook/handbook.html [accessed 2019-10-24]
29. Chunta K. An interventional study to provide telephone follow-up support to open-heart surgery patients during recovery. Appl Nurs Res 2016 Nov;32:41-43. [doi: 10.1016/j.apnr.2016.04.001] [Medline: 27969049]
30. Keeping-Burke L, Purden M, Frasure-Smith N, Cossette S, McCarthy F, Amsel R. Bridging the transition from hospital to home: effects of the VITAL telehealth program on recovery for CABG surgery patients and their caregivers. Res Nurs Health 2013 Dec;36(6):540-553. [doi: 10.1002/nur.21571] [Medline: 24242195]
31. Lin C, Yaseri M, Pakpour AH, Malm D, Broström A, Fridlund B, et al. Can a multifaceted intervention including motivational interviewing improve medication adherence, quality of life, and mortality rates in older patients undergoing coronary artery bypass surgery? A multicenter, randomized controlled trial with 18-month follow-up. Drugs Aging 2017 Feb;34(2):143-156. [doi: 10.1007/s40266-016-0429-3] [Medline: 28004259]
32. Arthur HM, Daniels C, McKelvie R, Hirsh J, Rush B. Effect of a preoperative intervention on preoperative and postoperative outcomes in low-risk patients awaiting elective coronary artery bypass graft surgery. A randomized, controlled trial. Ann Intern Med 2000 Aug 15;133(4):253-262. [doi: 10.7326/0003-4819-133-4-200008150-00007] [Medline: 10929164]

33. Qu J, Du J, Rao C, Chen S, Gu D, Li J, MISSION-1 Collaborative Group. Effect of a smartphone-based intervention on secondary prevention medication prescriptions after coronary artery bypass graft surgery: The MISSION-1 randomized controlled trial. Am Heart J 2021 Jul;237:79-89. [doi: 10.1016/j.ahj.2021.03.005] [Medline: 33689732]
34. Rollman BL, Belnap BH, LeMenager MS, Mazumdar S, Houck PR, Counihan PJ, et al. Telephone-delivered collaborative care for treating post-CABG depression: a randomized controlled trial. JAMA 2009 Nov 18;302(19):2095-2103 [FREE Full text] [doi: 10.1001/jama.2009.1670] [Medline: 19918088]
35. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: A randomized controlled trial. Am Heart J 2017 Jun;188:65-72. [doi: 10.1016/j.ahj.2017.02.016] [Medline: 28577682]
36. Watanabe E, Yamazaki F, Goto T, Asai T, Yamamoto T, Hirooka K, et al. Remote management of pacemaker patients with biennial in-clinic evaluation: continuous home monitoring in the Japanese at-home study: a randomized clinical trial. Circ Arrhythm Electrophysiol 2020 May;13(5):e007734 [FREE Full text] [doi: 10.1161/CIRCEP.119.007734] [Medline: 32342703]
37. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: A randomized controlled trial. Eur J Prev Cardiol 2020 Nov;27(16):1782-1792 [FREE Full text] [doi: 10.1177/2047487320905717] [Medline: 32106713]
38. Martorella G, Côté J, Racine M, Choinière M. Web-based nursing intervention for self-management of pain after cardiac surgery: pilot randomized controlled trial. J Med Internet Res 2012;14(6):e177 [FREE Full text] [doi: 10.2196/jmir.2070] [Medline: 23241361]
39. Yu C, Liu C, Du J, Liu H, Zhang H, Zhao Y, MISSION-2 Collaborative Group. Smartphone-based application to improve medication adherence in patients after surgical coronary revascularization. Am Heart J 2020 Oct;228:17-26. [doi: 10.1016/j.ahj.2020.06.019] [Medline: 32745732]
40. 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]
41. Bikmoradi A, Masmouei B, Ghomeisi M, Roshanaei G, Masiello I. Impact of telephone counseling on the quality of life of patients discharged after coronary artery bypass grafts. Patient Educ Couns 2017 Dec;100(12):2290-2296. [doi: 10.1016/j.pec.2017.06.010] [Medline: 28625860]
42. Danielsen SO, Moons P, Sandvik L, Leegaard M, Solheim S, Tønnessen T, et al. Impact of telephone follow-up and 24/7 hotline on 30-day readmission rates following aortic valve replacement - A randomized controlled trial. Int J Cardiol 2020 Feb 01;300:66-72. [doi: 10.1016/j.ijcard.2019.07.087] [Medline: 31387822]
43. Gomis-Pastor M, Mirabet Perez S, Roig Minguell E, Brossa Loidi V, Lopez Lopez L, Ros Abarca S, et al. Mobile health to improve adherence and patient experience in heart transplantation recipients: the mHeart trial. Healthcare (Basel) 2021 Apr 14;9(4):463 [FREE Full text] [doi: 10.3390/healthcare9040463] [Medline: 33919899]
44. Jeon HR, Park JS. [Development and application of a self-management program based on prothrombin INR monitoring for patients with cardiac valve replacement]. J Korean Acad Nurs 2015 Aug;45(4):554-564. [doi: 10.4040/jkan.2015.45.4.554] [Medline: 26364530]
45. Lindman BR, Gillam LD, Coylewright M, Welt FGP, Elmariah S, Smith SA, et al. Effect of a pragmatic home-based mobile health exercise intervention after transcatheter aortic valve replacement: a randomized pilot trial. Eur Heart J Digit Health 2021 Mar;2(1):90-103 [FREE Full text] [doi: 10.1093/ehjdh/ztab007] [Medline: 34048509]
46. Yadav S, Sethi R, Pradhan A, Vishwakarma P, Bhandari M, Gattani R, et al. 'Routine' versus 'Smart Phone Application Based - Intense' follow up of patients with acute coronary syndrome undergoing percutaneous coronary intervention: Impact on clinical outcomes and patient satisfaction. Int J Cardiol Heart Vasc 2021 Aug;35:100832 [FREE Full text] [doi: 10.1016/j.ijcha.2021.100832] [Medline: 34235246]
47. Zhu Z, Li C, Shen J, Wu K, Li Y, Liu K, et al. New internet-based warfarin anticoagulation management approach after mechanical heart valve replacement: prospective, multicenter, randomized controlled trial. J Med Internet Res 2021 Aug 13;23(8):e29529 [FREE Full text] [doi: 10.2196/29529] [Medline: 34397393]
48. van der Meij E, Anema JR, Otten RHJ, Huirne JAF, Schaafsma FG. The effect of perioperative e-health interventions on the postoperative course: a systematic review of randomised and non-randomised controlled trials. PLoS One 2016;11(7):e0158612 [FREE Full text] [doi: 10.1371/journal.pone.0158612] [Medline: 27383239]
49. Xu A, Wang Y, Wu X. Effectiveness of e-health based self-management to improve cancer-related fatigue, self-efficacy and quality of life in cancer patients: Systematic review and meta-analysis. J Adv Nurs 2019 Dec;75(12):3434-3447. [doi: 10.1111/jan.14197] [Medline: 31566769]
50. Zhou L, Zhou Y, Yu P, Meng F, Xu Y, Jiang Y. Effects of e-health interventions on health outcomes in patients with rheumatoid arthritis: a systematic review and meta-analysis. J Clin Nurs 2022 Jan 31. [doi: 10.1111/jocn.16236] [Medline: 35102628]
51. Butler S, Sculley D, Santos D, Fellas A, Gironès X, Singh-Grewal D, et al. Effectiveness of eHealth and mHealth interventions supporting children and young people living with juvenile idiopathic arthritis: systematic review and meta-analysis. J Med Internet Res 2022 Feb 02;24(2):e30457 [FREE Full text] [doi: 10.2196/30457] [Medline: 35107431]

52. Li J, Liu Y, Jiang J, Peng X, Hu X. Effect of telehealth interventions on quality of life in cancer survivors: A systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2021 Oct;122:103970. [doi: 10.1016/j.ijnurstu.2021.103970] [Medline: 34303269]
53. Larson JL, Rosen AB, Wilson FA. The effect of telehealth interventions on quality of life of cancer survivors: A systematic review and meta-analysis. Health Informatics J 2020 Jun;26(2):1060-1078 [FREE Full text] [doi: 10.1177/1460458219863604] [Medline: 31566455]
54. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: A systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
55. Chen Y, Guan B, Li Z, Li X. Effect of telehealth intervention on breast cancer patients' quality of life and psychological outcomes: A meta-analysis. J Telemed Telecare 2018 Apr;24(3):157-167. [doi: 10.1177/1357633X16686777] [Medline: 28081664]
56. Snoswell CL, Rahja M, Lalor AF. A systematic review and meta-analysis of change in health-related quality of life for interactive telehealth interventions for patients with asthma. Value Health 2021 Feb;24(2):291-302 [FREE Full text] [doi: 10.1016/j.jval.2020.09.006] [Medline: 33518036]
57. Su JJ, Yu DSF, Paguio JT. Effect of eHealth cardiac rehabilitation on health outcomes of coronary heart disease patients: A systematic review and meta-analysis. J Adv Nurs 2020 Mar;76(3):754-772. [doi: 10.1111/jan.14272] [Medline: 31769527]
58. Esfandiari E, Miller WC, Ashe MC. The effect of telehealth interventions on function and quality of life for older adults with pre-frailty or frailty: a systematic review and meta-analysis. J Appl Gerontol 2021 Nov;40(11):1649-1658. [doi: 10.1177/0733464820983630] [Medline: 33402043]
59. Zhao L, Chen J, Lan L, Deng N, Liao Y, Yue L, et al. Effectiveness of telehealth interventions for women with postpartum depression: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Oct 07;9(10):e32544 [FREE Full text] [doi: 10.2196/32544] [Medline: 34617909]
60. Thabrew H, Stasiak K, Hetrick SE, Wong S, Huss JH, Merry SN. E-Health interventions for anxiety and depression in children and adolescents with long-term physical conditions. Cochrane Database Syst Rev 2018 Aug 15;8:CD012489 [FREE Full text] [doi: 10.1002/14651858.CD012489.pub2] [Medline: 30110718]
61. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: A meta analysis. Int J Nurs Stud 2020 Jun;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]
62. Bingham JM, Black M, Anderson EJ, Li Y, Toselli N, Fox S, et al. Impact of telehealth interventions on medication adherence for patients with type 2 diabetes, hypertension, and/or dyslipidemia: a systematic review. Ann Pharmacother 2021 May;55(5):637-649. [doi: 10.1177/1060028020950726] [Medline: 32815400]
63. Jeminiwa R, Hohmann L, Qian J, Garza K, Hansen R, Fox BI. Impact of eHealth on medication adherence among patients with asthma: A systematic review and meta-analysis. Respir Med 2019 Mar;149:59-68. [doi: 10.1016/j.rmed.2019.02.011] [Medline: 30803887]
64. Kebapci A, Ozkaynak M, Lareau SC. Effects of eHealth-based interventions on adherence to components of cardiac rehabilitation: a systematic review. J Cardiovasc Nurs 2020;35(1):74-85. [doi: 10.1097/JCN.0000000000000619] [Medline: 31738217]
65. Pouls BPH, Vriezekolk JE, Bekker CL, Linn AJ, van Onzenoort HAW, Vervloet M, et al. Effect of interactive eHealth interventions on improving medication adherence in adults with long-term medication: systematic review. J Med Internet Res 2021 Jan 08;23(1):e18901 [FREE Full text] [doi: 10.2196/18901] [Medline: 33416501]
66. Aardoom JJ, Loheide-Niesmann L, Ossebaard HC, Riper H. Effectiveness of eHealth interventions in improving treatment adherence for adults with obstructive sleep apnea: meta-analytic review. J Med Internet Res 2020 Feb 18;22(2):e16972 [FREE Full text] [doi: 10.2196/16972] [Medline: 32130137]
67. Schulte MHJ, Aardoom JJ, Loheide-Niesmann L, Verstraete LLL, Ossebaard HC, Riper H. Effectiveness of eHealth interventions in improving medication adherence for patients with chronic obstructive pulmonary disease or asthma: systematic review. J Med Internet Res 2021 Jul 27;23(7):e29475 [FREE Full text] [doi: 10.2196/29475] [Medline: 34313593]
68. Rawstorn JC, Gant N, Direito A, Beckmann C, Maddison R. Telehealth exercise-based cardiac rehabilitation: a systematic review and meta-analysis. Heart 2016 Aug 01;102(15):1183-1192. [doi: 10.1136/heartjnl-2015-308966] [Medline: 26936337]

Abbreviations

GRADE: Grading of Recommendations, Assessment, Development and Evaluations
MD: mean difference
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
QoL: quality of life

Edited by T Leung, G Eysenbach; submitted 06.06.22; peer-reviewed by M Pritchard, X Xiong; comments to author 29.06.22; revised version received 12.07.22; accepted 26.07.22; published 16.08.22.

Please cite as:
Ni R, Liu M, Huang S, Yang J
Effects of eHealth Interventions on Quality of Life and Psychological Outcomes in Cardiac Surgery Patients: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(8):e40090
URL: https://www.jmir.org/2022/8/e40090
doi:10.2196/40090
PMID:35972792

Review

Digital Health Competencies Among Health Care Professionals: Systematic Review

Jessica Longhini[1], MSc; Giacomo Rossettini[2], DPhil; Alvisa Palese[1], DPhil

[1]Department of Medical Sciences, University of Udine, Udine, Italy
[2]School of Physiotherapy, University of Verona, Verona, Italy

Corresponding Author:
Alvisa Palese, DPhil
Department of Medical Sciences, University of Udine
Viale Ungheria, 20
Udine, 33100
Italy
Phone: 39 0432590926
Email: alvisa.palese@uniud.it

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/11/e43721/

Abstract

Background: Digitalization is not fully implemented in clinical practice, and several factors have been identified as possible barriers, including the competencies of health care professionals. However, no summary of the available evidence has been provided to date to depict digital health competencies that have been investigated among health care professionals, the tools used in assessing such competencies, and the effective interventions to improve them.

Objective: This review aims to summarize digital health competencies investigated to date and the tools used to assess them among health care professionals.

Methods: A systematic review based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist was performed. The MEDLINE, Cumulative Index to Nursing and Allied Health Literature, PsycINFO, and Scopus databases were accessed up to September 4, 2021. Studies assessing digital health competencies with quantitative designs, targeting health care professionals, and written in English were included. The methodological quality of included studies was evaluated using the Joanna Briggs Institute tools.

Results: A total of 26 studies, published from 1999 to 2021, met the inclusion criteria, and the majority were cross sectional in design, while only 2 were experimental study designs. Most studies were assessed with moderate to low methodological quality; 4 categories and 9 subcategories of investigated digital health competencies have been identified. The most investigated category was “Self-rated competencies,” followed by “Psychological and emotional aspects toward digital technologies,” “Use of digital technologies,” and “Knowledge about digital technologies.” In 35% (9/26) of the studies, a previously validated tool was used to measure the competencies assessed, while others developed ad hoc questionnaires.

Conclusions: Mainly descriptive studies with issues regarding methodology quality have been produced to date investigating 4 main categories of digital health competencies mostly with nonvalidated tools. Competencies investigated might be considered while designing curricula for undergraduate, postgraduate, and continuing education processes, whereas the methodological lacks detected might be addressed with future research. There is a need to expand research on psychological and emotional elements and the ability to use digital technology to self-learn and teach others.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021282775; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=282775

(J Med Internet Res 2022;24(8):e36414) doi:10.2196/36414

KEYWORDS

eHealth literacy; eHealth competencies; digital health; competencies; eHealth; health literacy; digital technology; health care professionals; health care workers; review; systematic review

Introduction

Background

Over the last few decades, the increasing technology development has led to a wide digitalization of several work processes in health care settings. The World Health Organization (WHO) has recently defined and categorized digital health interventions in the health care context as "a discrete function of the digital technology to achieve health care sector objectives" [1]. The framework developed by the WHO includes a wide range of digital tools and interventions, such as telemonitoring, the use of artificial intelligence, decision-making algorithms, and health data collection [1]. According to the evidence available, digitalization has improved the quality of care, affecting several outcomes at the system level (eg, safety in medication administration and length of in-hospital stay) and at the individual level (eg, increasing functional/cognitive abilities and patients' satisfaction) [2].

Despite its potential effectiveness, digitalization is not fully implemented in clinical practice. Several factors have been identified as possible barriers, including the availability of technology, financial resources, and health care professionals' skills in using digital technology [3]. To improve health care digitalization, health professionals have been recognized as a key factor in the digital transformation of the health care sector. Therefore, they should be equipped with digital health competencies, from basic (eg, computers, tablets) to more complex skills, such as teaching patients about the safe and appropriate use of digital data sources and technology [3].

Digital Health Competencies

Different terms have been established to date by the literature to refer to digital health competencies. The most common term is eHealth literacy, which has been defined as the ability to use information retrieved from an electronic source to solve a health problem [4]. Conceptual frameworks describing the concept and components of eHealth literacy have been developed to date for citizens and patients [5]. For example, Norman and Skinner's Lily framework [4] includes 6 literacy competencies, namely, health, traditional, information, scientific, computer, and media literacy. These competencies have been further expanded, with updated frameworks such as the "Patient Readiness to Engage in Health Internet Technology" (PRE-HIT) and the "eHealth Literacy Framework" (eHLF). These include different elements promoting or hindering eHealth literacy such as motivation, engagement, willingness, anxiety, expectations, and beliefs [6,7]. However, the concepts and components considered in these frameworks should be conceived differently when referring to health care professionals, given that they are expected to have the competencies required to solve patients' problems rather than a personal health problem [8]. As a result of this gap, and in light of the required competencies to overcome barriers in health care digitalization processes [3], an emergent area of investigation has been set around the digital health competencies of health care professionals.

Different frameworks have been developed also in this context, mostly targeting a specific profession, mainly nurses, and using the methodology of expert consultation, surveys, and consensus (eg, the Delphi study) [5]. Among the most recent frameworks, the Health Information Technology Competencies (HITCOMP) [9] framework and the Technology Informatics Guiding Education Reform (TIGER) version 2.0 framework [10] have both identified 33 areas of competence articulated in domains. Specifically, the HITCOMP framework [9] has provided 5 domains, namely, (1) administration, (2) research/biomedicine, (3) direct patient care, (4) informatics, and (5) engineering/information systems/information and communications technology (ICT).

The TIGER framework has described relevant competencies [10] for those who provide direct patient care, including communication, documentation, quality and safety management, teaching, training/education, and ethics in health information technology [10].

In this context, a recent review dared to summarize the digital health competencies expected by health care professionals by synthetizing 30 available frameworks [5]. According to the findings, discrepancies and overlapping are still present across available frameworks regarding the different categorization of the competencies, the methods used to conceptualize such frameworks, and the competencies included [5]. These inconsistencies rely on the different health care professions targeted, including health professionals not involved in direct care, such as engineers [10]. Moreover, half of the 30 frameworks [5] emerged from gray literature and 30% were developed with the involvement of students, thus with different expected responsibilities and competencies [5].

Furthermore, the development of the digital health competencies according to the emergence of new technologies requires a continuous updating of both competencies to consider relevant and methods to assess appropriately these competencies [5]. However, to the best of our knowledge, no recent systematic reviews have been performed on digital health competencies among health care professionals. Providing a systematic summary of literature might inform policymakers, managers, and educators about how to appropriately measure the level of competencies in health care sector and how to develop adequate training programs to fill the gap in the digital health competencies. Moreover, a summary of the available evidence may inform researchers about the gaps in this field of investigation. Therefore, this systematic review aims to summarize which digital health competencies have been studied in literature and with what tools they have been measured to date among health care professionals.

Methods

Research Questions

Two main research questions have been addressed: (1) Which digital health competencies have been investigated to date among health care professionals? (2) How have these competencies been assessed?

Study Design

We conducted a systematic review by adopting the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Multimedia Appendix 1) checklist [11] both in protocol development and in method and finding reporting. The protocol has been submitted for evaluation to the International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD42021282775).

Eligibility Criteria

Studies satisfying the following criteria were included: (1) assessing digital health competencies as an umbrella term (thus including terms related and similar to, eg, digital literacy [12], health informatics competencies [10], or eHealth competencies [13,14]); (2) targeting health care professionals; (3) adopting a quantitative design (eg, randomized control trial, quasi-experimental trial, longitudinal, cross-sectional studies); and (4) written in English. Therefore, qualitative studies, commentaries, editorials, letters, PhD dissertations, conference abstracts, and all studies that investigated technology accessibility were excluded.

Data Searching

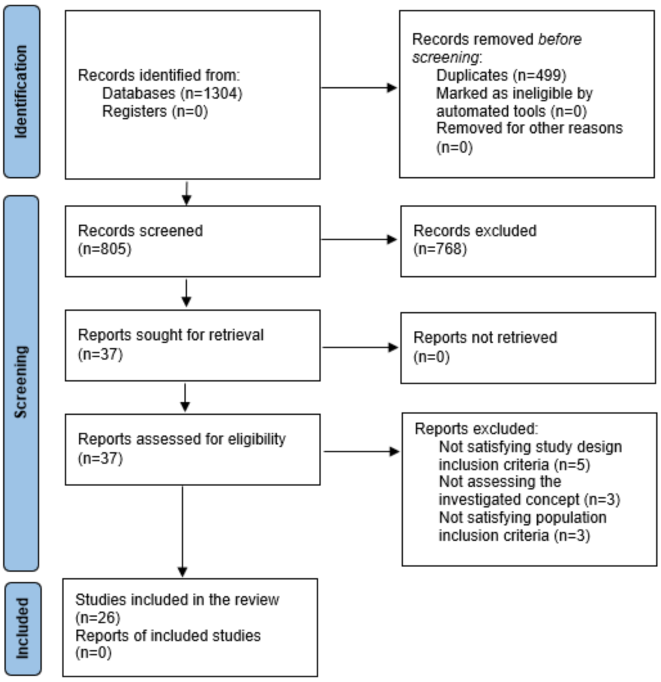
The search string was designed and developed with the support of an expert research librarian and then preliminarily piloted in a database to ensure its accuracy according to the review aims. The final string search included the following keywords: (1) “digital competencies” and “eHealth literacy” in their similar and affiliated terms (eg, digital Health Literacy,” “digital literac*,” “digital competenc*,” “digital skill*”; and (2) “health professionals” in its affiliated and similar term (“health care practice*,” “nurs*”) as fully reported in Multimedia Appendix 2. The search string was applied in the following databases: MEDLINE, Cumulative Index to Nursing and Allied Health Literature, PsycINFO, and Scopus up to September 4, 2021, with an English language restriction filter. In addition, the “TITLE-ABS-KEY” filter was adopted for the SCOPUS database to detect relevant studies. The reference lists of the included studies, the available trial registries, and the references of systematic reviews were screened by hand-searching to retrieve all relevant studies. Moreover, Mendeley Reference Manager was used to manage all references and delete duplicates.

Study Selection

The title, the abstract, and the full-text screening of eligible studies were performed by 2 researchers (JL and GR) independently, and disagreements were resolved by a third researcher (AP). Interrater reliability was assessed using Cohen κ statistics, and it resulted in a value of 0.83 (95% CI 0.73-0.93), meaning an almost perfect level of agreement [15].

The study selection process is summarized in Figure 1 according to the PRISMA flow diagram [11].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for new systematic reviews that included searches of databases and registers only (Page et al [11]).



Methodological Quality Assessment

Studies were assessed for their methodological quality by 2 researchers (JL and GR) independently, and a third researcher (AP) was consulted to resolve disagreements.

Joanna Briggs Institute tools for analytical cross-sectional [16], prevalence [17], and randomized control trial studies [18] were adopted according to the design used in the included studies. Specifically, regarding observational studies, we considered analytical cross-sectional studies when the statistical analysis was performed to identify associations between variables; otherwise they were considered prevalence studies [17].

For all quality assessment tools adopted, the scores applied were “Y” (yes) when the item was satisfied, “N” (no) when the item was not satisfied, and “U” (unclear) when the information contained in the study was not sufficient. Cut-off criteria were established through an agreement process among researchers based on previous evidence [19,20]. A moderate methodological quality level was identified when positive answers (= yes) were scored from 5 to 6 in analytical cross-sectional studies, from 6 to 7 in prevalence studies, and 10 or 11 in randomized control trials. Positive answers below and above these values were considered low and high methodological quality, respectively.

However, to comprehend all studies, their methodological quality was not considered an exclusion criterion.

Data Extraction, Analysis, and Synthesis

The following data were extracted from each included study: author(s); year of publication; country; study design; population characteristics (eg, age, work profile) and number of participants; investigated digital health competencies; definition(s) provided of the assessed competencies (as reported in the “Study Background” or in the “Methods” sections); tools; and data collection methods used to assess the competencies investigated.

Two Excel grids were developed to extract data from the included studies according to the study aims. The grids were piloted among 3 studies to ensure their feasibility, and consistency was also assessed among the researchers (JL and GR) who performed the data extraction.

After having extracted the data, first, the study characteristics were summarized according to the study design (analytical cross-sectional, prevalence, and randomized control trial studies), reporting their main features and methodological quality. Second, following the aims of this systematic review, digital health competencies were summarized by extracting and analyzing items as open- or closed-ended questions included in the tools used to assess such competencies in each study, irrespective of their formats [21]. The items that emerged were grouped into categories and, when needed, into subcategories through a content analysis [22]. In the content analysis, the researchers adopted a systematic coding and categorizing approach to textual information extracted from the studies to merge patterns, and structure them into main categories and subcategories, by also reporting the frequency [23]. Two researchers (JL and GR) independently performed the entire process, and disagreements were resolved by a third researcher (AP). From the analysis of 362 extracted items used to assess digital health competencies in the included studies, 4 main categories emerged, namely, “Self-rated competencies,” “Psychological and emotional aspects toward the use of digital technologies,” “Use of digital technologies,” and “Knowledge about digital technologies.” Then, the number of items used across studies and the number of studies that assessed each specific category of competence were counted. Furthermore, tools used to assess the competencies in included studies were summarized into their main features.

Results

Main Characteristics of Studies Identified

A total of 1304 studies were identified from literature searches, of which 26 met the inclusion criteria (Figure 1). The majority were cross-sectional studies, of which 11 [24-34] were considered prevalence data studies, and 13 [35-47] as analytical cross-sectional studies (Table 1). Among the remaining ones, 2 were experimental studies [48,49]. The studies included were conducted over a wide range of years, from 1999 [25] to 2021 [27], and more than 65% (17/26) of them [26-28,32,33,35,37,39,40,42,44] have been published in the last 5 years.

In total, 5 studies were conducted in the United States of America [25,30,38,40,41], while the others were performed in different European countries (eg, Germany [34,35] and Finland [25,26]), and in low-income countries (eg, Malawi [48] and Uganda [43]). In terms of the setting, 9 studies [32,33,36,37,43-46,49] were conducted in hospitals. By contrast, the others were performed in mixed settings (eg, acute care [26], local health departments [25], and community [48]).

A total of 8 studies [26,32,36,38,41,42,47,49] involved nurses and 7 [25,27,28,33,39,44,45] covered health care professionals, while the others involved specific roles (eg, psychiatrists [40], pharmacists [31], maternal and child professionals [30]; Tables 1 and 2). The sample size was variable across the studies, ranging from 36 [30] to 5209 participants [39] with a variable age mostly comprised between 30 [46] and 50 years [27].

Table 1. Characteristics of included analytical cross-sectional and prevalence studies.

Study type and reference	Country	Study design	Setting(s)	Sample and profession; age	Competencies assessed	Definition provided of the competencies assessed	Tools/data collection method(s) and items
Analytical cross-sectional studies							
Campbell and McDowell [38]	United States	Descriptive	Community hospital (100 beds)	112 registered nurses; 35 (31.2%) born in the 1960s	Self-perceived computer literacy	Computer literacy: "the skills necessary for accessing and using information, managing files, navigating an operating system, and using common applications, such as word processing" (source: "Background" section)	Gassert/McDowell Computer Literacy Survey (15 items)
Do et al [39]	Vietnam	Cross sectional	12 hospitals and 3 health centers	5209 HCPs[a], 905 (17.4%) aged between 41 and 60 years	eHealth literacy	N/A[b]	eHEALS[c] questionnaire [15] (8 items to measure consumers' combined knowledge, comfort, and perceived skills at finding, evaluating, and applying eHealth information to health problems)
Duffy et al [40]	United States	Cross sectional	Mixed settings	152 psychiatrists; 67 (44 %) aged between 50 and 64 (mean 56.9) years	• Comfort in using computers and other electronic devices for professional, personal, and clinical aims • Computer use for specific clinical tasks	N/A	Web-and-paper-based survey (open- and closed-ended questions)
Elhadi et al [37]	Libya	Cross sectional	Hospitals	673 specialists/senior physicians, physician trainees; 442 (65.7%) aged between 30 and 40 years	• Using computer ability • Awareness, knowledge, attitude, and computer skills about telemedicine	• Awareness: N/A • Knowledge: N/A • Attitude: N/A • Computer skills: level of "information technology and computer skills" (source: "Methods" section)	AKAS[d] questionnaire [27] (Awareness, 12 items; Knowledge, 11 items; Attitude, 11 items; information technology/computer skills, 13 items)
Gaumer et al [41]	United States	Cross sectional	Mixed settings	241 nurse practitioners; N/A	• Use of information technology (general and for specific function) • Benefits perceived from using technology (caregiving, time saving, patient safety) • Self-perceptions about information technology competence	N/A	• Questionnaire: Use of information technology (general: 1 item, specific functions: N/A) • Perceived benefit, 3 items • Self-perceptions about information technology competence, 1 item

Study type and reference	Country	Study design	Setting(s)	Sample and profession; age	Competencies assessed	Definition provided of the competencies assessed	Tools/data collection method(s) and items
Gürdaş Topkaya and Kaya [36]	Turkey	Cross sectional	Hospitals	688 nurses; 293 (42.6%) aged between 20 and 29 years	Computer literacy and attitudes toward computers in health care	Computer literacy: "briefly defined as the ability to use a computer" as well as "the ability to control [a] computer in achieving certain goals," "to use different computer applications," "to comprehend [the] economic, psychological and social effects of computer[s] on [the] individual and society," and "to use [a] computer [for] access to information, [for] communication and [in the] problem solution process" (source: "Background" section)	• Multicomponent Assessment of Computer Literacy, 24 items • Pre-test for Attitudes Towards Computers in Healthcare Assessment Scale version 2, 40 items
Hennemann et al [35]	Germany	Cross sectional	Rehabilitation facilities	149 participants (nurses, psychologists, physical therapists, physicians, patient administration, social workers, art/body/occupational therapists, nutritionists, medical technical assistants); mean 44.35 (SD 11.27) years	• Acceptance of eHealth intervention and of online aftercare • Information technology literacy • eHealth literacy • Performance expectancy • Effort expectancy • Internet anxiety • Knowledge of eHealth interventions	• Acceptance (operationalized according to the UTAUT[e]) "the intention to use eHealth interventions for patients' health promotion in work context, and adoption of online aftercare" • eHealth literacy: the ability to find, evaluate, and utilize internet-based health information to health problems" (source for both: "Methods" section) • N/A for others	• Self-administered web-based questionnaire (acceptance, 4 items; information technology literacy, 1 item; performance expectancy, 2 items; effort expectancy, 2 items; internet anxiety, 2 items; knowledge of eHealth interventions, 2 items) • eHEALS questionnaire [15] (8 items)
Kritsotakis et al [42]	Greece	Cross sectional	Secondary and primary general-care hospitals	200 nurses and nursing assistants; 70 (35%) aged 45-54 years	eHealth literacy	"The ability to find and assess health-related information online at the individual level" (source: "Methods" section)	eHEALS questionnaire [15] (8 items)

Study type and reference	Country	Study design	Setting(s)	Sample and profession; age	Competencies assessed	Definition provided of the competencies assessed	Tools/data collection method(s) and items
Olok et al [43]	Uganda	Cross sectional	Hospitals	68 doctors; 33 (48.5%) aged 31-40 years	• Attitudes toward eHealth • Level of ICT[f] use and skills	N/A	Questionnaire: internal consistence evaluated (level of ICT use and skills on the same 18 items—list of facilities and tools; attitudes, 25 items divided into relative advantages, compatibility, complexity, trialability—not considered, observability)
Shiferaw and Mehari [44]	Ethiopia	Cross sectional	Hospital	287 HCPs; mean 30.09 (SD 5.025) years	• Internet use (types and frequency) • eHealth literacy	• Internet use: “Health professionals’ practice of using the Internet for browsing health-related information to make sound decisions” • eHealth literacy: “participants’ ability to locate and use credible information from the Internet” (source: “Methods” section)	• Internet use, 15 items • eHEALS questionnaire [15] (8 items)
Tesfa et al [45]	Ethiopia	Cross sectional	Teaching hospitals	383 HCPs (nurses, doctors, midwives, pharmacists, laboratory technicians); mean 28.3 (SD 3.37) years	• Electronic health information resource utilization (information searching, technical skills) and purpose of use • Computer literacy • eHealth literacy • Awareness • Attitude • Motivational factors (perceived usefulness and use)	N/A	• Questionnaire (purpose of use, 5 items; N/A for others) • eHEALS questionnaire [15] (8 items)
Thapa et al [46]	Saudi Arabia	Cross sectional	Hospitals	218 physicians and nurses; 61 (28%) aged between 31 and 35 years	• Willingness to use digital health tools in patient care • Attitudes and self-efficacy toward using digital health tools • Digital health tools use perceived benefits and costs	• Willingness: N/A • Self-efficacy: “The belief in one’s own ability to successfully perform various specific actions related to the use of digital tools in patient care” • Attitude: “The perceived relevance/value of different functions of digital tools for active engagement of patients in their own treatment/care” • Perceived benefits: “Positive consequences of using digital tools” • Perceived costs: “Potential psychological, financial, technological and administrative burden” (source: “Methods” section)	Questionnaire (willingness, 1 item; self-efficacy, 12 items; attitude, 10 items; perceived benefit and costs, 20 items)

Study type and reference	Country	Study design	Setting(s)	Sample and profession; age	Competencies assessed	Definition provided of the competencies assessed	Tools/data collection method(s) and items
Vehko et al [47]	Finland	Cross sectional	Hospitals, primary care, private practice, social care, and others	3407 registered nurses; mean age 46.2 (SD 10.99) years	Nurses' informatics competence: classification competence; e-care competence; e-documentation competence; ethics competence	• Classification competence: "Planning, implementation and evaluation of care needs, and the use of the care process according to Finnish Care Classification" • E-care competence: "Use of eHealth tools in tailoring patient care" • E-documentation competence: "Electronic recording of patient data" • E-ethics competence: "Competence in the ethical and safe way to use patient information systems" (source: "Methods" section)	Questionnaire (16 items)
Prevalence studies							
Brady and Knox [24]	Northern Ireland	Cross sectional	N/A	98 psychiatric trainees/consultants (specialist registered, senior house officers, staff grades, consultants); age N/A	Self-rated computing skill levels	N/A	Questionnaire (6 items)
Hollander and Martin [25]	United States	Cross sectional	344 local health departments	Some of or all public health professional staff working in the local health departments; age N/A	Staff internet use and resources used	N/A	Questionnaire (N/A)
Kirchberg et al [34]	Germany	Cross sectional	N/A	93 physicians; 37 (40%) aged between 30 and 45 years	Level of knowledge of eHealth apps and data safety; mobile phone use; attitude toward (evaluation) medical apps for physician and patient use; evaluation of importance of medical app characteristics	N/A	Questionnaire (mobile phone use, 4 items; purpose of mobile phone use, 9 items, level of knowledge of eHealth apps and data safety, 9 items; evaluation of medical apps for physician use, list of 6 apps for patients and 5 apps for physicians; evaluation of importance of medical app characteristics, 7 items; evaluation of importance of privately used app characteristics, 7 items)

Study type and reference	Country	Study design	Setting(s)	Sample and profession; age	Competencies assessed	Definition provided of the competencies assessed	Tools/data collection method(s) and items
Kleib and Nagle [26]	Canada	Cross sectional	Acute care, community, other settings	2844 nurses (generalist registered nurses and registered psychiatric nurses); 1257 (44%) aged ≥50 years	Use of patient care technology; self-perceived informatics competencies (foundational ICT skills, information and knowledge management, professional and regulatory accountability, and use of ICT in delivery of patient care)	Developed framework	• Use of patient care technology (1 item) • Self-perceived informatics competencies: Canadian Nurse Informatics Competence Assessment Scale (21 items)
Kocher et al [27]	Switzerland	Explanatory sequential mixed method	University hospital, regional hospital, rheumatology outpatient clinics	47 professionals (registered nurses, physiotherapists, rheumatologists, occupational therapists, advanced practice nurses, general practitioners, psychologists, social workers, health policy); median age 41 (IQR 31-51) years	eHealth literacy (access, understand, appraise, apply)	• eHealth literacy: “people’s knowledge, motivation and competence to ‘access’, ‘understand’, ‘appraise’ and ‘apply’ health information from electronic sources to address or solve a health problem” • Access: “the ability to seek, find and obtain health information” • Understand: “the ability to comprehend information” • Appraise: “interpret and evaluate information” • Apply: “the ability to use health information to make informed decisions” (source: “Background” section)	Questionnaire based on previously validated instruments (access, 6 items; understand and appraise, 5 items; apply, 9 items)
Kujala et al [28]	Finland	Cross sectional	Public health organization	701 HCPs (nurses, social workers, physicians, dentists, ward secretaries, physiotherapists and other therapists, instrument or facility care personnel, health administration workers, psychologists); mean 44.1 (SD 11.9) years	Self-perceived eHealth competencies; actual patient guidance behaviors	eHealth competence: “A broad set of skills employing ICT and eHealth services, information management, multi-channel health coaching, patient communication, development and implementation” (source: “Background” section)	Questionnaire (self-perceived eHealth competencies, 9 items; actual patient guidance behaviors, 4 items)

Study type and reference	Country	Study design	Setting(s)	Sample and profession; age	Competencies assessed	Definition provided of the competencies assessed	Tools/data collection method(s) and items
MacLure and Stewart [29]	Scotland	Cross sectional	Community and hospital pharmacies	94 participants (pharmacists, reregistration pharmacy graduates, pharmacy technicians, dispensing assistants, medicine counter assistants); 34 (36.2%) aged ≤29 years	Self-reported digital literacy	The British Computer Society defines digital literacy as “Being able to make use of technologies to participate in and contribute to modern social, cultural, political and economic life”. A similar definition of digital literacy is adopted in the United States: “the ability to use information and communication technologies to find, evaluate, create, and communicate information; it requires both technical and cognitive skills” (source: “Background” section)	Self-reported digital literacy (1 item)
Polhamus et al [30]	United States	Cross sectional	N/A	36 maternal and child health professionals; 82% aged ≥40 years	Beliefs in the value of and confidence in using technology	Beliefs in the value of technology: “the extent to which they agreed with a set of questions about the value of a specific technology skill” (source: “Methods” section)	Questionnaire (beliefs, 3 items; confidence, 3 items)
Thomas and Rutter [31]	England	Cross sectional	Mixed settings	386 pharmacists, 83 (21.5%) aged between 50 and 59 years	Confidence in basic computer skills and use of key software applications	N/A	Questionnaire (16 items)
van Houwelingen et al [32]	The Netherlands	Cross sectional	Hospitals	1017 registered nurses; median age 41 (IQR 30-53) years	Confidence in their telehealth knowledge, skills, and attitudes	N/A	Questionnaire (31 items)
Zayapragassarazan and Kumar [33]	India	Cross sectional	Teaching hospitals	120 Health professional faculty working; 57 (40%) aged between 30 and 40 years	Using computer ability; awareness, knowledge, attitude, and computer skills about telemedicine	N/A	AKAS questionnaire [27] (awareness, 12 items; knowledge, 11 items; attitude, 11 items; information technology and computer skills, 13 items)

[a] HCP: health care professional.

[b] N/A: not available.

[c] eHEALS: eHealth Literacy Scale.

[d] AKAS: Awareness, Knowledge, Attitude, Skills.

[e] UTAUT: Unified Theory of Acceptance and Use of Technology.

[f] ICT: information and communications technology.

Table 2. Characteristics of included randomized controlled trial studies.

Reference	Country	Study design	Setting	Sample and profession; age	Competence(s) assessed and definitions	Intervention	Tools/data collection method(s)
Jouparinejad et al [49]	Iran	Interventional study	Hospitals	60 nurses; 26 (43.3%) aged between 30 and 40 years	• Nursing informatics competencies: Computer literacy: "The psychomotor skills to use computer tools, and knowledge of basic hardware and software functionality" • Informatics literacy: "Nurses' abilities to recognize, retrieve, evaluate and use information for patient care appropriately" • Information management skills: "apply the data to support clinical decisions, documentation, data integrity, confidentiality and security" (source: "Methods" section)	Three-day workshop with theory and practice to develop nursing informatics competencies	Questionnaires: Adapted Nursing Informatics Competence Assessment Tool (30 items): computer literacy (10 items), informatics literacy (13 items), information management skills (7 items)
Mastellos et al [48]	Malawi	Randomized controlled trial	Community	40 community health professionals; 23/39 (49%) aged ≥40 years	• Self-rated ICT[a] knowledge; attitudes toward using computers, tablets, and smartphones	3-week blended learning "Introduction to ICT and eHealth" course (intervention) versus traditional course (control) on same contents	Questionnaire including 10 items to assess self-rated ICT knowledge, and 10 items to assess attitudes

[a] ICT: information and communications technology.

Quality Assessment

The methodological quality was high in 5 analytical cross-sectional studies [35,39,42,46,47] (out of 12; Multimedia Appendix 3), in 1 prevalence study [26] (out of 11; Multimedia Appendix 4), and in 1 randomized control trial study [48] (out of 2; Multimedia Appendix 5). A total of 3 cross-sectional [38,41,43] and 7 prevalence studies [24,25,28-30,33,34] reported a low methodological quality. Among the former, no confounding factors were identified. By contrast, for all studies, the "Not applicable" option was assigned to the item regarding the use of "objective, standard criteria used for measurement of the condition." Among the prevalence data studies, the most unclear item (10/11 studies) was regarding the adequacy of the sample size. By contrast, the item most often scored as "No" (5/11 studies) was the sample description.

Digital Health Competencies Investigated

As many as 13/26 studies [27-30,35-38,42,44,46,47,49] reported the definitions of the concept assessed, which were retrieved from the "Methods" section in 8 studies [30,35,37,42,44,46,47,49].

As summarized in Table 3, "Self-rated competencies" were assessed with 140 items grouped into 4 subcategories. "Digital literacy" emerged as the first subcategory in terms of frequency (59 items, 14 studies) and included items used to assess the self-perceived level of competence in using technology without a specific health goal (eg, in using tablets and mobile phones [43], apps [24], the internet [48], digital cameras [43], and computer literacy [45]). The second was the "eHealth literacy" subcategory, which included the 40 items provided by the 8-item eHealth Literacy Scale (eHEALS) [15] adopted by 5 studies [35,39,42,44,45]. Then, "Patient-oriented competencies" (21 items, 4 studies) included items aimed at assessing the ability, for example, to train and advise patients about technology [32], suitable websites [44], and apps [27], to create confidentiality, to maintain an ethical attitude and convey empathy through videoconferencing [28,32], and to assess the needs of patients regarding telehealth [32]. Lastly, the "Process of care-oriented competencies" subcategory (20 items, 11 studies) included those items assessing the level of competence in retrieving, evaluating, and applying online information, as well as in using eHealth tools to inform the decision-making process in patient care [26,47,49].

The second category, "Psychological and emotional aspects toward digital technologies," was assessed with 110 items by 18 studies. The first subcategory, "Attitudes and beliefs" (82 items, 14 studies), included items assessing attitudes regarding the perceived benefits of the care delivered to and for patients (eg, quality of care and opportunity for self-care [46]); the work benefits perceived (eg, saving [41] and easy access to data [28]); the complexity [43]; the importance, value [28,30], and the feasibility in work [43] of using digital technologies and telemedicine [33,37]. Then, in the second subcategory, "Confidence" (21 items, 6 studies), most items were aimed at assessing the confidence in performing specific activities such as “searching the internet” [31] or “monitoring the patients’ health data using mobile apps” [46]. Finally, in the “Awareness” subcategory (7 studies, 4 items), items assessing the general level of awareness of telemedicine or health information resources and awareness meant as observability were included (eg, to observe the high use of information and communication technology in the workplace [43]).

The third category, “Use of digital technologies” (98 items, 13 studies), included the subcategory “General use of digital technologies” (51 items, 9 studies), which was adopted to investigate the extent to which health care professionals applied the digital technologies in general, for example, the use of computers, printers, the internet, email, and the “Use of digital technologies for specific functions” (47 items, 7 studies) for investigating specific functions as, for example, in documenting care [41], communicating with patients [40], or for research purposes [45].

Lastly, the fourth category, “Knowledge about digital technologies” (14 items, 5 studies), included items aimed at assessing knowledge regarding, for example, telemedicine [37], technical aspects [34], data protection and privacy requirements [32], security, and appropriateness of communication application (eg, WhatsApp, medCrowd) [34].

Table 3. Investigated areas of digital health competencies.

Category and subcategories	Item examples and references	Items, n (n=362)	Studies, n
Self-rated competencies		140	19
Digital literacy	• Self-rated level of computer skill on the application PowerPoint [43] • Level of skills in using body scanner [43]	59	14
eHealth literacy	• 8-item eHEALS[a] tool [15]	40	5
Patient-oriented competencies	• “Can put patients at ease when they feel insecure about using technology?” [32] • “Do you recommend apps to your patients that support them in a healthy lifestyle?” [28]	21	4
Process of care-oriented competencies	• “Can combine my nursing knowledge and experience effectively when using telehealth technology and decision-making” [32] • “I am able to recognize (at a distance) the needs of the patient and determine the care situation” [32]	20	11
Psychological and emotional aspects toward the use of digital technologies		110	18
Attitudes and Beliefs	• “I believe that using ICT[b] is cumbersome” [43] • “Using ICT is compatible with all aspects of my work” [43] • “Be a better caregiver by using information technology” [41]	82	14
Confidence	• “I believe I would be able to use a computer or mobile app to provide patient care” [48] • “Confidence using the Internet logging on” [31]	21	6
Awareness	• “Awareness of telemedicine” [37] • “ICT is very visible in the hospital where I work” [43]	7	4
Use of digital technologies		98	13
General use of digital technologies	• “Do you use and own a mobile phone?” [34] • “If you use the internet, how frequently do you use it?” [44]	51	9
Use of digital technologies for specific functions	• “Do you use the Internet regularly for medical/professional updates?” [44] • Using a computer for a specific clinical task: “Access online patient educational materials” [40]	47	7
Knowledge about digital technologies	• “Is it appropriate to use common email for professional communication in health systems?” [34] • “Do you think a legal obligation for external certification of medical apps is required?” [34]	14	5

[a] eHEALS: eHealth Literacy Scale.

[b] ICT: information and communications technology.

Tools Used to Assess Digital Health Competencies

In 9/26 studies [26,35-39,42,44,45], previously developed and validated tools were adopted to self-assess the competencies, with 5 studies [35,39,42,44,45] reporting the use of the e-HEALS tool [15], while the remaining used the Gassert/McDowell Computer Literacy Survey [38], the Canadian Nurse Informatics Competency Assessment Scale [26], the Awareness, Knowledge, Attitude, Skills tool [37], the Multicomponent Assessment of Computer Literacy, and the Pre-test for Attitudes Towards Computers in Healthcare Assessment tools [36]. The authors of the other studies developed ad hoc questionnaires, using 1 (eg, [32]) or multiple (eg, [49]) questionnaires with the number of items ranging from 1 [29] to 47 [33], mainly including several general dimensions (eg, Awareness, Self-efficacy, Attitudes) [32,33]. In most studies, tools were described in detail by reporting the dimensions of competencies under evaluation and the number of items; only in a few studies was the description poor (eg, [25]; Tables 1 and 2).

Discussion

The discussion has been developed under 2 main lines: around the principal findings emerged and the comparison of evidence emerged with available studies, by including in each the future directions recommended for both practice and research in this field.

Principal Findings

Despite the increased relevance of digital health competencies among health care professionals [50], in the last 20 years, only a few studies have been published, slightly more than 1 per year, with an increase in the last 5 years. Moreover, although there is an urgent need to equip health care professionals with appropriate competencies given the progressive digitalization [1], most studies available to date are cross sectional or prevalence in design and only 2 are experimental studies. In addition, a few studies have been conducted with high methodological quality, suggesting improvements in this research field.

Studies available have been conducted in developed (eg, United States, Europe) and developing (eg, Uganda) countries where different health digital transformations are in place. Therefore, our findings may help policymakers and educators to set competencies according to the stage of digitalization experienced regarding the infrastructures available. However, roughly half of the studies have been focused on hospitals, whereas the community settings and districts have been involved to a lesser extent despite their increased need to implement digitalization with competent health care professionals to address emerging inequalities and issues in terms of health care accessibility [51]. Moreover, studies have more often involved nurses, doctors, or mixed samples of health care professionals, suggesting that all health care profiles have been involved to date, albeit to a limited extend for some (eg, physiotherapists [8]). Given the progressive and expansive permeation of digitalization in the health care sector, all health care professionals should be involved in the assessment of digital health competencies aimed at tailoring educational strategies. Meriting attention is the variable age of participants involved in the studies, from new graduates to mature health care professionals close to retirement. The new generations, also called the “digital native generation” [52], have more attitudes toward digitalization [53], and this suggests the need to deepen this area of study by investigating in future studies specific digital health competencies, despite including other elements such as attitudes (eg, using a computer) that might be relevant only among mature health care professionals.

At the overall level, only half of the studies [27-30,35-38,42,44,46,47,49] reported the definitions of the competencies assessed, and these have been reported mainly in the “Methods” and “Background” sections.

This finding suggests that future studies should be strengthened in their conceptualization and grounded in their development on clear conceptual frameworks and definitions.

Four main categories of investigated areas regarding digital health competencies have emerged, along with 9 subcategories. The area most investigated to date is self-rated competencies, in line with available literature [4,9,10]. In particular, this area includes, among the others, competencies aimed at solving patients’ health or care plan issues. This point suggests an interest among the scientific community in investigating these competencies from innovative perspectives. Training, advising, and supporting patients in the appropriate and confident use of technologies and information retrieved from different ICTs, social media, and internet sources are crucial [54], as also underlined by the framework recently developed by a consortium of multiple European countries [14]. The interest in investigating psychological and emotional aspects of the use of digital technologies has grown increasingly over the years, being assessed in 18 studies. The perceived usefulness for smoothing the care processes, improving its quality and patient satisfaction, and understanding health conditions and the adherence to treatments are crucial elements. Attitudes, acceptance, and confidence [8,12] in using digital technologies, such as electronic prescriptions, remote monitoring, and electronic databases, have demonstrated a positive effect on care processes and patients’ outcomes [2].

A limited number of studies have investigated the use of and the knowledge regarding digital technologies. However, a review of frameworks on digital health competence identified these topics in almost 60% of them [5], suggesting an evident need to promote the awareness of these issues in future research, given the increasing threats to data safety from illegal hacking [55].

A lack of validated tools to measure digital health competencies has emerged. One-third of studies have used a validated tool, the eHEALS of Norman and Skinner [15], although it was developed for patients, thus requiring a specific validation process and adaptation in the field of health care professionals. Moreover, a propensity to develop ad hoc instruments rather than using those already validated has emerged. The reasons for this may rely on the limitations perceived by those available, as well as the rapid evolution of digital technologies and instruments that may require a continuous updating of the competencies to assess. Moreover, in all studies, the tools were intended to assess the perceptions of health care professions rather than measuring their digital health competencies objectively. Self-rated competencies might be useful while educational needs are investigated; however, the actual performance requires objective measurement systems that should be developed in this field.

Comparison With Prior Work

Comparing the categories of competencies emerged in available studies with frameworks established in this field might inform the future directions in both educational practice and research. At the overall level, similarities and divergences emerged. The most common competence between previous frameworks [5,7] and that emerged in our study included the technical skills and the ability to manage and understand information retrieved from technology, including the internet. Psychological and emotional aspects were also highly investigated [4,10,12,14] among the studies included in this review in line with Norgaard and colleagues' [7] eHLF for eHealth users. The engagement, the ability to take responsibility, the perception of feeling safe, and motivation were part of the framework as elements expressing the interaction between the person and the system [7]. Therefore, a debate on how these aspects may influence the digital health competence among health care professionals as well as how to transform them into professional competencies to evaluate merits further consideration.

A recent review indicated that most interventions that aimed to improve the digital health competencies of health care professionals focused on the capability rather than motivation in using eHealth [56]. Interventions promoting digital health competencies should also consider social and environmental factors, foreseeing participatory approaches, to bolster also the emotional and psychological factors toward the use of technology [56]. On the other side, discrepancies emerged regarding teaching, self-development, and learning abilities [12]. The National Health Service (NHS) framework on digital capability [12] embeds domains regarding the abilities, for example, to use digital technologies for personal learning and teaching others [12]. No similar elements emerged in our review. Therefore, future research should focus on the measurement of competencies regarding those aspects, while also considering increased use of blended learning and massive online open courses in continuing education [57].

As highlighted by a previous review [56], we also found that the competencies investigated are still mainly focused on health care professionals' perspectives. However, increased attention is required when considering the competencies to assess patients' needs, attitudes, barriers, facilitators, and potential benefits of being trained by health care professionals in the safe and appropriate use of technology and electronic information for health issues [47]. Therefore, from a self-perceived competence assessment mainly concerning general issues, efforts should now be addressed at developing patient-centered digital health care assessment tools capable of detecting all specific competencies involved in the entire process.

Strengths and Limitations

This systematic review has several limitations. First, despite the accuracy of the process preventing the risk of publication bias by screening 4 databases and the reference lists of the included studies, as well as the trial registries [58], some studies may have been missed given that we adopted the English language filter and gray literature has not been searched.

Second, we adopted "digital health competencies" as an umbrella term to refer to all concepts that emerged from the literature. Although the use of all possible terms (eg, "digital health literacy") in the search string and the inclusion process might have ensured inclusiveness, the summary provided under the same umbrella term might have introduced some limitations. Different aspects of digital health competencies, such as confidence, self-efficacy, attitude, and beliefs regarding digital technologies, have been considered relevant as affecting their use and appropriate adoption in the health care sector. Therefore, we included these elements as part of the umbrella term "digital health competencies," relying on the previous frameworks including them [6,7]. This process has been considered a strength of this review because of the consideration of the full range of competencies as assessed in available studies. Third, previous frameworks [5] mainly focused on the categorization according to technical skills or functions (eg, safety management or care coordination); the content analysis [22,23] performed allowed to include all competencies as documented in retrieved studies, not limiting them to just skills and behaviors. Therefore, we valued also self-concepts, values, personal traits, and motivation (eg, [43]) to map all factors involved. However, the content analysis conducted to categorize the competencies that emerged from included studies was performed by researchers with different backgrounds (eg, nursing, physiotherapy). Although carefully conducted and its reliability assessed with the interreliability rate, their interpretations might have influenced the final categorizations. Lastly, we have synthesized studies originating from different countries, thus differences in health care digitalization might affect the generalizability of the conclusion drawn on future directions for research and training of health care professionals. These should be targeted and adapted according to the characteristic of the countries by training health care professionals based on the technologies available at a local level.

Conclusion

Digital health competence among health care professionals is a new field of research that exploded in the last 5 years. However, studies conducted to date are mainly descriptive and have some methodological quality issues, suggesting lines of improvement. Moreover, with the increased decentralization of the health care sector, more studies are required in community settings, involving a wide range of health care professionals to assess the differences and commonalities in the competencies possessed and tailor specific educational strategies. Furthermore, with the increased size of the digital native generation among health care workers, specific digital health competencies instead of general ones should be investigated.

The different areas of competencies investigated to date might be considered while designing curricula for undergraduate, postgraduate, and continuing education processes. From the perspective of researchers, these competencies may drive the development of competence assessment tools, given the lack of validated instruments in this field, identifying more objective measures in addition to those based on self-perception. Furthermore, researchers should consider moving attention from the self-rated technical competencies to those embodying a patient-centered digital health care approach and related aspects that might affect the use of digital technologies.

In future frameworks and measurement tools, digital health competencies should be considered as a multicomponent competence, not limited to the technical skill, but rather expanded toward elements that might affect them. As our review showed, confidence, attitudes, beliefs, and awareness have been studied with increasing interest, suggesting the need to explore the relationships between different elements and understand how to train health care professionals properly. Curricula embedding the development of technical skills, knowledge, and psychological and emotional aspects of digital technology are recommended.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist [11].
[DOCX File , 24 KB - jmir_v24i8e36414_app1.docx]

Multimedia Appendix 2
Search strings according to searched databases.
[DOCX File , 14 KB - jmir_v24i8e36414_app2.docx]

Multimedia Appendix 3
Quality for analytical cross-sectional studies [16].
[DOCX File , 24 KB - jmir_v24i8e36414_app3.docx]

Multimedia Appendix 4
Quality assessment for prevalence data studies [17].
[DOCX File , 24 KB - jmir_v24i8e36414_app4.docx]

Multimedia Appendix 5
Quality assessment for randomized control trial [18].
[DOCX File , 22 KB - jmir_v24i8e36414_app5.docx]

References

1. World Health Organization (WHO). Classification of Digital Health Interventions v 1. WHO. Geneva, Switzerland: WHO; 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf [accessed 2022-07-28]
2. Huter K, Krick T, Domhoff D, Seibert K, Wolf-Ostermann K, Rothgang H. Effectiveness of Digital Technologies to Support Nursing Care: Results of a Scoping Review. J Multidiscip Healthc 2020;13:1905-1926. [doi: 10.2147/JMDH.S286193] [Medline: 33328736]
3. Brown J, Pope N, Bosco AM, Mason J, Morgan A. Issues affecting nurses' capability to use digital technology at work: An integrative review. J Clin Nurs 2020 Aug;29(15-16):2801-2819. [doi: 10.1111/jocn.15321] [Medline: 32416029]
4. 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]
5. Nazeha N, Pavagadhi D, Kyaw BM, Car J, Jimenez G, Tudor Car L. A Digitally Competent Health Workforce: Scoping Review of Educational Frameworks. J Med Internet Res 2020 Nov 05;22(11):e22706 [FREE Full text] [doi: 10.2196/22706] [Medline: 33151152]
6. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014 Jan 28;15:18 [FREE Full text] [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]
7. Norgaard O, Klokker BL, Hospital F, Astrid KD, Kayser L, Osborne R. Knowledge Management & E-Learning The eHealth literacy framework: A conceptual framework for characterizing eHealth users and their interaction with e- health systems. Knowledge Management & E-Learning 2015;7(4):522-540 [FREE Full text] [doi: 10.34105/j.kmel.2015.07.035]
8. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: A systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]

9. EU*US eHealth Work Project. Health Information Technology Competencies. Health Information Technology Competencies (HITCOMP). 2020. URL: http://hitcomp.org/competencies [accessed 2021-05-12]
10. Hübner U, Thye J, Shaw T, Elias B, Egbert N, Saranto K, et al. Towards the TIGER International Framework for Recommendations of Core Competencies in Health Informatics 2.0: Extending the Scope and the Roles. Stud Health Technol Inform 2019 Aug 21;264:1218-1222. [doi: 10.3233/SHTI190420] [Medline: 31438119]
11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
12. A Health and Care Digital Capabilities Framework. National Health Service (NHS). London, UK: NHS; 2018. URL: https://www.hee.nhs.uk/sites/default/files/documents/Digital%20Literacy%20Capability%20Framework%202018.pdf [accessed 2021-05-12]
13. Thye J, Shaw T, Hüsers J, Esdar M, Ball M, Babitsch B, et al. What Are Inter-Professional eHealth Competencies? Stud Health Technol Inform 2018;253:201-205. [Medline: 30147073]
14. Recommendations on a Common Framework for Mapping Health Professionals? eHealth Competencies. Joint Action to Support the eHealth Network. JAseHN (Joint Action to support the eHealth Network). 2018. URL: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5bc2cac91&appId=PPGMS [accessed 2021-05-12]
15. Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. Stat Med 2002 Jun 15;21(11):1635-1640 [FREE Full text] [doi: 10.1002/sim.1190] [Medline: 12111924]
16. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, et al. Chapter 7: Systematic reviews of etiology and risk. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, SA, Australia: JBI; Jan 2000.
17. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. Int J Evid Based Healthc 2015 Sep;13(3):147-153. [doi: 10.1097/XEB.0000000000000054] [Medline: 26317388]
18. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3. Systematic reviews of effectiveness. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, SA, Australia: JBI; Jan 2020.
19. Pluye P, Gagnon M, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
20. Danielis M, Povoli A, Mattiussi E, Palese A. Understanding patients' experiences of being mechanically ventilated in the Intensive Care Unit: Findings from a meta-synthesis and meta-summary. J Clin Nurs 2020 Jul;29(13-14):2107-2124. [doi: 10.1111/jocn.15259] [Medline: 32243007]
21. Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. BMJ 2004 May 29;328(7451):1312-1315 [FREE Full text] [doi: 10.1136/bmj.328.7451.1312] [Medline: 15166072]
22. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
23. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
24. Brady J, Knox S. Computer access, skills and training among consultants and psychiatric trainees in Northern Ireland. Psychiatr. bull 2018 Jan 02;28(11):415-417. [doi: 10.1192/pb.28.11.415]
25. Hollander SM, Martin ER. Public health professionals in the Midwest: a profile of connectivity and information technology skills. Bull Med Libr Assoc 1999 Jul;87(3):329-336 [FREE Full text] [Medline: 10427435]
26. Kleib M, Nagle L. Development of the Canadian Nurse Informatics Competency Assessment Scale and Evaluation of Alberta s Registered Nurses Self-perceived Informatics Competencies. CIN: Computers, Informatics, Nursing 2018;36(7):350-358. [doi: 10.1097/cin.0000000000000435]
27. Kocher A, Simon M, Dwyer AA, Blatter C, Bogdanovic J, Künzler-Heule P, et al. Patient and healthcare professional eHealth literacy and needs for systemic sclerosis support: a mixed methods study. RMD Open 2021 Sep;7(3):e001783 [FREE Full text] [doi: 10.1136/rmdopen-2021-001783] [Medline: 34475248]
28. Kujala S, Rajalahti E, Heponiemi T, Hilama P. Health Professionals' Expanding eHealth Competences for Supporting Patients' Self-Management. Stud Health Technol Inform 2018;247:181-185. [Medline: 29677947]
29. MacLure K, Stewart D. Self-Reported Digital Literacy of the Pharmacy Workforce in North East Scotland. Pharmacy (Basel) 2015 Oct 15;3(4):182-196 [FREE Full text] [doi: 10.3390/pharmacy3040182] [Medline: 28975912]
30. Polhamus B, Farel A, Trester A. Enhancing technology skills of maternal and child health professionals. Matern Child Health J 2000 Dec;4(4):271-275. [doi: 10.1023/a:1026651823204] [Medline: 11272347]
31. Thomas P, Rutter P. A computer literacy skills profile of pharmacists residing in two counties of England. Health Info Libr J 2008 Dec;25(4):288-294 [FREE Full text] [doi: 10.1111/j.1471-1842.2008.00767.x] [Medline: 19076675]
32. van Houwelingen CTM, Ettema RGA, Kort HSM, Ten Cate O. Hospital Nurses' Self-Reported Confidence in Their Telehealth Competencies. J Contin Educ Nurs 2019 Jan 01;50(1):26-34 [FREE Full text] [doi: 10.3928/00220124-20190102-07] [Medline: 30645656]

33. Zayapragassarazan Z. Awareness, Knowledge, Attitude and Skills of Telemedicine among Health Professional Faculty Working in Teaching Hospitals. JCDR 2016;10:JC01-4. [doi: 10.7860/jcdr/2016/19080.7431]
34. Kirchberg J, Fritzmann J, Weitz J, Bork U. eHealth Literacy of German Physicians in the Pre-COVID-19 Era: Questionnaire Study. JMIR Mhealth Uhealth 2020 Oct 16;8(10):e20099 [FREE Full text] [doi: 10.2196/20099] [Medline: 33064102]
35. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health Professionals' Acceptance and Adoption of eHealth Interventions in Inpatient Routine Care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
36. Gürdaş Topkaya S, Kaya N. Nurses' computer literacy and attitudes towards the use of computers in health care. Int J Nurs Pract 2015 May;21 Suppl 2:141-149. [doi: 10.1111/ijn.12350] [Medline: 24804813]
37. Elhadi M, Elhadi A, Bouhuwaish A, Bin Alshiteewi F, Elmabrouk A, Alsuyihili A, et al. Telemedicine Awareness, Knowledge, Attitude, and Skills of Health Care Workers in a Low-Resource Country During the COVID-19 Pandemic: Cross-sectional Study. J Med Internet Res 2021 Feb 25;23(2):e20812 [FREE Full text] [doi: 10.2196/20812] [Medline: 33600350]
38. Campbell CJ, McDowell DE. Computer literacy of nurses in a community hospital: where are we today? J Contin Educ Nurs 2011 Aug;42(8):365-370. [doi: 10.3928/00220124-20110215-01] [Medline: 21366161]
39. Do BN, Tran TV, Phan DT, Nguyen HC, Nguyen TTP, Nguyen HC, et al. Health Literacy, eHealth Literacy, Adherence to Infection Prevention and Control Procedures, Lifestyle Changes, and Suspected COVID-19 Symptoms Among Health Care Workers During Lockdown: Online Survey. J Med Internet Res 2020 Nov 12;22(11):e22894 [FREE Full text] [doi: 10.2196/22894] [Medline: 33122164]
40. Duffy FF, Fochtmann LJ, Clarke DE, Barber K, Hong S, Yager J, et al. Psychiatrists' Comfort Using Computers and Other Electronic Devices in Clinical Practice. Psychiatr Q 2016 Sep;87(3):571-584 [FREE Full text] [doi: 10.1007/s11126-015-9410-2] [Medline: 26667248]
41. Gaumer GL, Koeniger-Donohue R, Friel C, Sudbay MB. Use of information technology by advanced practice nurses. Comput Inform Nurs 2007;25(6):344-352. [doi: 10.1097/01.NCN.0000299656.59519.06] [Medline: 18000431]
42. Kritsotakis G, Andreadaki E, Linardakis M, Manomenidis G, Bellali T, Kostagiolas P. Nurses' ehealth literacy and associations with the nursing practice environment. Int Nurs Rev 2021 Sep 18;68(3):365-371. [doi: 10.1111/inr.12650] [Medline: 34004022]
43. Olok GT, Yagos WO, Ovuga E. Knowledge and attitudes of doctors towards e-health use in healthcare delivery in government and private hospitals in Northern Uganda: a cross-sectional study. BMC Med Inform Decis Mak 2015 Nov 04;15(1):87 [FREE Full text] [doi: 10.1186/s12911-015-0209-8] [Medline: 26537731]
44. Shiferaw KB, Mehari EA. Internet use and eHealth literacy among health-care professionals in a resource limited setting: a cross-sectional survey. Adv Med Educ Pract 2019 Jul;10:563-570 [FREE Full text] [doi: 10.2147/AMEP.S205414] [Medline: 31440113]
45. Tesfa GA, Kalayou MH, Zemene W. Electronic Health-Information Resource Utilization and Its Associated Factors among Health Professionals in Amhara Regional State Teaching Hospitals, Ethiopia. AMEP 2021 Mar;Volume 12:195-202. [doi: 10.2147/amep.s289212]
46. Thapa S, Nielsen JB, Aldahmash AM, Qadri FR, Leppin A. Willingness to Use Digital Health Tools in Patient Care Among Health Care Professionals and Students at a University Hospital in Saudi Arabia: Quantitative Cross-sectional Survey. JMIR Med Educ 2021 Feb 19;7(1):e18590 [FREE Full text] [doi: 10.2196/18590] [Medline: 33605896]
47. Vehko T, Hyppönen H, Puttonen S, Kujala S, Ketola E, Tuukkanen J, et al. Experienced time pressure and stress: electronic health records usability and information technology competence play a role. BMC Med Inform Decis Mak 2019 Aug 14;19(1):160 [FREE Full text] [doi: 10.1186/s12911-019-0891-z] [Medline: 31412859]
48. Mastellos N, Tran T, Dharmayat K, Cecil E, Lee H, Wong CCP, et al. Training community healthcare workers on the use of information and communication technologies: a randomised controlled trial of traditional versus blended learning in Malawi, Africa. BMC Med Educ 2018 Apr 02;18(1):61 [FREE Full text] [doi: 10.1186/s12909-018-1175-5] [Medline: 29609596]
49. Jouparinejad S, Foroughameri G, Khajouei R, Farokhzadian J. Improving the informatics competency of critical care nurses: results of an interventional study in the southeast of Iran. BMC Med Inform Decis Mak 2020 Sep 11;20(1):220 [FREE Full text] [doi: 10.1186/s12911-020-01244-5] [Medline: 32917187]
50. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital Education in Health Professions: The Need for Overarching Evidence Synthesis. J Med Internet Res 2019 Feb 14;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
51. Davies AR, Honeyman M, Gann B. Addressing the Digital Inverse Care Law in the Time of COVID-19: Potential for Digital Technology to Exacerbate or Mitigate Health Inequalities. J Med Internet Res 2021 Apr 07;23(4):e21726 [FREE Full text] [doi: 10.2196/21726] [Medline: 33735096]
52. Chicca J, Shellenbarger T. Connecting with Generation Z: Approaches in Nursing Education. Teaching and Learning in Nursing 2018 Jul;13(3):180-184 [FREE Full text] [doi: 10.1016/j.teln.2018.03.008]

53. Stevanin S, Palese A, Bressan V, Vehviläinen-Julkunen K, Kvist T. Workplace-related generational characteristics of nurses: A mixed-method systematic review. J Adv Nurs 2018 Jun;74(6):1245-1263. [doi: 10.1111/jan.13538] [Medline: 29424038]
54. Tsirintani M. Fake News and Disinformation in Health Care- Challenges and Technology Tools. Stud Health Technol Inform 2021 May 27;281:318-321. [doi: 10.3233/SHTI210172] [Medline: 34042757]
55. Grande D, Luna Marti X, Feuerstein-Simon R, Merchant RM, Asch DA, Lewson A, et al. Health Policy and Privacy Challenges Associated With Digital Technology. JAMA Netw Open 2020 Jul 01;3(7):e208285 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8285] [Medline: 32644138]
56. Brice S, Almond H. Health Professional Digital Capabilities Frameworks: A Scoping Review. J Multidiscip Healthc 2020;13:1375-1390 [FREE Full text] [doi: 10.2147/JMDH.S269412] [Medline: 33173300]
57. Longhini J, De Colle B, Rossettini G, Palese A. What knowledge is available on massive open online courses in nursing and academic healthcare sciences education? A rapid review. Nurse Educ Today 2021 Apr;99:104812. [doi: 10.1016/j.nedt.2021.104812] [Medline: 33610051]
58. Song F, Hooper, Loke Y. Publication bias: what is it? How do we measure it? How do we avoid it? OAJCT 2013 Jul;5:71-81. [doi: 10.2147/OAJCT.S34419]

Abbreviations

AKAS: Awareness, Knowledge, Attitude, Skills
eHEALS: eHealth Literacy Scale
eHLF: eHealth Literacy Framework
HCP: health care professional
HITCOMP: Health Information Technology Competencies
ICT: information and communications technology
N/A: not available
NHS: National Health Service
PRE-HIT: Patient Readiness to Engage in Health Internet Technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
TIGER: Technology Informatics Guiding Education Reform
UTAUT: Unified Theory of Acceptance and Use of Technology
WHO: World Health Organization

Edited by T Leung; submitted 13.01.22; peer-reviewed by U Bork, T Heponiemi; comments to author 25.04.22; revised version received 16.05.22; accepted 21.06.22; published 18.08.22.

Please cite as:
Longhini J, Rossettini G, Palese A
Digital Health Competencies Among Health Care Professionals: Systematic Review
J Med Internet Res 2022;24(8):e36414
URL: https://www.jmir.org/2022/8/e36414
doi:10.2196/36414
PMID:35980735

Review

The Barriers and Facilitators of eHealth-Based Lifestyle Intervention Programs for People With a Low Socioeconomic Status: Scoping Review

Isra Al-Dhahir[1], MSc; Thomas Reijnders[1], PhD; Jasper S Faber[2], MSc; Rita J van den Berg-Emons[3,4], PhD; Veronica R Janssen[1,5], PhD; Roderik A Kraaijenhagen[6,7], MD, PhD; Valentijn T Visch[2], PhD; Niels H Chavannes[8,9], MD, Prof Dr; Andrea W M Evers[1,10], Prof Dr

[1]Faculty of Social and Behavioral Sciences, Health, Medical and Neuropsychology Unit, Leiden University, Leiden, Netherlands
[2]Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands
[3]Department of Rehabilitation Medicine, Erasmus University Medical Centre, Rotterdam, Netherlands
[4]Capri Cardiac Rehabilitation, Rotterdam, Netherlands
[5]Department of Cardiology, Leiden University Medical Center, Leiden, Netherlands
[6]Vital10, Amsterdam, Netherlands
[7]NDDO Institute for Prevention and Early Diagnostics (NIPED), Amsterdam, Netherlands
[8]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands
[9]National eHealth Living Lab, Leiden University Medical Centre, Leiden, Netherlands
[10]Medical Delta, Leiden University, Delft University of Technology, Erasmus University, Delft, Netherlands

Corresponding Author:
Isra Al-Dhahir, MSc
Faculty of Social and Behavioral Sciences
Health, Medical and Neuropsychology Unit
Leiden University
Wassenaarseweg 52
Leiden, 2333 AK
Netherlands
Phone: 31 715273627
Email: i.al-dhahir@fsw.leidenuniv.nl

Abstract

Background: Promoting health behaviors and preventing chronic diseases through a healthy lifestyle among those with a low socioeconomic status (SES) remain major challenges. eHealth interventions are a promising approach to change unhealthy behaviors in this target group.

Objective: This review aims to identify key components, barriers, and facilitators in the development, reach, use, evaluation, and implementation of eHealth lifestyle interventions for people with a low SES. This review provides an overview for researchers and eHealth developers, and can assist in the development of eHealth interventions for people with a low SES.

Methods: We performed a scoping review based on Arksey and O'Malley's framework. A systematic search was conducted on PubMed, MEDLINE (Ovid), Embase, Web of Science, and the Cochrane Library, using terms related to a combination of the following key constructs: eHealth, lifestyle, low SES, development, reach, use, evaluation, and implementation. There were no restrictions on the date of publication for articles retrieved upon searching the databases.

Results: The search identified 1323 studies, of which 42 met our inclusion criteria. An update of the search led to the inclusion of 17 additional studies. eHealth lifestyle interventions for people with a low SES were often delivered via internet-based methods (eg, websites, email, Facebook, and smartphone apps) and offline methods, such as texting. A minority of the interventions combined eHealth lifestyle interventions with face-to-face or telephone coaching, or wearables (blended care). We identified the use of different behavioral components (eg, social support) and technological components (eg, multimedia) in eHealth lifestyle interventions. Facilitators in the development included iterative design, working with different disciplines, and resonating intervention content with users. Facilitators for intervention reach were use of a personal approach and social network, reminders, and self-monitoring. Nevertheless, barriers, such as technological challenges for developers and limited financial resources, may hinder intervention development. Furthermore, passive recruitment was a barrier to intervention reach. Technical difficulties and the use of self-monitoring devices were common barriers for users of eHealth interventions. Only limited data on barriers and facilitators for intervention implementation and evaluation were available.

Conclusions: While we found large variations among studies regarding key intervention components, and barriers and facilitators, certain factors may be beneficial in building and using eHealth interventions and reaching people with a low SES. Barriers and facilitators offer promising elements that eHealth developers can use as a toolbox to connect eHealth with low SES individuals. Our findings suggest that one-size-fits-all eHealth interventions may be less suitable for people with a low SES. Future research should investigate how to customize eHealth lifestyle interventions to meet the needs of different low SES groups, and should identify the components that enhance their reach, use, and effectiveness.

(J Med Internet Res 2022;24(8):e34229) doi:10.2196/34229

KEYWORDS

eHealth; lifestyle interventions; health behaviors; low socioeconomic status; intervention development; barriers; facilitators; prevention; intervention evaluation

Introduction

Chronic diseases, such as cardiovascular disease and type 2 diabetes, accounted for 74% of deaths globally in 2019 [1]. These diseases are often preventable and treatable. Adopting a healthy lifestyle, such as smoking cessation, increased physical activity, a balanced diet, and decreased alcohol consumption, can reduce the risk of developing a chronic disease [2]. Traditional lifestyle interventions have been shown to be effective in helping people adopt a healthy lifestyle [3,4]. However, these interventions mostly focus on the general population and often disregard vulnerable groups, such as those with a low socioeconomic status (SES; people with a low income or low education, or who are from deprived neighborhoods). There is firm evidence that people with a low SES often engage in more risky lifestyle behaviors and have an increased risk for various chronic diseases and premature death than those with a high SES [2,5-7]. Health inequalities for low SES are associated with a reduced life expectancy of 5 to 10 years and a reduced disability-free life expectancy of 10 to 20 years [8]. Furthermore, it seems that low income and poverty are more often associated with poorer mental health [9,10]. A systematic review by Bull et al [11] found that when lifestyle interventions focus on people with a low SES, most result in small and variable effects [11]. These findings may be due to designers not tailoring lifestyle interventions specifically to people with a low SES or not taking into account their specific characteristics and needs [12,13]. For instance, compared to the general population, individuals with a low SES living in poverty may focus more on coping with their current stressful everyday life (ie, money-related stress and unfavorable work environment). These stressors can lead to choosing unhealthy coping strategies, such as tobacco smoking [14], which could be detrimental to long-term health [14]. These challenges call for a different approach to help this priority group adopt and maintain a healthy lifestyle.

eHealth interventions could proactively support people with a low SES to adopt lifestyle changes [15]. The use of the latest information and communication technologies, such as websites, smartphones, email, text messaging, tablets, and smartwatches [16], offers health professionals and researchers more options to tailor intervention content to the specific needs and characteristics of the user [17]. Furthermore, eHealth interventions can provide users with the information, skills, and resources needed for a positive lifestyle change efficiently and interactively [18]. Health practitioners can reach diverse populations more easily with these interventions than with traditional interventions [18]. eHealth interventions can be supported by video or audio and delivered in an accessible manner to patients for use in their own time and home environment [19]. However, current eHealth interventions require users to have good digital skills and a high literacy level, which are often lacking in low SES groups. Moreover, such interventions must consider their different life situations, health care needs, and eHealth expectations [20]. When eHealth interventions do not consider the needs of this target group, intervention uptake can hinder and reinforce the inequitable use of eHealth, exacerbating health inequalities [21,22].

Studies have shown promising results for eHealth among people with a low SES [23]. For instance, Brown et al [23] showed that their eHealth lifestyle intervention for low SES individuals yielded small but significant changes in behavior. Hayba et al [24] suggested that even modestly effective interventions, sustainably deployed to target vulnerable groups (eg, low SES groups), would add value to the field of public health. Even though there is a growing body of research on eHealth lifestyle interventions for this vulnerable group, there is a lack of insight into how eHealth interventions are currently developed, used, and implemented for people with a low SES. Recently, there has been an increased focus on the specific needs and characteristics of low SES groups to bridge this digital divide. For example, the World Health Organization’s digital intervention guidelines for eHealth usage to improve patient care devoted special attention to the needs, preferences, and circumstances of vulnerable groups, such as people with low (digital) literacy skills [25]. However, current guidelines fall short for researchers and developers who want to develop eHealth lifestyle interventions tailored to people with a low SES. Therefore, this scoping review aims to identify intervention components, barriers, and facilitators in the development, reach, use, evaluation, and implementation of existing eHealth lifestyle interventions for low SES populations.

Methods

Scoping Review Methodology

We conducted a systematic scoping review from June to September 2019. In July 2021, we updated the search following the same procedures. There were no restrictions on the date of publication for articles retrieved upon searching the databases. Since the research area of eHealth lifestyle interventions for low SES groups is still in its infancy, a scoping review method was chosen because it is an appropriate methodology to map key concepts and identify knowledge gaps. A scoping review also offers the opportunity to review published literature with different methodological designs. It further examines the existing literature concerning the volume, nature, and characteristics of the primary research [26]. We used Arksey and O'Malley's methodological framework as a guide for the review [26].

Search Strategy and Eligibility Criteria

We defined the following 5 categories based on 2 frameworks used for the development process of eHealth interventions: development, reach, use, evaluation, and implementation [27,28]. The first framework is the Center for eHealth Research (CeHRes) roadmap, a framework for eHealth development, implementation, and evaluation that combines and uses aspects from approaches like human-centered design, persuasive technology, and business modeling [27]. The second framework is RE-AIM (reach, effectiveness, adoption, implementation, and maintenance), which describes the stages in intervention development and implementation [28]. The categories *development*, *use*, and *evaluation* were derived from the CeHRes roadmap, and the categories *reach* and *implementation* were derived from the RE-AIM framework.

After we defined the scope of the review, we developed a search strategy together with an experienced librarian and domain experts (Multimedia Appendix 1). We searched PubMed, MEDLINE (Ovid), Embase, Web of Science, and the Cochrane Library, using a combination of the following key constructs: eHealth, lifestyle (physical activity, nutrition, alcohol, smoking, and sleep), low SES, and development, reach, use, evaluation, and implementation. The definitions of these key constructs are provided in Multimedia Appendix 2. These databases were chosen because they cover a wide range of scientific articles on eHealth. For each construct, several keywords (spelling variations and synonyms) were used. Exclusion and inclusion criteria were defined based on relevant literature and in consultation with domain experts, after which relevant studies were selected (Textbox 1).

Textbox 1. Inclusion and exclusion criteria for selecting the studies.

Inclusion criteria

- Description of an eHealth/web-based intervention or mHealth/telemedicine intervention
- Definition of socioeconomic status (SES) as the position of an individual on a socioeconomic scale that measures factors by a single variable, such as education, income, or neighborhood status, or multiple variables
- Focus on at least one lifestyle component (physical activity, diet, alcohol, smoking, sleep, or overweight)
- Targeting of a low SES population (>18 years of age)
- Presentation of information on development, use, reach, evaluation, or implementation
- Publication of full text in English
- Any study type (included study protocols)

Exclusion criteria

- Measurement of SES using other variables (eg, race and ethnicity)
- Conference abstracts and reviews presenting filtered information, such as systematic reviews, scoping reviews, and narrative reviews

Data Extraction and Analysis

The eligibility criteria were used to review the articles. Initially, IA screened the titles and abstracts for the first selection of articles. Then, IA checked the bibliographic reference lists of publications that remained after full-text selection to identify any additional eligible publications. Any doubt about the included studies was discussed with the other authors. We extracted general study characteristics (eg, the year of publication and country), and details on SES, effectiveness, development, reach, use, evaluation, and implementation. Data were extracted as barriers or facilitators if they were related to the development, reach, use, evaluation, or implementation phases of the intervention and they were identified or mentioned as facilitators or barriers by the included studies. Even if the barriers and facilitators were mentioned in one of the included studies, they were eligible for inclusion. If there were uncertainties concerning under which phase the barriers and facilitators fell, they were discussed with the other authors.

Furthermore, we selected additional categories based on the CONSORT-EHEALTH checklist (V1.6), which provides helpful guidance on what eHealth studies should report [29]. These added categories were the (behavioral) theories or models used to develop and evaluate an eHealth intervention. The categories also included the level of human involvement in the intervention (eg, automated or human guidance) during the development, evaluation, and implementation (eg, health professionals and researchers).

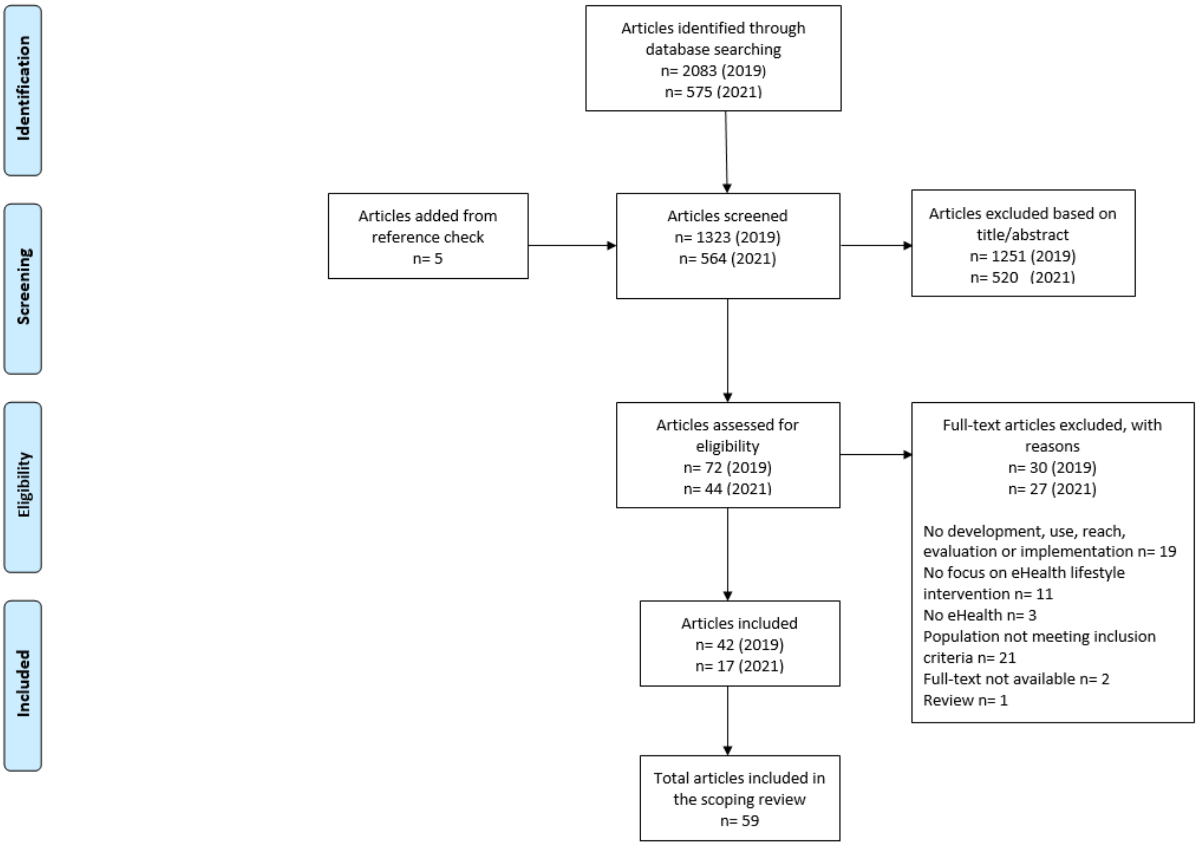
The selected articles were mapped, and data were recorded in Microsoft Excel (Multimedia Appendix 3). Data were synthesized narratively, and the findings were then summarized and grouped into themes as defined by the authors.

Results

Study Selection

The systematic search across the databases revealed 2083 potentially relevant citations. After removing duplicates (n=765) and screening 1323 titles and abstracts, 72 full-text articles were screened for eligibility. Of these, 42 articles met the eligibility criteria and were included in this review (Figure 1). The updated search led to 17 articles that were included in this review.

Figure 1. Flow diagram of the article selection process.



Description of the Included Studies

The studies included were randomized controlled trials (RCTs) [23,30-40], observational studies [41-56], and design studies [57-77]. Several RCTs and observational studies evaluated eHealth interventions on health outcomes (eg, BMI, blood pressure, and hemoglobin A_{1c}) [30, 34, 35-37, 40, 43, 45, 47, 48, 53, 78], nutrition-related behavior outcomes [32, 38, 39, 79, 80], physical activity–related outcomes, smoking-related outcomes [23,31,46,51,81,82], usage outcomes [33,49,54], and reach [44], as well as feasibility and acceptability outcomes [47,80] (Multimedia Appendix 3). Design studies examined recruitment [57], usability, feasibility [57-65], development, or acceptability of eHealth interventions [58-60,64,65,67,68,70].

The interventions were aimed at weight loss (n=9), physical activity (n=9), healthy eating (n=11), smoking (n=13), and alcohol use (n=2), and 17 interventions focused on multiple behaviors. The target audience of these interventions was mainly low SES participants; several studies also targeted a highly educated population [31-33,36,44,53,82].

The different studies assessed the education level [23,30-33,36,40,44,46,53-55,78,82], occupation [23,67], or income level of the participants [43,45,48,51,73,74]. In addition, the participants were recruited from a low SES neighborhood (residents who were unemployed, had a low education, or had a financial disadvantage) [60,64]. A summary of the study characteristics is presented in Table 1.

Table 1. Summary of the study characteristics (N=59).

First author, year	Study design	Target population	Type of eHealth technology
Aguilera [83], 2020	Protocol	Low income[a]	App (and SMS text messaging)
Aldoory [71], 2016	Design	Low income	SMS text messaging
Athavale [72], 2016	Design (part of RCT[b])	Low income[a]	Web-based
Atkinson [66], 2009	Design	Low income	Web-based
Bond [65], 2021	Design	Low income[a]	Web-based and SMS text messaging
Brown [23], 2014	RCT	Low SES[c]	Web-based
Griffin [48], 2020	Observational	Low income	SMS text messaging
Brunette [81], 2015	Quasiexperimental	Socioeconomically disadvantaged[d]	Web-based (single session)
Burner [80], 2020	Quasiexperimental	Low income[a]	SMS text messaging
Carolan-Olah [64], 2021	Design	Low SES neighborhoods	Web-based
Cavallo [47], 2021	Observational	Low income	Social media
Lepore [49], 2021	Observational (part of RCT)	Low income[a]	App
Stanczyk [31], 2013	RCT	Low, middle, and high education	Web-based
Clarke [39], 2019	RCT	Low income	App
van Dijk [84], 2021	Protocol	Low SES[a]	Web-based on smartphone
Brown [46], 2012	Observational	Low SES[c]	Web-based
Evans [62], 2019	Design	Low income	App
Flaherty [67], 2020	Design	Low SES[e]	App
Delrahim-Howlett [38], 2011	RCT	Low income	Web-based
Fontil [61], 2016	Design	Low income	Web-based
Garvin [73], 2019	Design	Low income	App
Golsteijn [85], 2017	Protocol	Low, middle, and high education	Web-based
Foley [37], 2016	RCT	Socioeconomically disadvantaged	SMS text messaging
Greene [54], 2021	Observational (secondary data analysis)	Low and middle educated	App
Cavallo [45], 2016	Observational	Low income	Web-based and social media
Tagai [50], 2020	Observational	Low income	SMS text messaging
Golsteijn [36], 2017	RCT	Low, middle, and high education	Web-based
Griffin [43], 2018	Observational	Low income	SMS text messaging
Kim [86], 2018	Nonrandomized design	Low income[a]	Web-based
Kothari [68], 2020	Design	Low income[a]	App
Leak [63], 2014	Design	Low income[a]	Social media
Kendzor [51], 2020	Observational	Low income	App
van Dongen [44], 2012	Observational	Low, middle, and high education	Web-based (email)
Lohse [52], 2013	Observational	Low income	Social media
Mayberry [74], 2016	Design	Low income	SMS text messaging
Michie [55], 2012	Design and observational	Low SES	Web-based
Neuenschwander [79], 2013	Block equivalence randomized trial	Low income	Web-based
Pathak [75], 2021	Design	Low income[a]	App (and SMS text messaging)
Patten [87], 2019	Nonrandomized design	Low income	SMS text messaging

First author, year	Study design	Target population	Type of eHealth technology
Radhakrishnan [53], 2016	Design and observational	Low, middle, and high education	App (mobile device video game)
Herring [35], 2014	RCT	Low income[a]	Web-based (social media) and mobile phone (text messaging)
Régnier [60], 2018	Design	Low SES neighborhoods	App
Ramirez [34], 2017	Pilot RCT	Low income[a]	Text messaging or voice
Silfee [56], 2018	Design and observational	Low income	Web-based
Silfee [42], 2019	Observational	Low income[a]	Web-based
Silk [41], 2008	Observational	Low income	Web-based vs games
Simons [58], 2018	Design	Lower education	App
Simons [40], 2018	RCT	Lower education	App
Spears [59], 2019	Design	Low SES	SMS text messaging
Schneider [33], 2012	RCT	Low, middle, and high education	Web-based
Stanczyk [82], 2014	Data from RCT	Low, middle, and high education	Web-based
Springvloet [32], 2015	RCT	Low, middle, and high education	Web-based
Stotz [76], 2018	Design	Low income	Web-based on smartphone
Tabak [77], 2018	Design	Low income[a]	SMS text messaging
Lohse [57], 2013	Design	Low income[a]	Social media
Wayne [30], 2015	RCT	Low SES	App
Wayne [78], 2014	Single arm	Low SES	App
Whittemore [70], 2020	Design	Low income	SMS text and MMS messaging
Yee [69], 2020	Design	Low income[a]	SMS text messaging

[a] Socioeconomic status was not specified in the study.

[b] RCT: randomized controlled trial.

[c] SES: socioeconomic status.

[d] Low education, unemployment, or living in poverty.

[e] Socioeconomic status was determined by the occupation and employment status of the household's primary income earner.

Intervention Development and Evaluation

In the various stages of the development and evaluation of the intervention (ie, problem definition, development, and implementation for the study) [48,50,59,60,64,66,69,70,84], several studies involved stakeholders, which included family members, experts, key informants [50,61,69,75,84,86], health professionals, and end users [48,54,64-66,68-70,73,75,83]. However, some studies provided little information on the identification of stakeholders and did not clarify the level of involvement of stakeholders and end users [37,44,50,56,66,70,78,79,81,83,87]. The studies used multiple methods, such as interviews, focus groups, and user testing [50,54,57,64-70,73,75,83], to gain insights from end users and stakeholders. Researchers used focus groups to map the needs and problems of the (potential) users [58,59,61,62,66,68,70,73] and to gain input from stakeholders to adapt existing interventions [61,70,77,86]. These methods also helped the researchers to gain insight into the challenges that participants experienced while using the intervention [65,73,86] and their thoughts on the requirements of successful participation [65,66,68,77,86]. Furthermore, other methods used the Community Engagement Studio [74], a consumer panel [32], and a collective discussion group [60]. The researchers used these methods to improve the accessibility of the interventions for the end users [60,74]. For the development phase, facilitators and barriers were related to technology and content factors. However, regarding the evaluation of the interventions, limited facilitators and barriers were mentioned. Several studies adapted existing interventions, which were developed and tested in different SES groups with various health concerns, such as diabetes, hypertension, mental disorders, and pregnancy [38,45,47,49,56,57,61,70,72,77,79,81,85-87].

Studies adapted these interventions and the delivery modality for use in different low SES groups. Nevertheless, many studies retained most of the content and components of the existing interventions [45,56,57,61,70,72,77,79,81,87]. Many studies chose to adapt the content of the intervention and apply linguistic and content simplification, such as using plain language and low content load through the use of images and videos [57,61,70,81,86,87]. Some studies also made cultural adaptations by using updated cultural components [61,86], translating the content into a second language (eg, Spanish), and employing bilingual coaches [61,86]. Although intervention adaptation was common, documentation of the adjustment process was scarce. Only 3 studies [70,77,85] described in detail the adaptation process and what changes they performed. Furthermore, 2 studies used frameworks (Stirman and the intervention mapping protocol) [77,85], and 1 study [87] used a model (Stage Model) to adapt the intervention.

Tailoring

The majority of the studies tailored the eHealth interventions in various ways to the characteristics and skills of people with a low SES [23, 31-37, 40, 45, 46, 50, 58, 61, 62, 64, 66, 69-72, 74-77, 81-86, 88]. One method of customizing the eHealth intervention matched the content delivery (eg, visual or text information) to the user's language and digital literacy skills [34,50,54,61,64,66,69,70,75,80,84,86]. Another method tailored feedback, advice, and information to the characteristics (eg, cultural adaptations and practical advice relevant to their situation) of individuals with a low SES [23,31,33,35-37,39,40,45-47,58,64,66,74,75,77,81-83,85] or the timing and type of text messages (eg, feedback) [83]. However, it is unknown how tailoring was applied (technology or human tailoring, or a combination of both methods). A few studies based tailoring on theoretical models of behavioral change [31-33,46,70] and gathered information through questionnaires [31-33,36,40,47,55,58,62,66,85], self-monitoring data [23,37,46,55,77,83], or intervention goals [32,74]. However, tailoring the intervention system to deliver feedback or advice proved challenging as it required technological expertise and financial recourses [33,40]. It is unclear whether tailoring led to better results. Because of tailoring the feedback, 1 study showed that lower-educated smokers were more likely to revisit the intervention website [82].

Reach

The included studies applied multiple strategies to recruit low SES participants. However, some studies (n=5; 8%) provided limited details on the strategies they used to reach their participants [35,38,39,62,79]. The recruitment strategies and places are summarized in Table 2.

Different methods were found to be helpful to reach low SES participants. Lohse et al [52] found that Facebook is an effective tool to reach low-income women. Furthermore, the studies that used a personal approach to recruit participants reported a higher enrollment rate [61,82,86,87]. For example, Patten et al [87] found that reaching the targeted population with a face-to-face outreach method was more successful compared to recruitment through flyers. Kim et al [86] found that personal or telephone approaches to recruiting participants were responsible for most of the enrollments in their study. Moreover, participants indicated that they were more receptive to participating in a study when their doctor had previously discussed it with them [86]. Another study found that smokers recruited through general practitioners were more likely to be lower educated and already living with smoking-related illnesses than participants recruited through the internet [82].

Some studies experienced challenges in reaching low SES groups. These studies reached mainly medium or highly educated [33,44] participants with stable incomes and relatively healthy lifestyles [33]. van Dongen et al [44] indicated that people with a low SES may be reached with the right strategies, such as integrating an eHealth intervention into standard midwifery care, increasing awareness about the intervention's existence by expanding mass media use, and involving key community representatives of the target group. Additionally, some studies recommended increasing the reach of lifestyle interventions by collaborating with other experts, such as designers and health professionals [44,61,62]. Tables 3 and 4 show the barriers and facilitators for reach found in the studies.

Table 2. Participant recruitment, places, and strategies.

Recruitment characteristic	Number of studies
Individuals involved in recruitment	
Health professionals [30,34,42,44,74,78,81-85].	11
Researchers [37,40,58,61,81] and research assistants [38,40,80]	8
Study coordinators [86], managers [40,41], organization staff [68], and agent assistance [48]	5
Paraprofessionals [63,79] and volunteers [60,66]	4
Snowballing (*participants recruited other participants*) [58,65,67,82]	4
Recruitment places	
Health care setting [30,35,44,45,53,54,59,61,64,65,68,75,78,80,81,83-86]	19
Federal Benefit and Assistance Program for low-income women [38,42,49,50,56,72,73] and families [43,47,48,57,79]	12
Workplaces [40,77] and care services [34,60]	4
Local communities [59,66,68,71,87]	5
Food bank distributors [39,43,62]	3
Public health insurance [70]	1
Local nongovernmental organizations [51,60]	2
Public places [47,67]	2
Research agencies [31,71]	2
Recruitment strategies	
Online [23,30,32,37,40,43,44,46,47,51,52,55,58,59,61,65,68,73,74,76-78,82,84-87]	27
Newspaper advertisements [33,85], banners [44], flyers, and posters [45,47,50,57,59,61,66,68,77,78,83,85,87]	16
Personal contact (face-to-face) [33,40,45,58,65,67,68,74,80,83] or via phone [51,66]	12
Postal invitation letters [32,37,57]	3
Local television campaigns [32,82]	2
Regional health authority [33]	1

Table 3. Overview of facilitators identified in the eHealth interventions.

Facilitators per phase	Studies
Development	
Iterative design of the intervention (user-centered approach)	[40,46,58,59,61,62,65,66,68,69,74-77,85]
Study staff collaborating with other experts or a digital health company	[53,54,61,62,65,69,88]
Broad number of data sources to inform development	[65,75]
Participants’ knowledge of technology	[60,65]
Providing devices	[39,51]
Concise and clear content	[54,56-59,61,63,64,66,69,70,75]
Use of visual and multimedia elements	[23,30,40,47,56,59,60,63-66,69,70,73]
Resonating content of the intervention with participants	[53,60,61,66,85]
Evaluation	
Conducting formative evaluation in the early stage of the intervention	[40,41,46,55,58,61,62,65,66,68,70,74-76,85]
Reach	
Recruitment through Facebook, and active recruitment through health care professionals and tailored recruitment strategies	[44,47,52,65,82,87]
Collaborating with other experts, such as designers and health professionals, and local community services	[44,61,62,66,75,76]
Use	
Social support (friends, family, and peers)	[34,39,45,50,56,59,60,62,63,69,74]
Self-monitoring	[34,45,48,49,58,61,66,67]
Human coach can be helpful for participants	[30,47,49,50,56,57,87]
Practical advice to incorporate a healthy lifestyle in daily life	[54,56,58,60,63,64,66,73]
Reminders	[33,45,53,59,67,69]
Trust (eg, have a familiar face posting on a social media page) and credible information	[54,60,63,66,69]
Recipes and meal ideas may be useful	[47,54,66,69,73]
Helping participants with technology use	[57,61,66,86]
User friendliness and simplicity	[64,66,70,73]
Interactive features	[64-66,68,69]
Providing incentives and rewards (eg, virtual or financial rewards)	[35,48,53,58]
Links to more information	[64,66,69]
Combining social media with face-to-face group sessions	[47,56]
Networking with others encourages participants’ use of social media interventions	[56,63]
Activities must focus on pleasure and not obligation	[61,66]
Incorporating affordable options	[61]
Implementation	
Supplying the intervention through different platforms	[36]
Increasing direct communication with the health coach	[61]
Training health care professionals	[70]
Collaborating with health insurance	[70]
Server support staff and marketing team continually monitoring the intervention for technical issues	[76]

Table 4. Overview of barriers identified in the eHealth interventions.

Barriers per phase	Studies
Development	
Technical challenges with the intervention software or prototype	[53,58,60,62,70,71,78,86]
Amount of information or visuals	[40,56,59,63,66,68,85]
Limited financial resources for the intervention	[39,53,58,72]
Optimal frequency for reminders or messages	[33,59,68]
Evaluation	
Evaluation is time-consuming	[62]
Slow iterations of the intervention in the academic field	[65]
Reach	
Introductory study presentations and sending reminders to clinicians had a limited effect on increasing referrals	[86]
Passive recruitment (flyers)	[87]
Use	
Technical difficulties using a self-monitoring device or eHealth intervention (eg, lack of internet access, problems with telephones, and poor signal)	[34,40,45,49-51,56,58,60,61,65,66,68,70,71,73,85]
Limited digital skills of users and lack of knowledge of innovative technologies	[34,60,61,65,66,86]
Not wanting extra push notifications and lost notifications among all the notifications from other apps	[40,58,68]
Not allowed to carry a smartphone during work or does not carry a phone	[58,75]
Literacy and not mastering the language	[32,34,60,66,70]
Lack of time in a low SES[a] group	[34,45,49,54,56,60,61,68,69,73,78]
Financial problems (eg, paying bills)	[60,68,69]
Lack of familiarity with other participants before using social media and trust in social media or the internet	[45,60]
Waning participant interest toward the end of the intervention period and low motivation	[40,49,61]
Implementation	
Limited time of staff or coaches	[39,53,74]
Limited financial resources	[39,72]
Difficulties getting medical data of participants from participating health care facilities	[86]
Limited ability of peer coaches	[72]

[a] SES: socioeconomic status.

Use of eHealth Interventions

Most studies did not mention how the participants used the eHealth lifestyle interventions. However, many studies gained insight into the intervention usage by evaluating the concepts of adherence, user engagement, and acceptance [23,40,46,47,49,53,54,56,58,59,61,71,72,74,80,81]. Most of the studies showed that participants with a low SES accepted the eHealth interventions [40,46,47,53,54,56,58,71,74,80]. When there was high adherence, usage, and user engagement, interventions seemed effective [23,81].

Several studies mentioned explicitly measuring intervention usage with Google Analytics (eg, user interactions with content) [40], log data [23,47,54,61], registration data [44], emails sent, quiz questions accessed [44], questionnaires [39,40,44,64], or self-monitoring questionnaires [39]. The data analysis demonstrated that interventions were used as intended [23,61,81]. However, Régnier et al [60] found that the intervention was used to a less extent due to different barriers, such as technical issues, lack of language skills, and searching for real contact. In addition, Simons et al [40] reported decreased use during the intervention because of lesser engagement with the intervention. It also emerged that there was a difference between users within the interventions [33,40,81]. For example, in a study, it was found that the users who received notifications with tips, facts, and feedback mostly used the intervention [40]. Using periodic email prompts significantly increased the reuse of the intervention [33]. Schneider et al [33] concluded that it is crucial to develop strategies that encourage engagement from people with a low SES. Furthermore, hedonic elements (eg, visual elements) in the intervention were significantly associated with increased use [54]. In another study, personal and nuisance factors were associated with lower intervention use, including lower educational achievement and perceived barriers (eg, no time or interest and technical problems) [49]. Barriers and facilitators for using the interventions were diverse and varied in terms of individual and technological factors (Tables 3 and 4).

Other studies have analyzed adherence to interventions [31,35,49,87] by measuring the numbers of messages sent by participants, completion of coach calls [35,87], or intention to visit or revisit the intervention, or using specific features of the intervention, self-monitoring data, and days that participants used the intervention [31,49]. Adherence to the intervention decreased gradually in certain studies [35,49]. Griffin et al [48] showed that noncompleters of the intervention had certain characteristics (were younger, were African American, had a high BMI, had a lower education [high school or lower], and had a low income) when compared with participants who completed the intervention. Engagement with interventions was measured through the self-management behavior of participants [86], the tracking of their behavior via self-monitoring devices [86], self-reporting [56,64], and the presence of several likes, comments, and posts or messages assessed throughout the intervention delivery [47,56,59,72,86], as well as by capturing the frequency of user logins [86]. At the time when intervention engagement was high in several studies [59,61,72], in other studies, engagement decreased during the use of the intervention [45,47,51,71]. In 1 study, participants were more engaged with text messages than voice messages [34]. Another study showed that participant contributions appeared to vary across time of the day and day of the week (more active in the beginning part of the week and during the middle of the day) [47].

Delivery Mode of the Intervention

There was a wide variation in the delivery mode of the intervention. Table 1 provides an overview of the modes used to deliver the intervention. The studies cited several reasons for using a certain delivery mode. Using the internet [23,31,33,42,44,45,55,56,61,63,66,79,81,86], smartphone apps [58,60,73,75,78], or text messages [34,35,50,70,71,75] offers many benefits. Internet-based [36,42,44,45,52,56,64,86] and text-based [34,43,71,77,87] interventions are good channels for reaching hard-to-reach groups and might be effective in changing healthy behavior [32,34,44,47,48,50,54,59,66,70]. They also help to investigate new channels or to deliver interventions to low SES individuals [33,45,47,53,56,57,74,79]. Finally, low SES individuals use the internet, which provides the interventions an opportunity to reach this target group [39,42,45,52,56-58,62,73,86].

Studies reporting on the effectiveness of the delivery mode demonstrated no unequivocal results. Three studies showed that a web-based intervention was more effective for low SES participants [35,41,79] than non-eHealth interventions, such as in-person counselling for low SES participants [79] and game-based versions [41], and website users had deeper knowledge and a higher intention to use the website henceforth [41]. Another study [53] demonstrated that a gamified intervention significantly improved heart failure self-management knowledge in low SES and high SES participants. Participants with lower education levels and older adults preferred a digital game to any other medium for receiving information on self-management of heart failure. One RCT reported no interaction effects between delivery strategy (video versus text advice) and education level in terms of message processing mechanisms and future use of a smoking cessation intervention [82].

Implementation

Most studies did not adequately describe how their respective eHealth lifestyle interventions were implemented, perhaps because almost all the interventions were pilot projects and were not implemented in practice after the study was completed. The few barriers and facilitators that were identified are listed in Tables 3 and 4.

Several studies reported that they collaborated with different disciplines for implementation [55,61,70,72,77,85]. Golsteijn et al [85] created a network of hospitals and radiotherapy institutes to implement the intervention. However, their results on implementation are unknown. Furthermore, it appears that health professionals play an important role in the implementation of interventions [70].

Very few studies discussed the cost of their eHealth interventions. Limited finances and staff time presented several challenges in implementing these interventions [39,72]. Tabak et al [77] considered practicality and sustainability of the intervention by choosing interventions that prevented higher cost, for example, providing automatic feedback instead of individualized feedback. Other examples include finding enough coaches with the expertise to guide participants [72] and working with their time constraints [74]. Studies that reported on how they evaluated the implementation of their interventions were scarce. However, 1 study [85] used intervention mapping to develop an implementation plan. Two studies plan to evaluate implementation in the future. Whittemore et al [70] aimed to document an implementation analysis, and Foley et al [37] aimed to evaluate implementation through the RE-AIM framework.

Effectiveness

A number of studies (n=19) [23, 30, 32, 33, 34, 35, 36, 38, 39, 41, 43, 45, 46, 48, 50, 79-82] investigated the effectiveness of interventions for smoking cessation [23,46,51,81], healthy eating behaviors [32,39,79], alcohol [38,80], weight loss [35,37,45,47], physical activity [34,36,40], and multiple lifestyle changes [30,33,43,48,78]. Three studies [23,46,81] were effective in achieving smoking cessation in the low SES group. Furthermore, some studies reported significant improvements in eating behaviors [32,39,43,79], reduction in weight [35,43,48], and increase in physical activity [34,36,48]. Two RCTs showed that interventions were more effective in high SES participants than in low SES participants [32,36]. One study [32] found educational differences in high-energy snack intake. In this previous study, the plus group (environmental-level factors) received information on the availability and location of healthy food in the home environment and the prices of healthy food products in the supermarkets that the participants usually shop at. The plus approach targeted higher-educated participants more effectively than the basic approach, which was more effective for lower-educated participants. The authors argued that higher-educated participants understood and applied the environmental-level information easier than the lower-educated participants. The intervention as described by Golsteijn et al [36] resulted in a significant improvement in self-reported physical activity. However, the highly educated group initially participated more on the web than their lower-educated peers. In contrast to a study, they found minor effects in low SES participants, but no effect in participants with a higher SES [23]. The authors stated that this is likely because the user testing of the intervention was conducted exclusively with smokers with a low SES, which contributed to its effectiveness in the low SES group.

Two studies [33,45] reported minor significant improvements and modest effects on reuse of a healthy lifestyle program [33]. Other studies reported an insignificant effect [40] due to lower user engagement and dropouts.

Intervention Components

Studies applied diverse components within the interventions. For example, they employed visual and multimedia elements, such as images, infographics, videos, and social support. To a lesser extent, there was human or virtual coaching, and incentives were used. Table 5 presents an overview of the components in eHealth lifestyle interventions.

Table 5. Overview of the eHealth lifestyle intervention components (N=59).

Components	Studies, n (%)
Multimedia (images, infographs, and videos) [23,30,31,35-38,41,46,47,49,55-57,59-66,68,70,76,78,79,84-86]	30 (51)
Self-monitoring [23,30,34-37,40,43,45-49,51,56,58,60,61,65,67,70,76-78,83,84,85,86]	28 (47)
Tips [23,38-40,43-45,48,49,53,54,58,60-62,64,66,68,71,75-77,79,84-88]	28 (47)
Social support [33,34,36,37,40,43,45,47,49,53,55,58,61,62,67,69,71-74,77,84-87]	25 (42)
Reminders [23,30,32,33,36,37,40,45,46,48,49,51,53,55,58,59,65,67,69,78,84-86]	23 (39)
Rewards/incentives [32,34,35,38,41,45-48,50,51,53,56,64,68,73,74,76,79,80,83,87]	22 (37)
Coach [30,35,37,47,51,56,61,66,70,72,74,75,77,78,84-87]	18 (31)
Theoretical frameworks [23,31-34,37,43,46,48-50,55,64,69,70,73-76,80,81,83,85]	23 (39)

Theoretical Frameworks

Several studies (n=23, 39%) stated that they used one or more theoretical frameworks in their interventions [23,31-34,37,43,46,48-50,55,64,69,70,73-76,80,81,83,85]. The frameworks most commonly used were the social cognitive theory [34,37,43,48,64,70,73,85], I-Change Model [31,33,85], and theory of planned behavior [32,81,85], followed by the Health Belief Model [69,76,85], theories of self-regulation [32,85], and Precaution Adoption Process Model [32,85]. However, several studies mentioned using the Techniques of Behavior Change [23,45,46,67,77]; the theories for the rest of the studies can be found in Multimedia Appendix 3. Few studies used the frameworks to develop, adapt, evaluate, or implement the eHealth interventions [37,56,76,79]. It is unclear whether these theories were associated with desirable effects. Although not all studies have reported why they chose the theories [33,53,82], a few mentioned using the constructs or determinants of the theories [32,37,43,81,85], due to their suitability and available evidence [23,34,40,46,55,58,85]. Furthermore, it appeared that some interventions included components, such as self-monitoring, reminders, and social support based on behavioral strategies or theoretical frameworks, to promote lifestyle change or maintain healthy behaviors.

Multimedia and Visual Elements

Many studies included multimedia in their interventions, such as videos [35, 36, 37, 61, 79, 85, 86] or images [23, 38, 40, 46, 53, 55, 59, 62, 63, 89]. Although it was unclear why studies included these materials; some used videos [35-37,61,79,85,86] to introduce the intervention components [37,61], provide skills training [37,79], give home exercise instructions [36,85], or introduce the participants to their coach [86]. Other studies used visual materials, such as images and videos, to increase engagement [56,76]. Interventions applied images because of their visual appeal and ease to recall [55], or to enhance learning and motivate users to continue using the program [76]. There was almost no mention of using graphic artists [62,76] or photographers [76] to create illustrations for the interventions. However, Evans et al [62] stated that selecting illustrations for the app was challenging because matching the main text with illustrations was not always easy and required more iterations to meet the criteria. It was also challenging to find the right graphic artist to design proper images based on the given assignment. In 1 study [56], long videos resulted in lower engagement with Facebook participants. Another qualitative study [60] found that participants who experience language barriers rely more on visual materials (ie, videos) than written materials. However, a study reported that illustrations crowded with visual details confused participants [62]. Another study [63] highlighted that participants emphasized the importance of photos and visual appeal. In the study by Silfee et al [56], participants were more likely to read and comment on Facebook posts containing messages with images. One study [40] made it possible for the participants to see their daily steps via graphs. Although participants appreciated graphs, they used them significantly less at the end of the intervention due to decreased interest and outdated graph data. Only 1 study chose audio to increase the media on the website and facilitate relapse prevention and coping [55].

Social Support

Participants' peers [35,45,47,56,61,63,86] or significant others [34,72,74,87] provided social support, online or offline [35,45,56,61,63,71,86]. Other studies only gave advice on how people can get social support to help each other to adopt new behaviors [40,59,70,85,88]. Three studies mentioned that participants had positive experiences with the social support provided by their significant others [34] or peers [47,56,71] (they perceived a sense of community and social [71], emotional, and instrumental support [34]) and that peers motivated them [56]. For others, the ability to network and interact with peers was an important reason to visit the social media of the intervention [63]. However, it is difficult to determine whether social support contributed to the increased effectiveness of some interventions.

Several studies provided support through social media [35,45,56,61,63,86]. Participants were part of an online social network where they could, for example, discuss their goals [61,86] and challenges [86], and offer each other social support [35,47,61]. However, the studies identified different challenges in supporting active participation in the social support component, such as lack of a connection with other participants before accessing the eHealth intervention, limited engagement with other participants on social media [45,61], and not receiving timely responses from other participants [63]. Furthermore, in a qualitative study [61], participants experienced their level of literacy as an obstacle to taking part in online discussions, while in another study, posting about themselves made some participants with low SES uncomfortable [56], and others did not want to share their unsuccessful weight loss [56]. Involving support persons in the intervention appears to be complicated; some participants with low SES had no support person or did not want to involve one [34,74]. Furthermore, Pathak et al [75] showed that participants who had no family disliked messages that alluded to family support, and the term was replaced by loved ones (similar to familial relations). The interventions [45,56,63] offered many solutions to encourage the use of social support on social media, such as team-building exercises and enlisting friends [45]. The majority of participants of a smoking intervention relapsed, nonrelapsers reported significantly less temptation to smoke, and the qualitative data showed that nonrelapsers were able to manage temptation and reported greater support [50].

Self-monitoring

Several studies used few self-monitoring devices, based on emerging evidence or previous studies [37,85], such as pedometers [34,43,45,56,61,85,86] and weight scales [37,43,45,56,61,86]. Participants with a low SES monitored their diets digitally [30,60], with a calorie-counting book [45], or kept paper records [34,56]. Physical activity was also tracked through Fitbit devices [40,58,77] and MyFitnessPal [56]. Participants entered self-monitoring data [30,43], or this was done automatically [30,37,40,56,77,86]. Simons et al [40] found that continuous engagement with a self-monitoring device was challenging, due to participants not wearing the tracker or forgetting to charge it. Few studies provided information about the participants' experiences, or why the studies chose self-monitoring devices. However, some studies mentioned that participants found self-monitoring devices easy [40,61] and comfortable to use [40,58]. In 2 studies, participants struggled to use tracking devices [34,56], while in another study, participants desired digital apps for calorie counting [45]. It is difficult to determine whether self-monitoring led to increased effectiveness of the intervention. However, 1 study found that food photo journaling improved dietary choices more than having a health coach only [30].

Reminders

Sending reminders to participants was used by many studies; however, it is unclear in some studies how they applied the reminders in their interventions [32,45,53,71]. Two studies applied reminders to improve the adoption of and adherence to healthy behaviors [30,78] and to improve heart failure self-management skills [53]. Other studies applied reminders to encourage participants with low and high SES to visit or revisit the intervention [32,33,86], to remind users about their goals [58,85], and to remind users to submit their self-monitoring information [37]. Reminders were often used in the form of automatic emails [33,46,86,89], push messages via smartphones [40,58,78,86], text messages via mobile phones [37,59], and news updates [85]. The majority of studies did not report on how the participants evaluated the reminders. However, 2 studies showed that participants with a low SES found reminders helpful [53,58]. Furthermore, 2 studies indicated that participants had a greater need for reminders [45,59]. Some interventions that employed reminders appeared to be effective [32,33]. For example, in an RCT, reminders increased revisits to the intervention [33].

Coaches

Several studies included a coaching component in the intervention [30,35,37,47,51,56,61,66,70,72,74,75,77,78,84-87]. The coaches provided guidance mainly by telephone [30,35,37,61,72,74,75,77,78,84,87], followed by face-to-face counselling [30,37,61,74,78,87], text messages, email [30,61,66,78], online counselling [36,49,56,85,86], or combinations of these methods [30,37,61,78]. This was done through health professionals [30,36,37,72,85], researchers [51,56,74,75,77,87], parahealth professionals [49,72], and automatic phone [37,72]. The roles of the coaches varied and included guiding participants in setting goals [35,37], helping to solve problems [85], and providing behavioral skills training [37,56], and they also stimulated discussions on the online platforms of the interventions [56,86]. Interactions with the coaches varied from single, daily, or regular monthly contact [30,35,37,56,72,77,87] to ad hoc, based on needs [85].

Some coaches were experienced in behavioral change methods [30,35,37,56,76,78,87], and 3 coaches applied motivational interviewing [37,72,87]. It is difficult to determine whether coaching led to increased effectiveness of the intervention. However in 3 studies, the coaching component seemed promising [47,49,56]. The coaching component was positively associated with intervention usage [49] or higher engagement [47]. Furthermore, several studies reported that participants with a low SES appreciated the coaches [56,61,74,87]. Moreover, in 1 study, after the coach stopped engaging on social media (eg, posting and commenting), intervention engagement considerably decreased and passive engagement increased [56].

Incentives

Many studies [23, 32, 34, 35, 38, 41, 45, 52, 53, 56, 71, 76, 79, 87] offered participants incentives (eg, gift cards) [23, 32, 34, 35, 38, 41, 45, 52, 53, 56, 71, 76, 79, 87] for completing the assessments [32,34,38,41,45,52,56,71,79,87] to improve response rates [23,52], when submitting their saliva [35,87] or sending their self-monitoring data [35]. Nonetheless, it is unclear whether incentives delivered positive results. In fact, Radhakrishnan et al [53] found that the rewards and incentives offered in a game intervention did not match the real-time behavior, while in another study, participants suggested a greater frequency of incentives [45].

Tips

Providing practical information as tips was mentioned in several studies [23, 38-40, 43-45, 48, 49, 53, 54, 58, 60-62, 64, 66, 68, 71, 75-77, 79, 84-88]. Various studies chose this practical component based on theories [55,68,69,75]. Participants appreciated tips or found it useful to receive practical solutions as tips [54,56,58,66,68]. However, tips have to fit into the socioeconomic and sociocultural realities of people with a low SES [60,61,68]. It is unclear whether tips led to increased effectiveness of the intervention. However, Greene et al [54] found that intervention use was significantly higher among those who found the “Tip of the Day” motivating.

Discussion

Principal Findings

This scoping review provides an overview of the most commonly applied components in eHealth lifestyle interventions (development, reach, use, evaluation, and implementation) for people with a low SES. It also investigates the most common barriers and facilitators for current eHealth lifestyle interventions. The components that emerged can be classified into behavioral components (such as basic theoretical foundation, coaching, social support, reminders, self-monitoring, and incentives) and technological components (such as visual multimedia, reminders, and self-monitoring). Nevertheless, we found considerable heterogeneity in components, barriers, and facilitators, showing significant variation between studies. Moreover, we believe that the majority of barriers and facilitators for development and use are related to technology (eg, technical difficulties) and environmental factors (eg, financial resources of the intervention developers or target group). However, there was limited reporting about the barriers or facilitators within specific interventions, partly because many authors did not always share the lessons learned within their interventions. We should note that the barriers and facilitators may not be generalizable across different lifestyle behaviors, and few may apply to all SES groups and not only to eHealth interventions for low SES groups.

The studies examined the effectiveness of eHealth lifestyle interventions and showed promising but inconsistent results. They showed small effects of smoking cessation, nutrition, increased physical activity, and weight loss. These studies provided limited information about which components contributed to the effectiveness of the intervention, making it difficult to conclude why these interventions worked when compared to those that were unsuccessful. This is in line with the results of the systematic review by Kohl et al [90], which found that effect sizes were small, variable, and unsustainable in eHealth lifestyle interventions for different SES populations and concluded that the efficacy of intervention elements were unclear.

Different Delivery Methods

The results of this review suggest that eHealth lifestyle interventions delivered via different delivery modes (ie, websites, SMS text messages, or apps) or combined with professional personal support seem to be accepted by people with a low SES. However, it is still unclear which delivery method is the most effective for this target group because each delivery mode has its advantages. Danaher et al [91] and Iribarren et al [92] suggested that interventions delivered via text messages may be an attractive option as they are inexpensive, suitable for most mobile phones, and require little user effort. Conversely, interventions delivered via websites or apps provide a visually pleasing option (ie, videos) for communicating the information and make the intervention interactive. However, it is crucial to consider the digital literacy levels of people with a low SES when choosing the delivery method of an intervention. Blended care (combination of face-to-face services with eHealth) offers people with a low SES timely guidance, which can promote engagement and adherence to the intervention. Therefore, we suggest that combinations of varied eHealth delivery modes and face-to-face elements (ie, human coaching) could engage people with a low SES successfully.

Reaching the Low SES Group

Overall, it was clear from the studies that it was difficult to reach low SES individuals for participation in eHealth interventions, which is typical for this group. Thus, a different approach to reach this group is crucial. For example, studies have been successful in reaching participants with active recruiting strategies, such as face-to-face or personal contact [31]. The personal approach may reduce the distance between intervention staff and potential users, create a sense of security, and increase engagement [93]. Long-term relationships build trust between health professionals and patients, and such an approach is needed to reach people with a low SES [93,94]. Moreover, with this rapport, individuals may perceive health professionals as more credible, especially within ethnic minorities [94]. Another promising strategy is collaboration with the social network of people with a low SES (eg, caregivers, relatives, and experts) [44]. Recent studies identified the importance of using a personal approach and connecting via existing networks (ie, community centers or ambassadors) to successfully recruit low SES populations for lifestyle interventions [20,93]. Furthermore, we found that social media may achieve this goal since it has a broad reach, but the lack of robust evidence makes it difficult to draw firm conclusions. Social media may be particularly effective to reach young people. However, reaching people with a low SES remains challenging as there is no clear reach strategy. A similar pattern of results was obtained in the systematic review by Bonevski et al [95], which found that proven strategies to reach socially disadvantaged groups were rare. This highlights the importance of tailoring reach strategies, both online and offline, to target different types of groups (eg, young populations and ethnic minorities) within the low SES population. Lessons can also be drawn from traditional lifestyle interventions that provide insight into reaching low SES groups [96].

First Phase of Intervention Design and Co-creation

We noted that few studies based their interventions on behavioral theories. When behavioral theories were reported, authors rarely elaborated on how they applied these theories. These results seem to be consistent with other research that found that behavioral theories were seldom applied in interventions [88,97,98]. One possible explanation for this might be that intervention components are developed with a practical viewpoint in mind or a pre-existing belief in the benefit of these components, since they have been used previously in effective interventions [98] Alternatively, it may be that certain behavioral theories were not found to be useful for the development of the intervention at hand and were therefore not applied [99]. However, using theories in interventions has been indicated to increase their effectiveness.

There are several issues to consider in the co-creation of eHealth interventions, such as how and when stakeholders and users get involved. In recent years, more attention has been paid to the role of stakeholders (including users) in public interventions; however, involving stakeholders (eg, people with a low SES and health professionals) from the beginning is time-consuming and expensive [32]. Follow-up research needs to explore the best way to actively involve low SES individuals in developing and evaluating interventions, as co-creating with end users seems promising.

Implementation

The results of this review show that the development, evaluation, and implementation of eHealth are difficult to distinguish from each other and that the implementation of the intervention takes place during its development. As advocated by Pieterse et al [100], eHealth development and implementation should be intertwined. Implementation should be accounted for from the start of the development process; this is especially true for people with a low SES, since their characteristics, such as low digital skills, may hinder the interventions' implementation [61].

A shortage of resources is also known to impact implementation. These findings are directly in line with previous findings. For example, Lau et al [101] and Ross et al [102] found in their reviews that available resources, including time, funding, and staff, can be both barriers and facilitators in the implementation of interventions.

Recommendation for Design and Research, and Limitations

There are still unanswered questions in the development, reach, use, evaluation, and implementation of eHealth interventions for a low SES population, as the research is in its infancy. Using existing guidelines (eg, the CONSORT checklist) or other frameworks could guide in reporting information comprehensively and clearly [29]. For instance, use of the behavior change technique taxonomy by Michie et al [89] can help researchers to report on the behavioral theories and techniques applied in the intervention. Furthermore, it is vital to report more detailed information on how participants use the components of eHealth interventions, which may help identify elements that contribute to the effectiveness of eHealth interventions. This information could be beneficial for future studies and interventions as it can guide developers in the design and implementation of effective eHealth interventions. Another recommendation is to collaborate with researchers, developers, and stakeholders (including users) in the development, evaluation, and implementation of eHealth lifestyle interventions, to fine-tune these to the target group's needs and requirements. Involving the social networks (eg, relatives and peers) of low SES participants in eHealth lifestyle interventions also seems promising. Research shows that engaging social networks can support low SES participants who experience problems with their digital skills [60]. It is therefore important to investigate what role social networks should play within eHealth lifestyle interventions. Finally, although many studies advised making the content of eHealth interventions accessible to people with low skills, clear recommendations for developers and researchers on how eHealth interventions for low SES populations can be developed, implemented, and evaluated were lacking. Future research should focus on how we can devise holistic eHealth guidelines that can assist developers and researchers with the creation of eHealth interventions that take the capabilities and requirements of this target group into account.

This review is the first to focus on state-of-the-art available knowledge about developing and evaluating eHealth lifestyle interventions, and reaching people with a low SES to realize behavioral change and improve health in these people. The barriers and facilitators that we found offer promising elements that eHealth developers can use as a toolbox to connect eHealth with low SES target groups. Further research on the method of using these tools is still needed. However, this review has some limitations. First, we only included studies on eHealth interventions that focused on lifestyle behaviors and excluded studies on interventions aimed at other relevant areas for low SES individuals (mental health, and medical, legal, and financial issues). These interventions may provide additional insights. Second, as the primary focus was to gain insight into how eHealth lifestyle interventions are developed and evaluated for low SES individuals, we did not assess the quality of the studies and their results (ie, systematic review). Finally, we focused on the low SES group in general and did not distinguish between subgroups. Although ethnicity is not an indicator of SES, ethnic minorities (eg, non-Western immigrants and African American individuals) were often mentioned as prominent groups in the studies. It is therefore important to consider the differences within the low SES population, with the aim of not further increasing health disparities.

Conclusions

This scoping review provides an overview of the available scientific knowledge on the behavioral and technological components, barriers, and facilitators in the development, evaluation, and implementation of eHealth lifestyle interventions. Although eHealth intervention development is diverse, contributing to the varying results in this review, certain factors may be beneficial for building and using eHealth interventions and reaching people with a low SES. Iterative design of interventions, use of visual and multimedia elements, and social support seem to be important facilitators for eHealth interventions. Technical challenges using eHealth interventions, lack of time in low SES groups, and limited resources appear to be key barriers for eHealth interventions. Understanding these barriers and facilitators may generate insights into how to optimize eHealth interventions for people with a low SES. Developing eHealth interventions for people with a low SES requires consideration of their specific needs and characteristics, and the involvement of users. This may contribute to the use of interventions and may facilitate their implementation.

Guidelines should be developed to aid stakeholders in developing and evaluating eHealth interventions. Moreover, high-quality studies are needed to investigate how eHealth lifestyle interventions can be customized to meet the needs of participants with a low SES. Future studies could benefit significantly from detailed reporting on eHealth interventions for this target group.

Acknowledgments

We acknowledge support from Medical Delta (MD 2.0 program), ZonMw (The Netherlands Organization for Health Research and Development), and the Netherlands Cardiovascular Research Initiative, an initiative with support of the Dutch Heart Foundation (CVON2016-12 BENEFIT). We would like to thank the librarian Jan W Schoones at Leiden University Medical Centre (LUMC) for his contribution to the development of the search strategy. We would like to thank Mike Keesman, PhD, for his valuable inputs during the initial phase of the research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strings per database.
[DOC File , 252 KB - jmir_v24i8e34229_app1.doc]

Multimedia Appendix 2
Key constructs and definitions for data extraction.
[DOC File , 32 KB - jmir_v24i8e34229_app2.doc]

Multimedia Appendix 3
Overview of included publications (characteristic studies and theories).
[XLSX File (Microsoft Excel File), 24 KB - jmir_v24i8e34229_app3.xlsx]

References

1. The top 10 causes of death. World Health Organization. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2022-04-19]
2. Artinian NT, Fletcher GF, Mozaffarian D, Kris-Etherton P, Van Horn L, Lichtenstein AH, American Heart Association Prevention Committee of the Council on Cardiovascular Nursing. Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: a scientific statement from the American Heart Association. Circulation 2010 Jul 27;122(4):406-441 [FREE Full text] [doi: 10.1161/CIR.0b013e3181e8edf1] [Medline: 20625115]
3. Lawrence M, Kerr S, McVey C, Godwin J. The effectiveness of secondary prevention lifestyle interventions designed to change lifestyle behavior following stroke: summary of a systematic review. Int J Stroke 2012 Apr;7(3):243-247. [doi: 10.1111/j.1747-4949.2012.00771.x] [Medline: 22405279]
4. Brown T, Avenell A, Edmunds LD, Moore H, Whittaker V, Avery L, et al. Systematic review of long-term lifestyle interventions to prevent weight gain and morbidity in adults. Obes Rev 2009 Nov;10(6):627-638. [doi: 10.1111/j.1467-789X.2009.00641.x] [Medline: 19754634]
5. Stringhini S, Carmeli C, Jokela M, Avendaño M, Muennig P, Guida F, LIFEPATH consortium. Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1·7 million men and women. Lancet 2017 Mar 25;389(10075):1229-1237 [FREE Full text] [doi: 10.1016/S0140-6736(16)32380-7] [Medline: 28159391]

6. Di Girolamo C, Nusselder WJ, Bopp M, Brønnum-Hansen H, Costa G, Kovács K, et al. Progress in reducing inequalities in cardiovascular disease mortality in Europe. Heart 2020 Jan;106(1):40-49 [FREE Full text] [doi: 10.1136/heartjnl-2019-315129] [Medline: 31439656]
7. Michie S, Jochelson K, Markham WA, Bridle C. Low-income groups and behaviour change interventions: a review of intervention content, effectiveness and theoretical frameworks. J Epidemiol Community Health 2009 Aug;63(8):610-622. [doi: 10.1136/jech.2008.078725] [Medline: 19386612]
8. Mackenbach JP, Valverde JR, Bopp M, Brønnum-Hansen H, Deboosere P, Kalediene R, et al. Determinants of inequalities in life expectancy: an international comparative study of eight risk factors. Lancet Public Health 2019 Oct;4(10):e529-e537 [FREE Full text] [doi: 10.1016/S2468-2667(19)30147-1] [Medline: 31578987]
9. Macintyre A, Ferris D, Gonçalves B, Quinn N. What has economics got to do with it? The impact of socioeconomic factors on mental health and the case for collective action. Palgrave Commun 2018;4(1):10. [doi: 10.1057/s41599-018-0063-2]
10. Silva M, Loureiro A, Cardoso G. Social determinants of mental health: A review of the evidence. The European Journal of Psychiatry 2016;30(4):259-292.
11. Bull ER, Dombrowski SU, McCleary N, Johnston M. Are interventions for low-income groups effective in changing healthy eating, physical activity and smoking behaviours? A systematic review and meta-analysis. BMJ Open 2014 Nov 28;4(11):e006046 [FREE Full text] [doi: 10.1136/bmjopen-2014-006046] [Medline: 25432903]
12. Bukman AJ, Teuscher D, Feskens EJM, van Baak MA, Meershoek A, Renes RJ. Perceptions on healthy eating, physical activity and lifestyle advice: opportunities for adapting lifestyle interventions to individuals with low socioeconomic status. BMC Public Health 2014 Oct 04;14:1036 [FREE Full text] [doi: 10.1186/1471-2458-14-1036] [Medline: 25280579]
13. Coupe N, Cotterill S, Peters S. Tailoring lifestyle interventions to low socio-economic populations: a qualitative study. BMC Public Health 2018 Aug 03;18(1):967 [FREE Full text] [doi: 10.1186/s12889-018-5877-8] [Medline: 30075716]
14. Sheehy-Skeffington J. The effects of low socioeconomic status on decision-making processes. Curr Opin Psychol 2020 Jun;33:183-188. [doi: 10.1016/j.copsyc.2019.07.043] [Medline: 31494518]
15. Reinwand DA, Schulz DN, Crutzen R, Kremers SP, de Vries H. Who Follows eHealth Interventions as Recommended? A Study of Participants' Personal Characteristics From the Experimental Arm of a Randomized Controlled Trial. J Med Internet Res 2015 May 11;17(5):e115 [FREE Full text] [doi: 10.2196/jmir.3932] [Medline: 25963607]
16. Ossebaard HC, Van Gemert-Pijnen L. eHealth and quality in health care: implementation time. Int J Qual Health Care 2016 Jun;28(3):415-419. [doi: 10.1093/intqhc/mzw032] [Medline: 27029590]
17. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013 Sep;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
18. Olson CM. Behavioral Nutrition Interventions Using e- and m-Health Communication Technologies: A Narrative Review. Annu Rev Nutr 2016 Jul 17;36:647-664. [doi: 10.1146/annurev-nutr-071715-050815] [Medline: 27022772]
19. Crutzen R, van der Vaart R, Evers A, Bode C. Public health, behavioural medicine and eHealth technology. In: van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multidisciplinary Approach. London, UK: Routledge; 2018:111-127.
20. Faber JS, Al-Dhahir I, Reijnders T, Chavannes NH, Evers AWM, Kraal JJ, et al. Attitudes Toward Health, Healthcare, and eHealth of People With a Low Socioeconomic Status: A Community-Based Participatory Approach. Front Digit Health 2021 Jul 8;3:690182 [FREE Full text] [doi: 10.3389/fdgth.2021.690182] [Medline: 34713165]
21. 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]
22. Latulippe K, Hamel C, Giroux D. Social Health Inequalities and eHealth: A Literature Review With Qualitative Synthesis of Theoretical and Empirical Studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
23. Brown J, Michie S, Geraghty AWA, Yardley L, Gardner B, Shahab L, et al. Internet-based intervention for smoking cessation (StopAdvisor) in people with low and high socioeconomic status: a randomised controlled trial. Lancet Respir Med 2014 Dec;2(12):997-1006 [FREE Full text] [doi: 10.1016/S2213-2600(14)70195-X] [Medline: 25262458]
24. Hayba N, Partridge SR, Nour MM, Grech A, Allman Farinelli M. Effectiveness of lifestyle interventions for preventing harmful weight gain among young adults from lower socioeconomic status and ethnically diverse backgrounds: a systematic review. Obes Rev 2018 Mar;19(3):333-346. [doi: 10.1111/obr.12641] [Medline: 29178423]
25. Recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2022-04-20]
26. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
27. 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]

28. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM Planning and Evaluation Framework: Adapting to New Science and Practice With a 20-Year Review. Front Public Health 2019;7:64 [FREE Full text] [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
29. 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]
30. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health Coaching Reduces HbA1c in Type 2 Diabetic Patients From a Lower-Socioeconomic Status Community: A Randomized Controlled Trial. J Med Internet Res 2015 Oct 05;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
31. Stanczyk NE, Crutzen R, Bolman C, Muris J, de Vries H. Influence of delivery strategy on message-processing mechanisms and future adherence to a Dutch computer-tailored smoking cessation intervention. J Med Internet Res 2013 Feb 06;15(2):e28 [FREE Full text] [doi: 10.2196/jmir.2153] [Medline: 23388554]
32. Springvloet L, Lechner L, de Vries H, Candel MJJM, Oenema A. Short- and medium-term efficacy of a Web-based computer-tailored nutrition education intervention for adults including cognitive and environmental feedback: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e23 [FREE Full text] [doi: 10.2196/jmir.3837] [Medline: 25599828]
33. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012 Mar 01;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]
34. Ramirez M, Wu S. Phone Messaging to Prompt Physical Activity and Social Support Among Low-Income Latino Patients With Type 2 Diabetes: A Randomized Pilot Study. JMIR Diabetes 2017 Jun 06;2(1):e8 [FREE Full text] [doi: 10.2196/diabetes.7063] [Medline: 30291094]
35. Herring SJ, Cruice JF, Bennett GG, Davey A, Foster GD. Using technology to promote postpartum weight loss in urban, low-income mothers: a pilot randomized controlled trial. J Nutr Educ Behav 2014;46(6):610-615 [FREE Full text] [doi: 10.1016/j.jneb.2014.06.002] [Medline: 25069621]
36. Golsteijn RHJ, Bolman C, Peels DA, Volders E, de Vries H, Lechner L. A Web-Based and Print-Based Computer-Tailored Physical Activity Intervention for Prostate and Colorectal Cancer Survivors: A Comparison of User Characteristics and Intervention Use. J Med Internet Res 2017 Aug 23;19(8):e298 [FREE Full text] [doi: 10.2196/jmir.7838] [Medline: 28835353]
37. Foley P, Steinberg D, Levine E, Askew S, Batch BC, Puleo EM, et al. Track: A randomized controlled trial of a digital health obesity treatment intervention for medically vulnerable primary care patients. Contemp Clin Trials 2016 May;48:12-20 [FREE Full text] [doi: 10.1016/j.cct.2016.03.006] [Medline: 26995281]
38. Delrahim-Howlett K, Chambers CD, Clapp JD, Xu R, Duke K, Moyer RJ, et al. Web-based assessment and brief intervention for alcohol use in women of childbearing potential: a report of the primary findings. Alcohol Clin Exp Res 2011 Jul;35(7):1331-1338. [doi: 10.1111/j.1530-0277.2011.01469.x] [Medline: 21410488]
39. Clarke P, Evans SH, Neffa-Creech D. Mobile app increases vegetable-based preparations by low-income household cooks: a randomized controlled trial. Public Health Nutr 2019 Mar;22(4):714-725 [FREE Full text] [doi: 10.1017/S1368980018003117] [Medline: 30472970]
40. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. Effect and Process Evaluation of a Smartphone App to Promote an Active Lifestyle in Lower Educated Working Young Adults: Cluster Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10003 [FREE Full text] [doi: 10.2196/10003] [Medline: 30143477]
41. Silk KJ, Sherry J, Winn B, Keesecker N, Horodynski MA, Sayir A. Increasing nutrition literacy: testing the effectiveness of print, web site, and game modalities. J Nutr Educ Behav 2008;40(1):3-10. [doi: 10.1016/j.jneb.2007.08.012] [Medline: 18174098]
42. Silfee VJ, Lopez-Cepero A, Lemon SC, Estabrook B, Nguyen O, Rosal MC. Recruiting low-income postpartum women into two weight loss interventions: in-person versus Facebook delivery. Transl Behav Med 2019 Jan 01;9(1):129-134 [FREE Full text] [doi: 10.1093/tbm/iby013] [Medline: 29474726]
43. Griffin JB, Struempler B, Funderburk K, Parmer SM, Tran C, Wadsworth DD. My Quest, an Intervention Using Text Messaging to Improve Dietary and Physical Activity Behaviors and Promote Weight Loss in Low-Income Women. J Nutr Educ Behav 2018 Jan;50(1):11-18.e1. [doi: 10.1016/j.jneb.2017.09.007] [Medline: 29325657]
44. van Dongen JM, van Poppel MNM, Milder IEJ, van Oers HAM, Brug J. Exploring the reach and program use of Hello World, an email-based health promotion program for pregnant women in the Netherlands. BMC Res Notes 2012 Sep 22;5:514 [FREE Full text] [doi: 10.1186/1756-0500-5-514] [Medline: 22999052]
45. Cavallo DN, Sisneros JA, Ronay AA, Robbins CL, Jilcott Pitts SB, Keyserling TC, et al. Assessing the Feasibility of a Web-Based Weight Loss Intervention for Low-Income Women of Reproductive Age: A Pilot Study. JMIR Res Protoc 2016 Feb 26;5(1):e30 [FREE Full text] [doi: 10.2196/resprot.4865] [Medline: 26920252]
46. Brown J, Michie S, Geraghty AWA, Miller S, Yardley L, Gardner B, et al. A pilot study of StopAdvisor: a theory-based interactive internet-based smoking cessation intervention aimed across the social spectrum. Addict Behav 2012 Dec;37(12):1365-1370. [doi: 10.1016/j.addbeh.2012.05.016] [Medline: 22795643]

47. Cavallo DN, Martinez R, Webb Hooper M, Flocke S. Feasibility of a social media-based weight loss intervention designed for low-SES adults. Transl Behav Med 2021 Apr 26;11(4):981-992 [FREE Full text] [doi: 10.1093/tbm/ibaa070] [Medline: 32716040]
48. Griffin JB, Struempler B, Funderburk K, Parmer SM, Tran C, Wadsworth DD. My Quest, a Community-Based mHealth Intervention to Increase Physical Activity and Promote Weight Loss in Predominantly Rural-Dwelling, Low-Income, Alabama Women. Fam Community Health 2020;43(2):131-140. [doi: 10.1097/FCH.0000000000000251] [Medline: 32079969]
49. Lepore SJ, Collins BN, Killam HW, Barry B. Supportive Accountability and Mobile App Use in a Tobacco Control Intervention Targeting Low-Income Minority Mothers Who Smoke: Observational Study. JMIR Mhealth Uhealth 2021 Jul 02;9(7):e28175 [FREE Full text] [doi: 10.2196/28175] [Medline: 34255698]
50. Tagai EK, Miller SM, Belfiglio A, Xu J, Wen KY, Hernandez E. Persistent Barriers to Smoking Cessation Among Urban, Underserved Women: A Feasibility Study of Tailored Barriers Text Messages. Matern Child Health J 2020 Oct;24(10):1308-1317 [FREE Full text] [doi: 10.1007/s10995-020-02963-x] [Medline: 32557133]
51. Kendzor DE, Businelle MS, Waring JJC, Mathews AJ, Geller DW, Barton JM, et al. Automated Mobile Delivery of Financial Incentives for Smoking Cessation Among Socioeconomically Disadvantaged Adults: Feasibility Study. JMIR Mhealth Uhealth 2020 Apr 15;8(4):e15960 [FREE Full text] [doi: 10.2196/15960] [Medline: 32293568]
52. Lohse B. Facebook is an effective strategy to recruit low-income women to online nutrition education. J Nutr Educ Behav 2013;45(1):69-76. [doi: 10.1016/j.jneb.2012.06.006] [Medline: 23305805]
53. Radhakrishnan K, Toprac P, O'Hair M, Bias R, Kim MT, Bradley P, et al. Interactive Digital e-Health Game for Heart Failure Self-Management: A Feasibility Study. Games Health J 2016 Dec;5(6):366-374 [FREE Full text] [doi: 10.1089/g4h.2016.0038] [Medline: 27976955]
54. Greene EM, O'Brien EC, Kennelly MA, O'Brien OA, Lindsay KL, McAuliffe FM. Acceptability of the Pregnancy, Exercise, and Nutrition Research Study With Smartphone App Support (PEARS) and the Use of Mobile Health in a Mixed Lifestyle Intervention by Pregnant Obese and Overweight Women: Secondary Analysis of a Randomized Controlled Trial. JMIR Mhealth Uhealth 2021 May 12;9(5):e17189 [FREE Full text] [doi: 10.2196/17189] [Medline: 33978597]
55. Michie S, Brown J, Geraghty AWA, Miller S, Yardley L, Gardner B, et al. Development of StopAdvisor: A theory-based interactive internet-based smoking cessation intervention. Transl Behav Med 2012 Sep;2(3):263-275 [FREE Full text] [doi: 10.1007/s13142-012-0135-6] [Medline: 24073123]
56. Silfee VJ, Lopez-Cepero A, Lemon SC, Estabrook B, Nguyen O, Wang ML, et al. Adapting a Behavioral Weight Loss Intervention for Delivery via Facebook: A Pilot Series Among Low-Income Postpartum Women. JMIR Form Res 2018 Sep 10;2(2):e18 [FREE Full text] [doi: 10.2196/formative.9597] [Medline: 30684423]
57. Lohse B, Arnold K, Wamboldt P. Evaluation of About Being Active, an online lesson about physical activity shows that perception of being physically active is higher in eating competent low-income women. BMC Womens Health 2013 Mar 13;13:12 [FREE Full text] [doi: 10.1186/1472-6874-13-12] [Medline: 23496893]
58. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. A Smartphone App to Promote an Active Lifestyle in Lower-Educated Working Young Adults: Development, Usability, Acceptability, and Feasibility Study. JMIR Mhealth Uhealth 2018 Feb 20;6(2):e44 [FREE Full text] [doi: 10.2196/mhealth.8287] [Medline: 29463491]
59. Spears CA, Bell SA, Scarlett CA, Anderson NK, Cottrell-Daniels C, Lotfalian S, et al. Text Messaging to Enhance Mindfulness-Based Smoking Cessation Treatment: Program Development Through Qualitative Research. JMIR Mhealth Uhealth 2019 Jan 07;7(1):e11246 [FREE Full text] [doi: 10.2196/11246] [Medline: 30617043]
60. Régnier F, Dugré M, Darcel N, Adamiec C. Providing a Smart Healthy Diet for the Low-Income Population: Qualitative Study on the Usage and Perception of a Designed Cooking App. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e11176 [FREE Full text] [doi: 10.2196/11176] [Medline: 30470677]
61. Fontil V, McDermott K, Tieu L, Rios C, Gibson E, Sweet CC, et al. Adaptation and Feasibility Study of a Digital Health Program to Prevent Diabetes among Low-Income Patients: Results from a Partnership between a Digital Health Company and an Academic Research Team. J Diabetes Res 2016;2016:8472391 [FREE Full text] [doi: 10.1155/2016/8472391] [Medline: 27868070]
62. Evans SH, Clarke P. Resolving design issues in developing a nutrition app: A case study using formative research. Eval Program Plann 2019 Feb;72:97-105 [FREE Full text] [doi: 10.1016/j.evalprogplan.2018.10.010] [Medline: 30321688]
63. Leak TM, Benavente L, Goodell LS, Lassiter A, Jones L, Bowen S. EFNEP graduates' perspectives on social media to supplement nutrition education: focus group findings from active users. J Nutr Educ Behav 2014;46(3):203-208. [doi: 10.1016/j.jneb.2014.01.006] [Medline: 24613446]
64. Carolan-Olah M, Vasilevski V, Nagle C, Stepto N. Overview of a new eHealth intervention to promote healthy eating and exercise in pregnancy: Initial user responses and acceptability. Internet Interv 2021 Sep;25:100393 [FREE Full text] [doi: 10.1016/j.invent.2021.100393] [Medline: 34026565]
65. Bond MH, Bunge EL, Leykin Y, Barrera AZ, Wickham RE, Barlow MR, et al. Development and usability of a Spanish/English smoking cessation website: lessons learned. Mhealth 2021;7:30 [FREE Full text] [doi: 10.21037/mhealth-19-255] [Medline: 33898599]

66. 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]
67. Flaherty SJ, McCarthy MB, Collins AM, McCafferty C, McAuliffe FM. A phenomenological exploration of change towards healthier food purchasing behaviour in women from a lower socioeconomic background using a health app. Appetite 2020 Apr 01;147:104566. [doi: 10.1016/j.appet.2019.104566] [Medline: 31866355]
68. Kothari A, Godleski S, Abu B. Mobile-based consortium of parenting resources for low-income and underserved mothers and caregivers: app development, testing and lessons learned. Health and Technology 2020;10:1603-1608. [doi: 10.1007/s12553-020-00481-y]
69. 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]
70. Whittemore R, Vilar-Compte M, Burrola-Méndez S, Lozano-Marrufo A, Delvy R, Pardo-Carrillo M, et al. Development of a diabetes self-management + mHealth program: tailoring the intervention for a pilot study in a low-income setting in Mexico. Pilot Feasibility Stud 2020;6:25 [FREE Full text] [doi: 10.1186/s40814-020-0558-7] [Medline: 32082611]
71. Aldoory L, Yaros RA, Prado AA, Roberts E, Briones RL. Piloting Health Text Messages for Rural Low-Income Mothers: Effects of Source Similarity and Simple Action Steps. Health Promot Pract 2016 Sep;17(5):702-710. [doi: 10.1177/1524839915627457] [Medline: 26921364]
72. Athavale P, Thomas M, Delgadillo-Duenas AT, Leong K, Najmabadi A, Harleman E, et al. Linking High Risk Postpartum Women with a Technology Enabled Health Coaching Program to Reduce Diabetes Risk and Improve Wellbeing: Program Description, Case Studies, and Recommendations for Community Health Coaching Programs. J Diabetes Res 2016;2016:4353956 [FREE Full text] [doi: 10.1155/2016/4353956] [Medline: 27830157]
73. Garvin TM, Chiappone A, Boyd L, Stern K, Panichelli J, Edwards Hall LA, et al. Cooking Matters Mobile Application: a meal planning and preparation tool for low-income parents. Public Health Nutr 2019 Aug;22(12):2220-2227. [doi: 10.1017/S1368980019001101] [Medline: 31084663]
74. Mayberry LS, Berg CA, Harper KJ, Osborn CY. The Design, Usability, and Feasibility of a Family-Focused Diabetes Self-Care Support mHealth Intervention for Diverse, Low-Income Adults with Type 2 Diabetes. J Diabetes Res 2016;2016:7586385 [FREE Full text] [doi: 10.1155/2016/7586385] [Medline: 27891524]
75. Pathak LE, Aguilera A, Williams JJ, Lyles CR, Hernandez-Ramos R, Miramontes J, et al. Developing Messaging Content for a Physical Activity Smartphone App Tailored to Low-Income Patients: User-Centered Design and Crowdsourcing Approach. JMIR Mhealth Uhealth 2021 May 19;9(5):e21177 [FREE Full text] [doi: 10.2196/21177] [Medline: 34009130]
76. Stotz S, Lee JS. Development of an Online Smartphone-Based eLearning Nutrition Education Program for Low-Income Individuals. J Nutr Educ Behav 2018 Jan;50(1):90-95.e1. [doi: 10.1016/j.jneb.2016.12.008] [Medline: 29325667]
77. Tabak RG, Strickland JR, Stein RI, Dart H, Colditz GA, Kirk B, et al. Development of a scalable weight loss intervention for low-income workers through adaptation of interactive obesity treatment approach (iOTA). BMC Public Health 2018 Nov 16;18(1):1265 [FREE Full text] [doi: 10.1186/s12889-018-6176-0] [Medline: 30445939]
78. Wayne N, Ritvo P. Smartphone-enabled health coach intervention for people with diabetes from a modest socioeconomic strata community: single-arm longitudinal feasibility study. J Med Internet Res 2014 Jun 06;16(6):e149 [FREE Full text] [doi: 10.2196/jmir.3180] [Medline: 24907918]
79. Neuenschwander LM, Abbott A, Mobley AR. Comparison of a web-based vs in-person nutrition education program for low-income adults. J Acad Nutr Diet 2013 Jan;113(1):120-126. [doi: 10.1016/j.jand.2012.07.034] [Medline: 23092741]
80. Burner E, Zhang M, Terp S, Ford Bench K, Lee J, Lam CN, et al. Feasibility and Acceptability of a Text Message-Based Intervention to Reduce Overuse of Alcohol in Emergency Department Patients: Controlled Proof-of-Concept Trial. JMIR Mhealth Uhealth 2020 Jun 04;8(6):e17557 [FREE Full text] [doi: 10.2196/17557] [Medline: 32496203]
81. Brunette MF, Gunn W, Alvarez H, Finn PC, Geiger P, Ferron JC, et al. A pre-post pilot study of a brief, web-based intervention to engage disadvantaged smokers into cessation treatment. Addict Sci Clin Pract 2015 Feb 01;10:3 [FREE Full text] [doi: 10.1186/s13722-015-0026-5] [Medline: 25638283]
82. Stanczyk NE, Bolman C, Smit ES, Candel MJJM, Muris JWM, de Vries H. How to encourage smokers to participate in web-based computer-tailored smoking cessation programs: a comparison of different recruitment strategies. Health Educ Res 2014 Feb;29(1):23-40. [doi: 10.1093/her/cyt104] [Medline: 24287402]
83. Aguilera A, Figueroa CA, Hernandez-Ramos R, Sarkar U, Cemballi A, Gomez-Pathak L, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. BMJ Open 2020 Aug 20;10(8):e034723 [FREE Full text] [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]
84. van Dijk W, Oosterman M, Jansen I, de Vente W, Huizink A. Stress- and smoke free pregnancy study protocol: a randomized controlled trial of a personalized eHealth intervention including heart rate variability-biofeedback to support pregnant women quit smoking via stress reduction. BMC Public Health 2021 May 12;21(1):905 [FREE Full text] [doi: 10.1186/s12889-021-10910-w] [Medline: 33980201]

85. Golsteijn RHJ, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Development of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: OncoActive. BMC Cancer 2017 Jun 26;17(1):446 [FREE Full text] [doi: 10.1186/s12885-017-3397-z] [Medline: 28651586]
86. Kim SE, Castro Sweet CM, Gibson E, Madero EN, Rubino B, Morrison J, et al. Evaluation of a digital diabetes prevention program adapted for the Medicaid population: Study design and methods for a non-randomized, controlled trial. Contemp Clin Trials Commun 2018 Jun;10:161-168 [FREE Full text] [doi: 10.1016/j.conctc.2018.05.007] [Medline: 30035241]
87. Patten CA, Fu S, Vickerman K, Bock MJ, Nelson D, Zhu S, et al. Support person interventions to increase use of quitline services among racially diverse low-income smokers: A pilot study. Addict Behav Rep 2019 Jun;9:100171 [FREE Full text] [doi: 10.1016/j.abrep.2019.100171] [Medline: 31193750]
88. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
89. 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]
90. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
91. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: Outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 01;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
92. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
93. Stuber JM, Middel CNH, Mackenbach JD, Beulens JWJ, Lakerveld J. Successfully Recruiting Adults with a Low Socioeconomic Position into Community-Based Lifestyle Programs: A Qualitative Study on Expert Opinions. Int J Environ Res Public Health 2020 Apr 16;17(8):2764 [FREE Full text] [doi: 10.3390/ijerph17082764] [Medline: 32316344]
94. Bukman AJ, Teuscher D, Ben Meftah J, Groenenberg I, Crone MR, van Dijk S, et al. Exploring strategies to reach individuals of Turkish and Moroccan origin for health checks and lifestyle advice: a mixed-methods study. BMC Fam Pract 2016 Jul 21;17:85 [FREE Full text] [doi: 10.1186/s12875-016-0476-1] [Medline: 27439610]
95. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
96. van den Brand FA, Magnée T, de Haan-Bouma L, Barendregt C, Chavannes NH, van Schayck OCP, et al. Implementation of Financial Incentives for Successful Smoking Cessation in Real-Life Company Settings: A Qualitative Needs Assessment among Employers. Int J Environ Res Public Health 2019 Dec 16;16(24):5135 [FREE Full text] [doi: 10.3390/ijerph16245135] [Medline: 31888195]
97. Webb T, 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]
98. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris MF, IMPACT Study Group. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192 [FREE Full text] [doi: 10.1136/bmjopen-2017-019192] [Medline: 30158214]
99. Riley W, Rivera D. Methodologies for optimizing behavioral interventions: introduction to special section. Transl Behav Med 2014 Sep;4(3):234-237 [FREE Full text] [doi: 10.1007/s13142-014-0281-0] [Medline: 25264463]
100. Pieterse M, Kip H, Cruz-Martínez R. The complexity of eHealth implementation: A theoretical and practical perspective. In: van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multidisciplinary Approach. London, UK: Routledge; 2018:247-270.
101. Lau R, Stevenson F, Ong BN, Dziedzic K, Treweek S, Eldridge S, et al. Achieving change in primary care--causes of the evidence to practice gap: systematic reviews of reviews. Implement Sci 2016 Mar 22;11:40 [FREE Full text] [doi: 10.1186/s13012-016-0396-4] [Medline: 27001107]
102. 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]

Abbreviations

CeHRes: Center for eHealth Research
RCT: randomized controlled trial
RE-AIM: reach, effectiveness, adoption, implementation, and maintenance

SES: socioeconomic status

Edited by A Mavragani; submitted 13.10.21; peer-reviewed by H Mehdizadeh, G Porter, F Velayati, H Ayatollahi; comments to author 28.12.21; revised version received 24.03.22; accepted 31.03.22; published 24.08.22.

Please cite as:
Al-Dhahir I, Reijnders T, Faber JS, van den Berg-Emons RJ, Janssen VR, Kraaijenhagen RA, Visch VT, Chavannes NH, Evers AWM
The Barriers and Facilitators of eHealth-Based Lifestyle Intervention Programs for People With a Low Socioeconomic Status: Scoping Review
J Med Internet Res 2022;24(8):e34229
URL: https://www.jmir.org/2022/8/e34229
doi:10.2196/34229
PMID:36001380

Review

Randomized Controlled Trials of Artificial Intelligence in Clinical Practice: Systematic Review

Thomas Y T Lam[1,2*], PhD; Max F K Cheung[3*], BSc; Yasmin L Munro[3], MSc; Kong Meng Lim[3], MSc; Dennis Shung[4], MHS, MD; Joseph J Y Sung[3], MD, PhD

[1]The Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[2]Stanley Ho Big Data Decision Analytics Research Centre, The Chinese University of Hong Kong., Hong Kong, Hong Kong
[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[4]Department of Medicine (Digestive Diseases), Yale School of Medicine, New Haven, CT, United States
[*]these authors contributed equally

Corresponding Author:
Joseph J Y Sung, MD, PhD
Lee Kong Chian School of Medicine
Nanyang Technological University
Lee Kong Chian School of Medicine, Nanyang Technological University
11 Mandalay Road
Singapore, 308232
Singapore
Phone: 65 65138886
Fax: 65 63392889
Email: josephsung@ntu.edu.sg

Abstract

Background: The number of artificial intelligence (AI) studies in medicine has exponentially increased recently. However, there is no clear quantification of the clinical benefits of implementing AI-assisted tools in patient care.

Objective: This study aims to systematically review all published randomized controlled trials (RCTs) of AI-assisted tools to characterize their performance in clinical practice.

Methods: CINAHL, Cochrane Central, Embase, MEDLINE, and PubMed were searched to identify relevant RCTs published up to July 2021 and comparing the performance of AI-assisted tools with conventional clinical management without AI assistance. We evaluated the primary end points of each study to determine their clinical relevance. This systematic review was conducted following the updated PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines.

Results: Among the 11,839 articles retrieved, only 39 (0.33%) RCTs were included. These RCTs were conducted in an approximately equal distribution from North America, Europe, and Asia. AI-assisted tools were implemented in 13 different clinical specialties. Most RCTs were published in the field of gastroenterology, with 15 studies on AI-assisted endoscopy. Most RCTs studied biosignal-based AI-assisted tools, and a minority of RCTs studied AI-assisted tools drawn from clinical data. In 77% (30/39) of the RCTs, AI-assisted interventions outperformed usual clinical care, and clinically relevant outcomes improved with AI-assisted intervention in 70% (21/30) of the studies. Small sample size and single-center design limited the generalizability of these studies.

Conclusions: There is growing evidence supporting the implementation of AI-assisted tools in daily clinical practice; however, the number of available RCTs is limited and heterogeneous. More RCTs of AI-assisted tools integrated into clinical practice are needed to advance the role of AI in medicine.

Trial Registration: PROSPERO CRD42021286539; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=286539

(J Med Internet Res 2022;24(8):e37188) doi:10.2196/37188

KEYWORDS

artificial intelligence; randomized controlled trial; systematic review; clinical; gastroenterology; clinical informatics; mobile phone

Introduction

Background

Artificial intelligence (AI) was first described in the 1950s as a theory of human intelligence being exhibited by machines, including but not limited to learning, reasoning, and problem-solving [1]. With an exponential increase of computational power, reduced cost of data storage, improved algorithmic sophistication, and increased availability of health data from electronic health records, the era of AI has arrived in different specialties of medicine [2-4]. AI-assisted tools have been successfully applied in various clinical settings to assist diagnosis [5], improve therapy [6], and predict risk of mortality [7]. To date, 64 AI-powered medical devices and algorithms have been approved by the Food and Drug Administration in the United States [8].

The number of AI-related articles (using the Medical Subject Headings term, "artificial intelligence" as the search keyword) in the health care literature has increased dramatically from 6802 articles in 2016 to 21,160 in 2020. However, only a minority of these are prospective clinical studies, and there are few randomized controlled trials (RCTs). Several systematic reviews have been conducted to summarize the performance of recent AI-assisted tools in specific clinical settings, such as AI-assisted adenoma detection during colonoscopy [9], AI-assisted mammography in detecting breast cancer [10], AI-assisted intracranial hemorrhage recognition on computed tomography head imaging [11], AI-assisted glycemic control for patients with diabetes, and AI-assisted diagnosis of diabetes and its related complications [12]. A recent systematic review examined all studies of AI application in clinical practice, but was limited by restriction to English language and only searching full manuscripts published between January 2010 and May 2020 [13].

Objectives

To date, no systematic review has been restricted to RCTs regarding the clinical performance of AI-assisted tools in real-life practice. As RCTs represent the best clinical evidence to examine the effects of an intervention while controlling for unmeasured confounding factors, a comprehensive search of all RCTs studying AI-assisted tools in clinical practice would provide information regarding areas of opportunity for AI to affect real-world patient care [14]. We conducted a systematic review of all RCTs studying AI-assisted tools in clinical care.

Methods

Search Strategy

The systematic review was conducted following the updated PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [15]. We comprehensively searched CINAHL, Cochrane Central, Embase, MEDLINE, and PubMed from inception to July 14, 2021, to identify RCTs of AI-based tools across all medical specialties. Details of the full search strategy are provided in Multimedia Appendix 1. The search strategy included a combination of keywords and standardized Medical Subject Headings terms: "Artificial Intelligence," "Deep Learning," "Computer-Assisted Diagnosis," "Computer Assisted Diagnosis," "Computational Intelligence," "Computer Reasoning," "Computer Vision System," "Knowledge Acquisition," "Knowledge Representation," "Machine Intelligence" or "Machine Learning" or "Transfer Learning" or "Hierarchical Learning." The search was limited to RCTs. We also hand searched the references of the included studies to identify additional studies of interest. To include as many previous endeavors in this research area as possible, our search was not limited to peer-reviewed information. Conference abstracts and preprints were also included. The authors had no funding source for this study. This study was registered on PROSPERO (CRD42021286539).

Study Selection

After removing duplicates, two study authors (TYTL and MFKC) independently screened the title, abstract, and full text (if available) of each article to determine their eligibility. Unresolved disagreements were resolved by consulting the senior author (JJYS). Discrepancies were resolved by consensus. The complete manuscript was downloaded if the study met the inclusion criteria. We included studies that met the following inclusion criteria: (1) application of AI-assisted tools in clinical practice, which is defined as diagnosis, treatment, and prognostication on medical conditions that are seen and managed in daily clinical practice in hospitals or clinics. This does not include cellular or tissue cultures, animal studies, or experimental conditions such as induced cardiac arrhythmia and metabolic abnormalities. We classified the tool as AI-powered if the expressions, "artificial intelligence," "AI," "machine learning," "deep learning," "deep neural network," and "neural network" were used to describe the tool within the articles or other publicly available information resources; (2) patients or health care providers must be involved; (3) study design must be an RCT; and (4) control group must be without AI assistance. The exclusion criteria were as follows: (1) studies without implementation of clinical AI-assisted tools for patient management; (2) studies that were not conducted as original RCTs, for example, secondary analysis of a published RCT; and (3) clinical outcome not clearly defined. Reasons for exclusion were also recorded.

Data Extraction

After identifying relevant studies, the same two authors (TYTL and MFKC) independently extracted the data from each included study. Study design (racial information, sample size, RCT setting and design, and AI intervention and control) and AI-assisted tool characteristics (AI-assisted tool name, AI subtype, data type, and training and validation data) were documented. If AI development–related data were not available in the included articles, previously published articles of the same AI-assisted tool were reviewed to obtain the relevant information. Study end points (performance metrics used in primary and secondary end points) were listed. Clinically relevant end points were defined as whether the AI-assisted tools led to subsequent clinical interventions focusing on specific end points: (1) further diagnostic workup and investigation of the medical conditions, (2) changes in treatment strategy, (3) requirement of hospitalization, (4) escalation of care to the intensive care unit, and (5) influence on survival and mortality. Two independent researchers (TYTL and MFKC) resolved disagreements through discussion. If there were unresolved disagreements, consultation from senior author (JJYS) was sought.

Assessment of Risk of Bias

Risk of bias was assessed using the Cochrane risk-of-bias tool for randomized trials [16]. We specifically assessed the risk of bias of randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported results. The overall risk of bias was classified as low, some concerns, or high.

Results

Overview

The search performed on July 14, 2021, yielded 11,839 articles (n=2232, 18.85% from MEDLINE; n=1406, 11.88% from Embase; n=2264, 19.12% from PubMed; n=5229, 44.17% from Cochrane Central; and n=708, 5.98% from CINAHL); of these, 6823 (57.63%) were screened after removal of duplicates (n=5016, 42.37%). After screening the titles and abstracts, 6676 articles were excluded, because they did not fulfill the inclusion criteria. A total of 147 full manuscripts were individually assessed, of which 34 (23.1%) met the inclusion criteria. In addition, 4 more articles were identified by examining the references of the listed articles and manual searches (Figure 1). A total of 39 articles were included in this systematic review [6,17-54] as listed in Table 1.

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

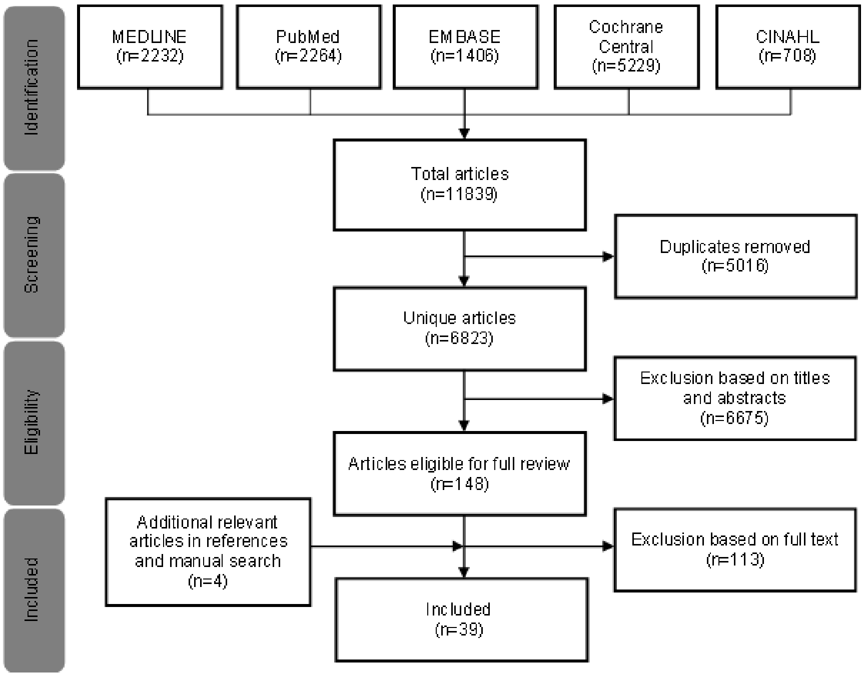


Table 1. Publications.

Author (publication year)	Title	Country	Article type	Specialty
El Solh et al [17], 2009	Predicting optimal CPAP by neural network reduces titration failure: a randomized study	The United States	Original article	Respiratory medicine
Shimabukuro et al [18], 2017	Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial	The United States	Original article	Anesthesiology
Labovitz et al [6], 2017	Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy	The United States	Original article	Neurology
Gracey et al [19], 2018	Improving medication adherence by better targeting interventions using artificial intelligence-a randomized control study	The United States	Abstract	Family medicine
Liu et al [20], 2018	Evaluating the impact of an integrated computer-based decision support with person-centered analytics for the management of hypertension: a randomized controlled trial	China	Abstract	Cardiology
Vennalaganti et al [21], 2018	Increased detection of Barrett’s esophagus–associated neoplasia using wide-area trans-epithelial sampling: a multicenter, prospective, randomized trial	The United States	Original article	Gastroenterology and hepatology
Biester et al [22], 2019	DREAM5: An open-label, randomized, cross-over study to evaluate the safety and efficacy of day and night closed-loop control by comparing the MD-Logic automated insulin delivery system to sensor augmented pump therapy in patients with type 1 diabetes at home	Germany, Israel, and Slovenia	Original article	Endocrinology, diabetes, and metabolism
Pouska et al [23], 2019	The use of HPI (Hypotension probability indicator) during major intracranial surgery; preliminary results of a prospective randomized trial	Czech Republic	Abstract	Anesthesiology
Lin et al [24], 2019	Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in Eye Clinics: a multicenter randomized controlled trial	China	Original article	Ophthalmology
Kamdar et al [25], 2019	A randomized controlled trial of a novel artificial intelligence–based smartphone application to optimize the management of cancer-related pain	The United States	Abstract	Clinical oncology
Persell et al [26], 2020	Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial	The United States	Original article	Cardiology
Voss et al [27], 2019	Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial	The United States	Original article	Psychiatry
Wang et al [28], 2019	Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study	China	Original article	Gastroenterology and hepatology
Wu et al [29], 2019	Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy	China	Original article	Gastroenterology and hepatology
Pavel et al [30], 2020	A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial	Ireland, the Netherlands, Sweden, and the United Kingdom	Original article	Neurology
Alfonsi et al [31], 2020	Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized control trial	Canada	Original article	Endocrinology, diabetes, and metabolism
Auloge et al [32], 2020	Augmented reality and artificial intelligence–based navigation during percutaneous vertebroplasty: a pilot randomised clinical trial	France	Original article	Orthopedics and traumatology
Avari et al [33], 2020	Safety and feasibility of the PEPPER adaptive bolus advisor and safety system; a randomized control study	The United Kingdom and Spain	Original article	Endocrinology, diabetes, and metabolism

Author (publication year)	Title	Country	Article type	Specialty
Chen et al [34], 2020	Comparing blind spots of unsedated ultrafine, sedated, and unsedated conventional gastroscopy with and without artificial intelligence: a prospective, single-blind, 3-parallel-group, randomized, single-center trial	China	Original article	Gastroenterology and hepatology
Gong et al [35], 2020	Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study	China	Original article	Gastroenterology and hepatology
Liu et al [36], 2020	The single-monitor trial: an embedded CADe system increased adenoma detection during colonoscopy: a prospective randomized study	China	Original article	Gastroenterology and hepatology
Nicolae et al [37], 2020	Conventional vs machine learning-based treatment planning in prostate brachytherapy: results of a phase I randomized controlled trial	Canada	Original article	Clinical oncology
Repici et al [38], 2020	Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial	Italy	Original article	Gastroenterology and hepatology
Su et al [39], 2020	Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos)	China	Original article	Gastroenterology and hepatology
Wang et al [40], 2020	Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study	China	Original article	Gastroenterology and hepatology
Wang et al [41], 2020	Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study	China	Original article	Gastroenterology and hepatology
Wijnberge et al [42], 2020	Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial	The Netherlands	Original article	Anesthesiology
Weisinger et al [43], 2021	Artificial intelligence-powered non-invasive and frequency-tuned electromagnetic field therapy improves upper extremity motor function in sub-acute stroke patients: a pilot randomized controlled trial	Israel	Abstract	Neurology
Blomberg et al [44], 2021	Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: a randomized clinical trial	Denmark	Original article	Emergency medicine
Browning et al [45], 2021	The clinical effectiveness of using a predictive algorithm to guide antidepressant treatment in primary care (PReDicT): an open-label, randomised controlled trial	The United Kingdom, Spain, Germany, France, and the Netherlands	Original article	Psychiatry
Jayakumar et al [46], 2021	Comparison of an artificial intelligence-enabled patient decision aid vs educational material on decision quality, shared decision-making, patient experience, and functional outcomes in adults with knee osteoarthritis: a randomized clinical trial	The United States	Original article	Orthopedics and traumatology
Kamba et al [47], 2021	A multicentre randomized controlled trial to verify the reducibility of adenoma miss rate of colonoscopy assisted with artificial intelligence–based software	Japan	Abstract	Gastroenterology and hepatology
Luo et al [48], 2021	Artificial intelligence-assisted colonoscopy for detection of colon polyps: a prospective, randomized cohort study	China	Original article	Gastroenterology and hepatology
Rafferty et al [49], 2021	A novel mobile app (Heali) for disease treatment in participants with irritable bowel syndrome: randomized controlled pilot trial	The United States	Original article	Gastroenterology and hepatology
Repici et al [50], 2021	Artificial intelligence and colonoscopy experience: lessons from two randomised trials	Italy and Switzerland	Original article	Gastroenterology and hepatology

Author (publication year)	Title	Country	Article type	Specialty
Strömblad et al [51], 2021	Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources: a randomized clinical trial	The United States	Original article	Surgery
Wu et al [52], 2021	Evaluating the effects of an artificial intelligence system on endoscopy quality and preliminarily testing its performance on detecting early gastric cancer: a randomized controlled trial	China	Original article	Gastroenterology and hepatology
Yao et al [53], 2021	Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial	The United States	Original article	Cardiology
Brown et al [54], 2021	Deep learning computer-aided polyp detection reduces adenoma miss rate: a US multi-center randomized tandem colonoscopy study (CADeT-CS trial)	The United States	Original article	Gastroenterology and hepatology

Study Characteristics

There were very few RCTs on AI-assisted medicine published until 2017. There was 1 RCT published in 2009, and the remaining 38 were published in the past 5 years (2 in 2017, 3 in 2018, 7 in 2019, 14 in 2020, and 12 in the first half of 2021; Figure 2).

These RCTs were conducted across 16 countries in North America, Europe, and Asia, with most of them conducted in the United States (13/39, 33%) and China (12/39, 31%). Furthermore, 18% (7/39) of the RCTs were published as conference abstracts only. Of these 39 publications, 16 (41%) were related to gastroenterology, whereas other specialties included anesthesiology (n=3, 7.7%), cardiology (n=3, 7.7%), endocrinology (n=3, 7.7%), psychiatry (n=2, 5%), neurology (n=3, 7.7%), orthopedics (n=2, 5%), oncology (n=2, 5%), surgery (n=1, 2.6%), ophthalmology (n=1, 2.6%), respiratory medicine (n=1, 2.6%), family medicine (n=1, 2.6%), and emergency medicine (n=1, 2.6%; Figure 3).

Figure 2. Number of randomized controlled trials of artificial intelligence–assisted medicine per year.

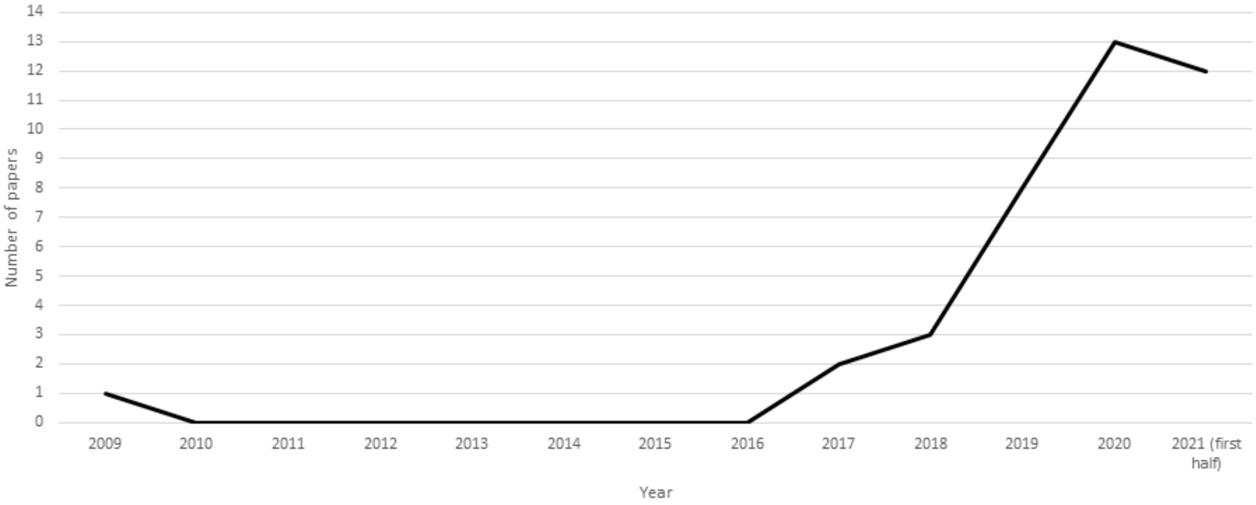


Figure 3. Distribution of original of publications and specialty.

Study Design

Multimedia Appendix 2 shows the study design and AI-assisted tool characteristics of each selected RCT. Most studies were single centered with a limited number of patients. Of these 39 studies, 35 (90%) had a sample size <1000 participants, 11 (28%) studies recruited fewer than 100 participants, 2 (5%) studies had a sample size of >1000 participants, and only 2 (5%) studies recruited >10,000 participants. More than half of the included RCTs (23/39, 59%) were conducted in a single center, 36% (14/39) of the studies were conducted across multiple centers, and 5% (2/39) of the studies did not mention how many centers were involved. A total of 16 open-label studies were conducted. Only 7 studies mentioned the racial information of the participants. A total of 13 blinded randomized trials were identified, of which 4 (31%) were double blinded and 9 (69%) were single blinded. The remaining 10 studies did not mention the level of blinding. Furthermore, 8 studies had a crossover study design. Most RCTs (36/39, 92%) compared the AI-assisted tools to control arms using the standard of care. Furthermore, 5% (2/39) of the studies used a sham treatment without AI assistance as the control group. A study used a mobile app without AI assistance as the control arm.

AI-Assisted Tool Characteristics

Biosignal-based AI tools are more common than clinical data–based tools. A total of 26 AI-assisted tools were biosignal based. Endoscopic images were the most commonly used biosignal (15/26, 58%). Furthermore, 50% (13/26) of the AI-assisted tools used clinical or biochemical data for analysis (patients' demography, self-administered questionnaire, and other relevant clinical data such as blood test results, blood pressure, and continuous positive airway pressure). No AI-assisted tool used both biosignal and clinical data combined as source data in the algorithm.

Most AI-assisted tools relied on static data (34/39, 87%) input to build the algorithm instead of dynamic data input (5/39, 13%). Static data refer to a snapshot of image or data of patients at a specific time point, whereas dynamic data are those captured continuously over a certain period during the study. For example, still images of the intestinal lumen captured during colonoscopy for AI-assisted adenoma detection are static data, whereas hourly captured vital signs and selected available laboratory tests for AI-assisted prediction of severe sepsis are dynamic data [18].

Approximately half of the studies (19/39, 49%) reported the AI-assisted tools development process. Of these, three AI-assisted tools in 8 studies, namely, GI-Genius, EndoScreener, and CC-Cruiser, were developed using data from multiple centers, whereas others were developed using data from a single center. A total of 35 studies reported the AI developer. Of these, 18 (51%) AI-assisted tools were developed by industry and 17 (49%) were developed by academic institutions.

Study End Points

Table 2 presents the study objectives and end points. Approximately half of the studies (18/39, 46%) used diagnostic accuracy as primary end point. The most common diagnostic end point is adenoma or polyp detection rate during colonoscopy. A total of 13 studies measured treatment response after AI-assisted intervention. Quality assurance of interventions was examined in 7 studies. End point measures of 27 studies were considered clinically relevant: 19 (70%) led to further investigation, 6 (22%) indicated the need for change in treatment, 1 (4%) reported in-hospital mortality and length of hospitalization, and 1 (4%) reported hospital admission.

Table 2. Study objectives and end points (primary and secondary).

Author (publication year)	Study objective	Primary end point	Secondary end point	Clinical relevance
El Solh et al [17], 2009	To test the effectiveness of an ANN[a] application for CPAP[b] titration on the time required to achieve an optimal CPAP pressure and CPAP titration failure	Time to optimal CPAP pressure	Titration failure rate	Change of treatment
Shimabukuro et al [18], 2017	To test the use of a machine learning–based severe sepsis prediction system for reductions in average length of stay and in-hospital mortality rate	Average hospital length of stay	In-hospital mortality rate; ICU[c] length of stay	Mortality and hospital and ICU length of stay
Labovitz et al [6], 2017	To evaluate the use of an artificial intelligence platform on mobile devices in measuring and increasing medication adherence in patients with stroke on anticoagulation therapy	Medication adherence	Nil	Nil
Gracey et al [19], 2018	To evaluate the effectiveness of using artificial intelligence to target which patients should receive interventions compared with traditional targeting approaches to improve medication adherence	Medication adherence	Nil	Nil
Liu et al [20], 2018	To assess the effects of clinical decision support system of graph-based machine learning algorithms on blood pressure management and economic burden of disease	Blood pressure reduction in patients with hypertension	Economic burden	Change of treatment
Vennalaganti et al [21], 2018	To evaluate the use of WATS[d] as an adjunct to biopsy sampling for the detection of HGD[e] or EAC[f] in a referral population with BE[g]	Rate of detection of HGD or EAC	Neoplasia detection rates based on the procedure order (WATS vs biopsy sampling first) of each procedure separately and the additional time required for WATS	Further investigation
Biester et al [22], 2019	To evaluate the safety and efficacy of 60-hour glucose control using the MD-Logic system in individuals with type 1 diabetes at home for day and night use, particularly without remote monitoring	Percentage of glucose sensor readings within 70 to 180 mg/dL (3.9-10 mmol/L)	Percentage of glucose sensor readings <60 to 70 mg/dL (3.3-3.9 mmol/L), percentage of glucose sensor readings >180 to 240 mg/dL (10-13.3 mmol/L), average and SD of glucose sensor readings, and overnight percentage of readings (“overnight” defined as 11:00 PM-7:00 AM) <70 mg/dL (3.9 mmol/L)	Further investigation
Pouska et al [23], 2019	To assess the use of HPI[h] to avoid hypotension in major intracranial surgery	Number of hypotension events; duration of hypotension events	Number of hypotension events in maintenance phase of anesthesia	Further investigation
Lin et al [24], 2019	To compare the diagnostic efficacy and treatment decision-making capacity between CC-Cruiser and ophthalmologists in real-world clinical settings	Accuracy of the diagnosis normal lens versus cataract	Evaluation of the disease severity; time required for making the diagnosis; patient satisfaction	Further investigation
Kamdar et al [25], 2019	To examine the impact of ePAL on cancer pain severity, attitudes toward cancer pain, and health care use	Pain severity	Attitudes toward cancer treatment (Barriers Questionnaire II); anxiety (General anxiety Disorder-7); pain-related hospital admissions	Hospitalization
Persell et al [26], 2020	To evaluate the effectiveness of an artificial intelligence smartphone coaching app to promote hypertension self-management	SBP[i] measured at 6 months	Self-reported medication adherence; home monitoring and self-management practices; self-efficacy related to BP[j] and BMI; self-reported health behaviors	Change of treatment

Author (publication year)	Study objective	Primary end point	Secondary end point	Clinical relevance
Voss et al [27], 2019	To test the efficacy of a wearable machine learning tool for intervention on a core ASD[k] deficit in the natural home environment	SRS-II[l] total score; Vineland Adaptive Behavioural Scales, Second edition; Developmental Neuropsychological Assessment, Second edition; Emotion Guessing Game	Moderator analysis; child behavior checklist; and the Vineland Adaptive Behavioural Scales, Second edition adaptive composite score	Nil
Wang et al [28], 2019	To investigate whether a high-performance real-time automatic polyp detection system can increase polyp and ADRs[m] in the real clinical setting	ADR	PDR[n]; mean number of polyps detected per colonoscopy; mean number of adenomas detected per colonoscopy; rate of false positives and false negatives	Further investigation
Wu et al [29], 2019	To evaluate the effectiveness of WISENSE to monitor blind spots, time the procedure, and automatically generate photo documentation during EGD[o] and thus raise the quality of everyday endoscopy	Blind spot rate	Inspection time; completeness of photo documentation generated by endoscopists; completeness of photo documentation generated by WISENSE in WISENSE group; completeness of photo documentation generated by WISENSE and endoscopists in WISENSE group; the percentage of patients being ignored in each site	Further investigation
Pavel et al [30], 2020	To evaluate the performance of the ANSeR[p] algorithm in real time by assessing the diagnostic accuracy for the detection of neonatal electrographic seizures with and without the use of ANSeR as a support tool for clinicians at the cot side	Diagnostic accuracy (sensitivity, specificity, and false detection rate) of health care professionals to identify neonates with electrographic seizures and seizure hours with and without the support of the ANSeR algorithm	Summary measures of seizure burden (total seizure burden, maximum hourly seizure burden, and median seizure duration); number of inappropriate antiseizure medications given	Change of treatment
Alfonsi et al [31], 2020	To test the app's usability and potential impact on carbohydrate counting accuracy	Carbohydrate counting accuracy	Quality of life for youth; self-care; patient or parent responsibility	Nil
Auloge et al [32], 2020	To evaluate technical feasibility, accuracy, safety, and patient radiation exposure granted by a novel navigational tool integrating augmented reality and artificial intelligence during percutaneous vertebroplasty of patients with vertebral compression fractures	Technical feasibility of trocar placement using augmented reality or artificial intelligence guidance	Comparison between groups A and B in terms of accuracy, procedural safety, time for trocar placement, and patient radiation exposure (dose area product and fluoroscopy time)	Nil
Avari et al [33], 2020	To evaluate the safety and efficacy of the PEPPER system compared with a standard bolus calculator	Difference in change in percentage time in range (3.9-10.0 mmol/L; 70-180 mg/dL) between the intervention arm that receives the PEPPER safety system with adaptive bolus advice and the control arm	Percentage time spent in euglycemia, hypoglycemia, and hyperglycemia; number of episodes of serious hypoglycemia; episodes of hypoglycemia within 5 hours postprandially; severe hypoglycemia (defined as a hypoglycemia event requiring third party assistance); postprandial mean area under the curve at 5 hours (expressed as mmol/L min); glycemia risk and variability measures	Change of treatment
Chen et al [34], 2020	To compare blind spots of sedated C-EGD[q], unsedated U-TOE[r], and unsedated C-EGD with and without the assistance of ENDOANG	The blind spot of 3 types of EGD with the assistance of ENDOANGEL	Blind spot rate of unsedated U-TOE and unsedated and sedated C-EGD with or without the assistance of ENDOANGEL; consistency between ENDOANGEL and endoscopists' review	Further investigation

Author (publication year)	Study objective	Primary end point	Secondary end point	Clinical relevance
Gong et al [35], 2020	To evaluate whether the ENDOANGEL system could improve polyp yield during colonoscopy	ADR	The ADR for adenomas of different sizes (diminutive [≤5 mm], small [>5 to<10 mm], and large [≥10 mm]); locations (cecum, ascending colon, transverse colon, descending colon, sigmoid colon, and rectum); PDR; PDR for polyps of different sizes; locations; mean number of adenomas per patient; mean number of polyps per patient; withdrawal time (time spent viewing as the endoscope is withdrawn during a colonoscopy, excluding biopsy or treatment time); adverse events and serious adverse events	Further investigation
Liu et al [36], 2020	To investigate whether the integration of a CADe[s] system into the primary monitor used during colonoscopy may increase polyp and adenoma detection without increasing physician fatigue	ADR	PDR; polyps per colonoscopy and adenomas per colonoscopy	Further investigation
Nicolae et al [37], 2020	To evaluate the noninferiority of day 30 dosimetry between a machine learning–based treatment planning system for prostate low-dose-rate brachytherapy and the conventional manual planning technique	The 1-month postoperative follow-up results between expert-planned low-dose-rate treatments (conventional) and the PIPA[t] machine learning approach	The efficiency of the PIPA approach in a standardized preoperatively planned workflow; total treatment planning time; need and extent of modifications	Nil
Repici et al [38], 2020	To assess the safety and efficacy of a CADe system in detection of colorectal neoplasias during real-time colonoscopy	ADR	Proximal ADR; total number of polyps detected; sessile serrated lesion detection rate; mean number of adenomas per colonoscopy; cecal intubation rate; withdrawal time	Further investigation
Su et al [39], 2020	To develop an automatic quality control system and assess whether it could improve polyp and adenoma detection in clinical practice	ADR	PDR; mean number of adenomas detected per colonoscopy; mean number of polyps detected per colonoscopy; withdrawal time (biopsy time was excluded by stopping the clock); adequate bowel preparation rate, defined as the percentage of colonoscopies with each segmental BBPS[u] score 2	Further investigation
Wang et al [40], 2020	To compare adenoma miss rates of CADe colonoscopy vs routine white-light colonoscopy.	Adenoma miss rate	Polyp miss rate; miss rate of advanced adenomas; sessile serrated adenoma or polyps; patient miss rate; ADR for the first pass; adenoma per colonoscopy; polyp per colonoscopy	Further investigation
Wang et al [41], 2020	To perform a double-blinded study using a sham control to more rigorously assess the effectiveness of a CADe system for improving detection of colon adenomas and polyps. We also aimed to analyze the characteristics of polyps missed by endoscopists	ADR	PDR; number of polyps per colonoscopy; number of adenomas per colonoscopy; sensitivity; specificity of the 3 skilled endoscopists	Further investigation
Wijnberge et al [42], 2020	To test whether the clinical application of the early warning system in combination with a hemodynamic diagnostic guidance and treatment protocol reduces intraoperative hypotension	Time-weighted average of hypotension during surgery	Incidence of hypotension (the number of hypotensive events per patient): total time with hypotension and percentage of time spent with hypotension during surgery; incidence of hypertension (the number of hypotensive events per patient): total time with hypertension and percentage of time spent with hypertension during surgery	Further investigation
Weisinger et al [43], 2021	To explore the benefit of BrainQ's novel and noninvasive, artificial intelligence–powered, frequency-tuned ELF-EMF[v] treatment (BQ) in improving upper extremity motor function in a population with subacute ischemic stroke	Fugl-Meyer Assessment-Upper Extremity score	Modified Rankin Scale; Action Research Arm Test; Box and Block Test; NIHSS[w]	Nil

Author (publication year)	Study objective	Primary end point	Secondary end point	Clinical relevance
Blomberg et al [44], 2021	To examine how a machine learning model trained to identify OHCA[x] and alert dispatchers during emergency calls affected OHCA recognition and response	Rate of dispatchers' recognition of subsequently confirmed OHCA	Dispatchers' time to recognition of OHCA; rate of DA-CPR[y]	Change of treatment
Browning et al [45], 2021	To assess the clinical effectiveness of using a predictive algorithm based on behavioral tests of affective cognition and subjective symptoms and to guide antidepressant treatment	Treatment response of depression symptoms	Change in anxiety scores at week 8 (measured using the Generalized Anxiety Disorder Assessment, 7 item version [30]); remission of depression at week 8 (defined as QIDS-SR-16[z] score of ≤5); change in the individual item scores from the QIDS-SR-16 measuring restlessness and sadness at week 8; change in symptoms of depression (treated as a continuous variable) across 12 months (measured using QIDS-SR-16); change in observer-reported symptoms of depression (treated as dichotomous response and as a continuous variable and measured using the MADRS[aa] at week 8); change in functional outcome across 12 months (measured using the SAS[ab] screener); patients also completed detailed health economic, acceptability, and cognitive functioning measures that will be reported separately	Nil
Jayakumar et al [46], 2021	To assess the effect of an artificial intelligence–enabled patient decision aid that includes education, preference assessment, and personalized outcome estimations (using patient-reported outcome measurements) on decision quality, patient experience, functional outcomes, and process-level outcomes among individuals with advanced knee osteoarthritis considering total knee replacement in comparison with education only	Decision process score of the knee decision quality instrument questions 3.1 to 3.5	Level of shared decision-making (assessed using the CollaboRATE survey); patient satisfaction with the consultation (numerical rating scale); condition-specific symptoms and functional limitations (Knee Injury and Osteoarthritis Outcome Score, Joint Replacement); duration of consultation in minutes; total knee replacement rates (proportion of patients undergoing surgery); treatment concordance (knee decision quality instrument question 1.6)	Nil
Kamba et al [47], 2021	To clarify whether adenoma miss rate could be reduced with the CADe assistance during screening and surveillance colonoscopy	Adenoma miss rate	Polyp miss rate; sessile serrated lesion miss rate; ADR	Further investigation
Luo et al [48], 2021	To explore whether artificial intelligence–assisted colonoscopy could improve the PDR in the actual clinical environment	PDR	Number of polyps detected; the number of diminutive polyps (diameter <6 mm); the number of polyps of each Paris type detected; the number of false positive results	Further investigation
Rafferty et al [49], 2021	To determine whether Heali, a novel artificial intelligence dietary mobile app can improve adherence to the LFD[ac], IBS[ad] symptom severity and quality of life outcomes in adults with IBS or IBS-like symptoms over a 4-week period	Adherence to the LFD	IBS symptom severity; quality of life outcomes	Nil
Repici et al [50], 2021	To assess the efficacy of a CADe system in detection of colorectal neoplasias in a nonexpert setting to challenge the CADe impact in a real-life scenario	ADR	Proximal ADR; total number of polyps detected; sessile serrated lesion detection rate; mean number of adenomas per colonoscopy; cecal intubation rate; withdrawal time	Further investigation
Strömblad et al [51], 2021	To assess accuracy and real-world outcome from implementation of a machine learning model that predicts surgical case duration	Accurate prediction of the duration of each scheduled surgery	Effects on patients and systems were measured by start time delay of following cases; time between cases; the time patients spent in presurgical area	Nil
Wu et al [52], 2021	To verify the effectiveness of ENDOANGEL in improving endoscopy quality and pretest its performance in detecting EGC[ae] in a multicenter randomized controlled trial	Number of blind spots	Performance of ENDOANGEL in predicting early gastric cancer in a clinical setting	Further investigation

Author (publication year)	Study objective	Primary end point	Secondary end point	Clinical relevance
Yao et al [53], 2021	To assess whether an ECG-based, artificial intelligence–powered clinical decision support tool enables early diagnosis of low EF[af], a condition that is underdiagnosed but treatable	Rate of newly diagnosed low EF, defined as EF≤50% within 90 days	Completion of an ECG within 90 days; other findings (eg, valvular heart disease), except low EF present on ECGs	Further investigation
Brown et al [54], 2021	To assess the comparative adenoma miss rate for CADe-assisted colonoscopy when compared with high-definition white light colonoscopy alone	Adenoma miss rate	Polyp miss rate; hyperplastic polyp miss rate; sessile serrated lesion miss rate; ADR; PDR; adenoma per colonoscopy; polyp per colonoscopy; sessile serrated lesion per colonoscopy	Further investigation

[a] ANN: artificial neural network.

[b] CPAP: continuous positive airway pressure.

[c] ICU: intensive care unit.

[d] WATS: wide-area transepithelial sampling.

[e] HGD: high-grade dysplasia.

[f] EAC: esophageal adenocarcinoma.

[g] BE: Barrett’s esophagus.

[h] HPI: hypotension probability indicator.

[i] SBP: systolic blood pressure.

[j] BP: blood pressure.

[k] ASD: autism spectrum disorder.

[l] SRS-II: Social Responsiveness Scale II.

[m] ADR: adenoma detection rate.

[n] PDR: polyp detection rate.

[o] EGD: esophagogastroduodenoscopy

[p] ANSeR: Algorithm for Neonatal Seizure Recognition.

[q] C-EGD: conventional esophagogastroduodenoscopy.

[r] U-TOE: ultrathin transoral endoscopy.

[s] CADe: computer-assisted detection.

[t] PIPA: prostate implant planning algorithm.

[u] BBPS: Boston Bowel Preparation Scale.

[v] ELF-EMF: extremely low frequency and low intensity electromagnetic fields.

[w] NIHSS: National Institutes of Health Stroke Scale.

[x] OHCA: out-of-hospital cardiac arrest.

[y] DA-CPR: dispatcher-assisted cardiopulmonary resuscitation.

[z] QIDS-SR-16: Quick Inventory of Depressive Symptomatology (16-Item) (Self-Report).

[aa] MADRS: Montgomery-Åsberg Depression Rating Scale.

[ab] SAS: Social Adjustment Scale.

[ac] LFD: low fermentable oligo-, di-, mono-saccharides and polyols diet.

[ad] IBS: irritable bowel syndrome.

[ae] ECG: electrocardiogram.

[af] EF: ejection fraction.

Study Outcomes

Table 3 shows the study results and limitations of each RCT. Of the 39 RCTs, 30 (77%) reported a positive study outcome where AI-assisted interventions outperformed the control arms. Of these 30 studies with positive outcomes, 22 (73%) AI-assisted interventions were biosignal based, and 8 (27%) studies used clinical data–based AI-assisted intervention for clinical outcome improvement. In addition, 21 of these 30 (70%) studies reported positive results of clinically relevant end points. Of these, 18 (86%) led to further investigations, 1 (5%) led to change in treatment, and 2 (9%) reduced the length of hospitalization.

Table 3. Study results and limitations.

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
El Solh et al [17], 2019	Time to optimal CPAP[a] pressure: AI[b] mean 198.7 (SD 143.8) minutes versus control mean 284.0 (SD 126.5) minutes	Titration failure: AI 16% versus control 36%; drop of residual obstructive apnea–hypopnea events and oxygen desaturations	Positive	No	Maximizing the time to achieve optimal CPAP and in reducing CPAP titration failure	Single center only; possible analysis bias as technologists were not blinded
Shimabukuro et al [18], 2017	AI 10.3 days versus control 13 days	In-hospital mortality: AI 8.96% versus control 21.3%; ICU[c] length of stay: AI 6.31 days versus control 8.40 days	Positive	No	Significant decrease in the hospital LOS[d] and in-hospital mortality	Small sample size; heterogenous population; trial was conducted in the 2 ICUs only; metrics were not monitored prospectively during the study because of the likely misrepresentation of such results; false positive rate, sensitivity, and prediction rate may be affected as clinicians may have initiated treatment before severe sepsis onset owing to advanced notice from the predictive algorithm; the use of overall metrics, LOS, and in-hospital mortality for all comers may underestimate the impact of the intervention on outcomes for patients with sepsis; potential for competing risks in the selected end points, mortality may shorten a patient's LOS; this study was patient-outcome oriented
Labovitz et al [6], 2017	Mean (SD) cumulative adherence based on pill count was 97.2 (4.4%) for the AI platform group and 90.6% (5.8%) for the control group. Plasma drug concentration levels indicated that adherence was 100% (15/15) and 50% (6/12) in the intervention and control groups, respectively	Nil	Positive	Unknown	Real-time monitoring has the potential to increase adherence and change behavior, particularly in patients on direct oral anticoagulant therapy	Not mentioned
Gracey et al [19], 2018	Likelihood of being adherent: AI>control, 6.11%; likelihood of being adherent: AI>traditional, 7.8%; no significant difference in likelihood of being adherent	Nil	Positive	Unknown	Using AI to target interventions can increase the effectiveness of medication adherence intervention programs	Not mentioned

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Liu et al [20], 2018	AI versus control: no significant difference	Economic burden of disease—AI versus control (all): no significant difference; economic burden of disease—AI: 46,006 (SD 40,831) yuan (US $6901 [SD 6125]) versus control (in surgical dept): 64,192 (SD 67,968) yuan (US $9629 [SD 10195])); benefit-cost ratio of AI: 1.15; net present value of benefit-cost of AI: 5792 yuan; direct medical costs—AI: 43,467 (SD 39.716) versus control: 61,205 (SD 66,576) yuan	Negative	Unknown	A clinical decision support system based on the graph-based machine learning algorithms changed the antihypertensive prescriptions and reduced the medical expense among patients with hypertension	Not mentioned
Vennalaganti et al [21], 2018	HGD[e] or EAC[f] detection—WATS[g] alone: 29 versus control alone: 7; AI (alone)>control (alone) 4.2 times	Neoplasia detection rates: not mentioned; average time required for WATS; additional time required for WATS: 11 minutes 26 seconds versus control: 6 minutes 55 seconds	Positive	Unknown	WATS increases the detection of HGD and EAC in a high-risk BE[h] surveillance population when used as an adjunct to biopsy sampling compared with biopsy sampling alone	Single center research only; potential of population bias as study population (20%) was enriched with patients with BE with a known history of dysplasia or referred for endoscopic therapy; no long-term follow-up
Biester et al [22], 2019	AI: 66.6% versus control: 59.9%	Percentage <60 mg/dL—AI: 0.64% versus control: 0.38%; percentage <70 mg/dL—AI: 2.31% versus control: 1.45%; percentage >180 mg/dL—AI: 28.32% versus control: 36.43%; percentage >240 mg/dL—AI: 8.53% versus control: 8.71%; Mean —AI: median (IQR) 153.11 (142.33-174.81) versus control: 163.84 (150.17-186.54); SD—AI: median (IQR) 52.71 (44.75-66.39) versus control: 54.95 (46.19-69.19)	Positive	No	The MD-Logic system was safe and associated with better glycemic control than SAP[i] therapy for day and night use. The absence of remote monitoring did not lead to safety signals in adapting basal rates nor in administration of automated bolus corrections	High rate of communication errors between the tablet computer running the algorithm and the insulin pump
Pouska et al [23], 2019	No significant difference in number of hypotension events between 2 groups (4/20 vs 2/20)	AI: 10 versus control: 4	Negative	Unknown	On the basis of our data, it seems that the inclusion of HPI[j] into a goal-directed treatment strategy could lower the incidence of hypotension within maintenance phase of anesthesia	Not mentioned

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Lin et al [24], 2019	Accuracy—AI: 87.4% versus control: 99.1%	No significant difference in evaluation of the disease severity between AI and control; AI: 2.79 minutes versus control: 8.53 minutes; rating of overall satisfaction—AI: mean 3.47 (SD 0.501) versus control: mean 3.38 (0.554)	Negative	No	CC-Cruiser exhibited less accuracy compared with senior human consultants in diagnosing childhood cataracts and making treatment decisions, but it has the capacity to assist human physicians in clinical practice in its current state	Patients without symptoms were less willing to participate in; patients with slightly opaque lens may have missed; CC-Cruiser provided treatment suggestions without considering the patients' general conditions; lack of internet accessibility limited the implementation of CC-Cruiser in low-income areas; possibly sufficient statistic power because cluster RCT[k] was adopted in trial, whereas RCT was used in sample size calculation
Kamdar et al [25], 2019	Difference of BPI[l] between AI and control: ß=−.09	Difference of BQ-II[m] between AI and control: ß=−.037; difference of General Anxiety Disorder-7 between AI and control: ß=.21; AI: 4 versus control: 20	Positive	Unknown	AI significantly decreases pain scores and pain-related hospitalizations in patients with cancer-related pain	Not mentioned
Persell et al [26], 2020	AI: mean systolic blood pressure (SD) 132.3 (15.0) mm Hg versus control: 135 (13.9) mm Hg	Significant improvement in self-reported medication adherence in AI group than control; no significant difference between home monitoring and self-management practices; AI group has 26.7 minutes per week (−5.4 to 58.8) more than control group in self-reported physical activity	Negative	Unknown	Adults with hypertension randomized to a coaching app plus home monitor had similar SBP[n] compared with controls receiving a tracking app and home monitor	Blinding to participants and research staff is impossible; some outcomes were self-reported; not specifically select participants who were likely to use a health-coaching app; small sample size; the app used in the study was a beta version; the AI and machine learning technology used here in this app gains information with larger numbers of users contributing data; cannot exclude the possibility that some patients may have well-controlled hypertension; limited generalizability because only iOS device users were recruited
Voss et al [27], 2019	SRS-II[o] showed large, not significant, positive mean changes in treatment participants; the VABS-II[p] socialization subscale score significantly increased between the start and end of the intervention in treatment-to-control comparisons	Moderator analyses showed a moderation effect for girls showing greater improvement; no significant changes from intake to posttest 1 were observed on Child Behaviour Checklist; the VABS-II adaptive composite score showed slightly greater improvement in younger participants	Positive	Yes	This study underscores the potential of digital home therapy to augment the standard of care	According to the poststudy empirical variance, this study may be underpowered by a factor of 2; low treatment adherence; bias in recruitment of participants; bias owing to the inherent demographic and behavioral heterogeneity of patients; second posttest appointments were not available for control participants before crossing over into treatment

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Wang et al [28], 2019	AI: 29.1% versus control: 20.3%	AI: 0.45 versus control: 0.29 (OR[q] 1.995, 95% CI 1.532-2.544); AI: 0.97 versus control: 0.51; AI: 0.53 versus control: 0.31; false positive rate of AI: 0.075 per colonoscopy; false negative rate of AI: not mentioned	Positive	No	In a low prevalent ADR[r] population, an automatic polyp detection system during colonoscopy resulted in a significant increase in the number of diminutive adenomas detected as well as an increase in the rate of hyperplastic polyps	Endoscopists were not blinded; lack of external validity; despite low false positive rates, potential distraction during the procedure could also be caused; fatigue level of participating endoscopists were not controlled; inadequate sample size of colonoscopies performed by junior endoscopists; only Olympus colonoscopy equipment was used
Wu et al [29], 2019	AI: 5.86% versus control: 22.46%	AI: 5.03 minutes versus control: 4.24 minutes; AI: 71.87% versus control: 79.14%; AI: 90.64% versus control: 79.14%; AI: 92.91% versus control: 79.14%; percentage of patients being ignored in majority gastric sites were significantly lower than control	Positive	No	WISENSE greatly reduced blind spot rate, increased inspection time, and improved the completeness of photo documentation	Only Olympus and Fujifilm endoscopes were used in this trial; the withdrawal time in this trial was generally less than recommended 7 minutes of EGD[s] in the guideline
Pavel et al [30], 2020	Diagnostic accuracy (sensitivity, specificity, and false detection rate) for recognition of a neonate with seizures were not significantly different between the 2 groups; sensitivity of seizure hours—AI: 66% versus control: 45.3%; false detection rate of seizure hours was not mentioned	No significant differences found in seizure characteristics; AI: 37.5% versus control: 31.6%; difference 5.9%	Negative	No	In conclusion, this clinical investigation was the first to assess the performance of a machine learning algorithm for neonatal seizure detection in real time and in the real-world setting of busy neonatal ICUs throughout Europe	Excluded seizures with a duration of <30 seconds from both groups; analysis was done using seizure hour instead of looking at each individual seizure
Alfonsi et al [31], 2020	Absolute error at 3-month follow-up—AI: 27.45% (10.90%) versus control: 38.00% (14.74%); error>10 g at 3-month follow-up—AI: 21.43% (16.82%) versus control: 32.27% (16.31%)	No significant difference between groups A and B in terms of accuracy, procedural safety, time for trocar placement, and patient radiation exposure (dose area product and fluoroscopy time)	Positive	No	The data suggest that use of iSpy is associated with improved carbohydrate counting and that usability and acceptability of the app is quite positive	Single tertiary pediatric center only; the number of foods recognized by iSpy is not all encompassing; detailed information about other factors that can influence care such as education level, socioeconomic status data, family dynamics, or details of treatment regimen were not acquired; text reminders to the control participants was not provided

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Auloge et al [32], 2020	Group A technical feasibility was 100% with successful segmentation and generation of safe or accurate trajectory in all cases	No significant difference in accuracy; no complications or unintended effects were observed in either group—AI: mean 642 (SD 210) seconds, range 300-963 versus control: mean 336 (SD 60) seconds, range 240-438; DAP[t]—AI: mean 182.6 (SD 106.7) mGy cm^2, range 27-355 versus control: mean 367.8 (SD 184.7) mGy cm^2, range 115-644; fluoroscopy time—AI: 5.2 (SD 2.6) seconds, range 1.6-8.7 versus control: mean 10.4 (SD 4.1) seconds, range 4.2-17.9	Positive	Yes	Augmented reality or AI–guided percutaneous vertebroplasty appears feasible, accurate and safe and facilitates lower patient radiation exposure compared with standard fluoroscopic guidance	Small sample size; surgeon bias due to inherent nonblinding; lack of power to assess differences in vertebroplasty complication rates; accuracy of final trocar position was estimated on augmented fluoroscopic images rather than CBCT[u]; no clinical follow-up is presented
Avari et al [33], 2020	AI: 62.5% (52.1%-67.8%) versus control: 58.4% (49.6%-64.3%)	No significant difference for percentage of time in euglycemia, hypoglycemia, and hyperglycemia; no episode of serious hypoglycemia; no episodes of hypoglycemia within 5 hours postprandially; case of severe hypoglycemia; AI: 0 versus control: 1; no significant difference in glycemic risk	Negative	Yes	The PEPPER system was safe but did not change glycemic outcomes compared with control	The potential need for additional time required for the adaptive insulin recommender system to be effective; the algorithm is likely to be most beneficial to individuals maintaining regular work patterns rather than shift workers; the algorithm only adapts for bolus insulin and assumes that the basal insulin has been optimized; the system is dependent on meal scenarios where the user has not ingested a significant snack or taken an insulin bolus correction within 5 hours of a meal for revision
Chen et al [34], 2020	Sedated C-EGD versus unsedated U-TOE[v] versus unsedated C-EGD: 3.42% (0.89/26) versus 21.77% (5.66/26) versus 31.23% (8.12/26), respectively	Blind spot rate of Sedated C-EGD—AI: 3.42 versus control: 22.46; blind spot rate of unsedated U-TOE—AI: 21.77 versus control: 29.92; blind spot rate of unsedated C-EGD—AI: 31.23 versus control: 42.46; the average accuracy, sensitivity, and specificity of EN-DOANGEL in sedated C-EGD were 88.3%, 92.6%, and 90.2%, respectively; in unsedated U-TOE, were 91.3%, 84.5%, and 90.1%, respectively; and in unsedated C-EGD, were 87.8%,8 2.8%, and 87.8%, respectively	Positive	No	In summary, our study showed that the number of blind spots in conventional sedated EGD was the lowest compared with unseated U-TOE and unsedated EGD, and the addition of ENDOANGEL had a maximal effect on unsedated C-EGD	Single-center study; endoscopist were not blinded

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Gong et al [35], 2020	ITT[w]—AI: 16% (58/355) versus control: 8% (27/349); PP[x]—AI: 17% (54/224) versus control: 8% (26/318)	ITT diminutive—AI: 46 (13%) versus control: 25 (7%); ITT small—AI: 4 (1%) versus control: 1 (<1%); ITT large—AI: 10 (3%) versus control: 1 (<1%); no significant differences were found comparing adenoma locations: ITT—AI: 47% (166/355) versus control: 34% (118/349); ITT diminutive—AI: 158 (45%) versus control: 114 (33%); ITT small—AI: 9 (3%) versus control: 7 (2%); ITT large—AI: 11 (3%) versus control: 3 (1%); significant different was only found in sigmoid colon—AI: 79 (22%) versus control: 48 (14%); AI: 0.18 versus control: 0.08; AI: 1.17 versus control: 0.68; AI: 6.38 minutes versus control: 4.76 minutes; no adverse and serious adverse events	Positive	No	In conclusion, the ENDOANGEL system is a quality improving system for colonoscopy that uses computer vision, real-time monitoring of withdrawal speed, and timing of colonoscopy intubation and withdrawal and provides reminders to endoscopists of blind spots, in addition to live tracking previously seen frames during colonoscopy	AI was validated at 1 center only; the withdrawal speed was artificially divided into safe, alarm, and dangerous by assessing videos from Renmin Hospital of Wuhan University; the difference between assisted and unassisted colonoscopy for adenomas of 6-9 mm was not significant, which could be attributable to small numbers
Liu et al [36], 2020	AI: 29.01% versus control: 20.91%	AI: 47.07% versus control: 33.25%; AI: 1.07 versus control: 0.51; AI: 0.48 versus control: 0.29	Positive	No	In conclusion, real-time visual alarms provided by a high-performance CADe[y] system embedded into the primary colonoscopy monitor, with nearly unnoticeable latency, have been shown to cause a significant improvement in ADR because of an increased detection of diminutive adenomas without increasing physician fatigue level during colonoscopy	Open-labeled study; the fatigue score was subjective and susceptible to factors other than the visual alarms; whether a polyp was first detected by CADe before the endoscopist was based on the operating endoscopist's own judgment; the fact that the CADe system detected a polyp before the endoscopists does not necessarily mean that the endoscopists would have missed that lesion
Nicolae et al [37], 2020	No significant difference in CTV[z] V_{100}, CTV D_{90}, and Rectum V_{100} at 1-month postoperative follow-up	AI: mean 2.38 (SD 0.96) minutes versus control: mean 43.13 (SD 58.70) minutes; no significant difference in need and extent of modifications	Positive	Yes	A machine learning–based planning workflow for prostate LDR[aa] brachytherapy has the potential to offer significant time savings and operational efficiencies, while producing noninferior postoperative dosimetry to that of expert, conventional treatment planners	Single-center study; examining only preoperatively planned cases

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Repici et al [38], 2020	AI: 54.8% versus control: 40.4%	AI: 123 versus control: 97; AI: 353 out of 262 patients versus control: 243 out of 198 patients; AI: 7% versus control: 5.2%; AI: 1.07 versus control: 0.71; AI: 95.6% versus control: 98.5%; withdrawal time: not mentioned	Positive	No	In a multicenter, randomized trial, we found that including CADe in real-time colonoscopy significantly increases ADR and adenomas detected per colonoscopy without increasing withdrawal time	Psychological bias could not be excluded; the equivalence in withdrawal time excludes a somewhat reduced degree of mucosal exposure in the control arm; low detectors and inexperienced or nongastroenterologist endoscopists were not involved in this study
Su et al [39], 2020	AI: 28.90% versus control: 16.51% (OR 2.055, 95% CI 1.397-3.024; *P*<.001)	AI: 38.31% versus control: 25.40%; AI: 0.367 versus control: 0.178; AI: 0.575 versus control: 0.305; AI: mean 7.03 (SD 1.01) minutes versus control: mean 5.68 (SD 1.26) minutes; AI: 87.34% versus control: 80.63%	Positive	No	In summary, AQCS[ab], an automatic quality control system, could be used in real time for timing, supervising withdrawal stability, evaluating BBPS[ac], and detecting polyp	Single endoscopic center; some false prompts occurred with the AQCS; fatigue level of participating physicians was not controlled; used 4 intraprocedural quality metrics to form the AQCS, without performing preliminary testing to evaluate whether just 2 or 3 or 4 of these metrics had the same quality improvement; did not test the sole effect of colonoscopy stability; the DCNNs[ad] were trained only on images obtained from a Pentax imaging system
Wang et al [40], 2020	AI: 13.89% versus control: 40%	AI: 12.98% versus control: 45.90%; no statistical differences in the miss rate of advanced adenomas and sessile serrated adenoma or polyps; no significant difference in patient miss rate; no significant difference in ADR for the first pass; no significant difference in adenoma per colonoscopy; no significant difference in polyp per colonoscopy	Positive	No	The results from this study suggest a significantly lower AMR[ae] when a CADe technology is used compared with routine white light colonoscopy. The detection of diminutive and small adenomas with nonadvanced histology and nonpedunculated shape could be effectively improved by CADe colonoscopy	AMR obtained in the tandem study cannot reflect the absolute miss rate; subjective bias in open-labeled trial; tandem colonoscopy in each patient was performed by the same endoscopist; study population was not restricted to screening-only participants according to guidelines; only skilled endoscopists were allowed to participate in this study; subjected bias may be introduced as the judgments made by the panel of 3 experts who reviewed the video record were not a gold standard as pathology
Wang et al [41], 2020	AI: 165 (34%) versus sham control: 132 (28%)	AI: 252 (52%) versus sham control: 176 (37%); AI: 1.04 versus sham control: 0.64; AI: 0.58 versus sham control: 0.38	Positive	No	The CADe system is a safe and effective method to increase ADR during colonoscopy	Potential bias in the presence of a second senior endoscopist; bias in patient recruitment; the actual alert numbers of the sham system should have been measured in the trial to show equivalence

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Wijnberge et al [42], 2020	AI: median 0.10, IQR 0.01-0.43 mm Hg versus control: median 0.44, IQR 0.23-0.72 mm Hg	AI: 3.00, IQR 1.00-8.00 versus control: 8.00, IQR 3.50-12.00; AI: 8.00, IQR 1.33-26.00 minutes versus control: 32.67, IQR 11.50-59.67 minutes; AI: 2.8%, IQR 0.8%-6.6% versus control: 5.6%, IQR 3%-9.4%; AI: 2.0 (0.0 to 3.0) versus control: 0.0 (−1.0 to 0.0); AI: 4.0 (0.0 to 10.7) minutes versus control: −0.7 (−4.3 to 0.7) minutes; AI: 1.5% (0.0 to 3.3) versus control: −0.2% (−1.4 to 0.3)	Positive	No	In this single-center preliminary study of patients undergoing elective noncardiac surgery, the use of a machine learning–derived early warning system compared with standard care resulted in less intraoperative hypotension. Further research with larger study populations in diverse settings is needed to understand the effect on additional patient outcomes and to fully assess safety and generalizability	Single center only; small sample size; patient may have their own personal minimal MAP to be maintained during surgery; depth of anesthesia was not measured; the early warning system is validated only for invasive continuous blood pressure monitoring; an observer being present in the operating room may have influenced protocol adherence
Weisinger et al [43], 2021	Fugl-Meyer Assessment-Upper Extremity: week 4—AI: mean 23.2 (SD 3.91) versus control: mean 9.9 (SD 3.2); week 8—AI: mean 31.5 (SD 2.97) versus control: mean 23.1 (SD 4.99)	AI: 2.5 (0.18) points versus control: 1.3 (0.16) points; significance improved: Action Research Arm Test-Pinch subscale; significance improved: Box and Block Test; significance improved: NIHSS[ae]	Positive	Unknown	BQ treatment significantly improves upper extremity motor function in a population with subacute ischemic stroke across multiple clinical metrics. Further studies are planned and ongoing with larger study populations and in related indications	Nil
Blomberg et al [44], 2021	AI: 93.1% (296/318) versus control: 90.5% (304/336)	AI: 1.72 (1.52) minutes versus control: 1.70 (1.63) minutes; AI: 64.8% versus control: 61.9%	Negative	Yes	This randomized clinical trial did not find any significant improvement in dispatchers' ability to recognize cardiac arrest when supported by machine learning even though AI did surpass human recognition	Not 100% compliance with the machine learning model; the servers analyzing the phone calls had downtime, because the server was underdimensioned
Browning et al [45], 2021	QIDS-SR-16[af] at week 8—AI: 55.9% versus control: 51.8	Generalized Anxiety Disorder Assessment, 7 item version (week 8)—AI: −5.44 versus control: −6.12	Negative	Yes	Use of a predictive algorithm to guide antidepressant treatment improves symptoms of anxiety and functional outcomes provides initial support for the use of personalized medicine approaches in the treatment of depression	The accuracy of the predictive algorithm was modest at 57.5%; effectiveness was focused rather than efficacy, requesting but not requiring clinicians to alter treatment in response to a prediction of nonresponse; randomization occurred at the level of the patient rather than the site, and thus, the treatment as usual arm may have been influenced by behavior learned in the active arm

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Jayakumar et al [46], 2021	AI: mean 68.9 (SD 19.8) versus control: mean 48.8 (SD 14.5)	CollaboRATE median—AI: 8 of 69 versus control: 28 of 60; number of patient-rated satisfaction scores lower than the median value of 10—AI: 9 of 69 versus control: 19 of 60; no significant difference in duration of consultation in minutes; no significant difference in TKR[ag] rates and treatment concordance	Positive	No	In this randomized clinical trial, an AI-enabled decision aid significantly improved decision quality, level of shared decision-making, satisfaction, and physical limitations without significantly impacting consultation times, TKR rates, or treatment concordance in patients with knee osteoarthritis considering TKR. Decision aids using a personalized, data-driven approach can enhance shared decision-making in the management of knee osteoarthritis	Single-center study; surgeons were not masked; we did not assess the effect of the decision aid on patient knowledge; the typical course of a formal osteoarthritis in-clinic diagnosis possesses a general limitation in limiting the time frame over which the tool may be applied
Kamba et al [47], 2021	AI first: 13.8% versus control first: 35.7%	AI first: 14.2% versus control first: 40.6%; AI first: 13% versus control first: 38.5%; AI first: 64.5% versus control first: 53.6%	Positive	No	The reduction of AMR by assisting with CADe based on deep learning in a multicenter randomized controlled trial	Nil
Luo et al [48], 2021	AI: 38.7% versus control: 34%	The number of polyps detected in the control group and the research group was 80 and 105, respectively; AI: 91 versus control: 69; polyp type 0-IIa—AI: 87 versus control: 61; polyp type 0-Is—AI: 5 versus control: 8; polyp type 0-Ip—AI: 13 versus control: 11; 52 false positive result in AI group; in average, 0.35 false positive per colonoscopy	Positive	No	This study shows that an AI system based on deep learning and its real-time performance led to significant increases in colorectal PDR[ah]	Single center study; small sample size; AI has different effects on improving the PDR among different physicians; ADR was not compared between 2 groups in this trial
Rafferty et al [49], 2021	IBS[ai] symptom score—AI: −170 versus control: −138; quality of life score—AI: 31.1 versus control: 11.8	No significant difference; AI: 8.3 (4.4-13.1) versus control: 10.4 (7.4-14.0)	Negative	Yes	Results showed that the Heali app was able to significantly increase quality of life outcomes in IBS participants over a 30-day intervention period	Small sample size; self-reporting bias in survey may resulted owing to lack of blinding; stratification was not done; participants were not randomized to groups until study day 10, which was after the collection of baseline data; although anthropometric measures (bodyweight and height) were collected at baseline, they were not collected at the end of the trial, and it is possible that changes in body weight influenced the outcome variables; adherence may be affected by social impacts of the COVID-19 pandemic

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Repici et al [50], 2021	AI: 176/330, 53.3% versus control: 146/330, 44.2%	AI: 41.5% versus control: 36.1%; AI: 1.98 (range 0-15) versus control: 1.61 (range: 0-17); AI: 3.3% versus control: 5.2%; AI: 1.26 (SD 1.82) versus control: 1.04 (SD 1.75); 100% in both groups after excluding patients with inadequate bowel preparation; AI: mean 815 (SD 1.6) versus control: mean 7.98 (SD 1.5)	Positive	No	CADe in real-time colonoscopy significantly increases ADR and adenomas detected per colonoscopy in a nonexpert setting	No comparison of AI assistance with alternative educational interventions among inexpert endoscopists; this study design was not fit to assess the sensitivity or specificity of the device; no power calculations were done for any of our secondary outcomes
Strömblad et al [51], 2021	Mean absolute error—AI: 49.5 minutes {66} versus control: 59.3 minutes {72}	Mean patient wait time: overall—AI: 16.3 minutes versus control: 49.4 minutes (67.1% improvement); turnover time: overall—AI: 69.1 minutes versus control: 70.6 minutes (2% improvement); patient time in facility—AI: 148.1 versus control: 173.3 (14.5% improvement)	Positive	No	Implementing machine learning–generated predictions for surgical case durations may improve case duration accuracy, presurgical resource use, and patient wait time, without increasing surgeon wait time between cases	Small sample size; prediction accuracy may be affected if the submitted procedure codes deviate significantly from the procedures that are performed; a less common occurrence were multipanel cases in which multiple surgeons from different services operated on the same patient during the same case; there was no stratification by days
Wu et al [52], 2021	AI: 5.38 (SD 4.32) versus control: 9.82	AI: 5.40 (SD 3.82) minutes versus control: 4.38 (SD 3.91) minutes; the median percentage of patients with blind spots at each site—AI: 21% versus control: 38.9%; per-lesion accuracy: 84.7%; sensitivity: 100%; specificity: 84.3%	Positive	No	ENDOANGEL was an effective and robust system to improve the quality of EGD and has the potential to detect electrocardiogram in real time	We only conducted a feasibility analysis on real-time detection of gastric cancer based on deep learning in a clinical setting; the enrolled patients were not followed up for a long time; statisticians were not blinded
Yao et al [53], 2021	In overall cohort—AI: 2.1% versus control 1.6%; among 1356 patients who had a positive result—AI: 19.5% versus control: 14.5%	No significant between AI and control on disease discovery	Positive	No	An AI algorithm run on existing electrocardiograms enabled the early diagnosis of low ejection fraction in a large cohort of patients managed in routine primary care practices. Because electrocardiography is a low-cost test that is frequently performed for a variety of purposes, the algorithm could potentially improve early diagnosis and treatment of a condition that is often asymptomatic but has effective treatments and thus reduce the disease burden in broad populations	Echocardiogram may not be ordered by clinician as nearly all the patients had insurance coverage; study was not designed to determine the long-term clinical impact; for example, heart failure hospitalizations and mortality

Author (publication year)	Primary end point result	Secondary end point result	Study outcome	Underpowered	Study conclusion	Limitations
Brown et al [54], 2021	AMR—CADe first: 20.12% versus HDWL[aj] first: 31.25%	PDR—CADe first: 20.7% versus HDWL first: 33.71%; HPMR[ak]—no significant difference in the hyperplastic polyp miss rate; SSLMR[al]—CADe first: 7.140% versus HDWL first: 42.11%; no statistically significant difference in ADR during first pass, second pass, and whole process; no statistically significant difference in PDR during first pass, second pass, and whole process; adenoma per colonoscopy during first pass—CADe first: 1.19 versus HDWL first: 0.90; no statistically significant difference during second pass and whole process; polyp per colonoscopy during first pass—CADe first: 2.0 versus HDWL first: 1.59; polyp per colonoscopy during second pass—CADe first: 0.52 versus HDWL first: 0.81; no statistically significant difference during whole process; SSLPC[am] during second pass—CADe first: 0.01 versus HDWL first: 0.07; no statistically significant difference during first pass whole process	Positive	No	This study showed a decrease in AMR with the use of a deep learning CADe system when compared with HDWL colonoscopy alone and a decrease in polyp and sessile serrated lesion miss rates and an increase in first-pass adenomas per colonoscopy	Not powered to detect a difference in ADR; the tandem colonoscopy design limited in terms of generalizability to the real-world clinical setting; only included experienced endoscopists with a high baseline ADR at US academic medical centers; used a second monitor adjacent to the primary endoscopy monitor

[a] CPAP: continuous positive airway pressure.

[b] AI: artificial intelligence.

[c] ICU: intensive care unit.

[d] LOS: length of stay.

[e] HGD: high-grade dysplasia.

[f] EAC: esophageal adenocarcinoma.

[g] WATS: wide-area transepithelial sampling.

[h] BE: Barrett's esophagus.

[i] SAP: sensor-augmented pump.

[j] HPI: hypotension probability indicator.

[k] RCT: randomized controlled trial.

[l] BPI: Brief Pain Inventory.

[m] BQ-II: Barriers Questionnaire II.

[n] SBP: systolic blood pressure.

[o] SRS-II: Social Responsiveness Scale II.

[p] VABS-II: Vineland Adaptive Behavioural Scales, Second edition.

[q] OR: odds ratio.

[r] ADR: adenoma detection rate.

[s] EGD: esophagogastroduodenoscopy.

[t] DAP: dose–area product.

[u] CBCT: cone-beam computed tomography.

[v] U-TOE: ultrathin transoral endoscopy.

[w] ITT: intention to treat.

[x] PP: per protocol.

[y] CADe: computer-assisted detection.

[z] CTV: clinical target volume.

[aa] LDR: low-dose rate.

[ab] AQCS: automatic quality control system.

[ac] BBPS: Boston bowel preparation scale.

[ad] DCNN: deep convolutional neural networks.

[ae] AMR: adenoma miss rate.

[af] QIDS-SR-16: Quick Inventory of Depressive Symptomatology (16-Item) (Self-Report).

[ag] TKR: total knee replacement.

[ah] PDR: polyp detection rate.

[ai] IBS: irritable bowel syndrome.

[aj] HDWL: high-definition white light.

[ak] HPMR: Hyperplastic polyp miss rate.

[al] SSLMR: sessile serrated lesion miss rate.

[am] SSLPC: sessile serrated lesion per colonoscopy.

Study Limitations

The most common limitations listed by the authors among these studies were single-center study design (22/39, 56%) and small sample size (n<1000; 33/39, 85%). This limits the generalizability and statistical power of the AI-assisted tools in different studies. There were 7 studies which were underpowered because of small sample size. Of these, 5 (71%) studies included <100 participants. Another common limitation is the open-label design (15/39, 38%).

Assessment of Risk of Bias

Detailed assessment results of the risk of bias using the second version of the Cochrane risk-of-bias tool for randomized trials are reported in Table 4. On the basis of the overall risk-of-bias assessment, 20% (8/39) of the trials had a low risk of bias, 31% (12/39) trials had some concerns, and 49% (19/39) had a high risk of bias. Missing outcome data and outcome measurements were the most common risk factors.

Table 4. Quality assessment outcome based on the second version of the Cochrane risk-of-bias tool for randomized trials.

Author (publication year)	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall bias
El Solh et al [17], 2009	Low	Low	Low	Low	Low	Low
Shimabukuro et al [18], 2017	Low	Low	Low	Low	Low	Low
Labovitz et al [6], 2017	Some concerns	Low	Low	Low	Low	Some concerns
Gracey et al [19], 2018	Some concerns	High	High	High	High	High
Liu et al [20], 2018	Some concerns	Some concerns	High	High	Some concerns	High
Vennalaganti et al [21], 2018	Some concerns	Low	Low	High	Low	High
Biester et al [22], 2019	Some concerns	Some concerns	High	Low	High	High
Pouska et al [23], 2019	Some concerns	High	Low	Low	Low	High
Lin et al [24], 2019	Low	Low	Low	Low	Low	Low
Kamdar et al [25], 2019	Some concerns	Some concerns	High	High	Some concerns	High
Persell et al [26], 2020	Low	High	Low	High	Low	High
Voss et al [27], 2019	High	Some concerns	Low	High	Low	High
Wang et al [28], 2019	Some concerns	Low	Low	Low	Low	Some concerns
Wu et al [29], 2019	Some concerns	Low	Low	Low	Low	Some concerns
Pavel et al [30], 2020	Some concerns	Low	Low	Low	Low	Some concerns
Alfonsi et al [31], 2020	Some concerns	Low	Low	High	Low	High
Auloge et al [32], 2020	Some concerns	Some concerns	Low	Low	Low	Some concerns
Avari et al [33], 2020	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Chen et al [34], 2020	Some concerns	Low	Low	Low	Low	Some concerns
Gong et al [35], 2020	Low	Low	Low	Low	Low	Low
Liu et al [36], 2020	Some concerns	Low	Low	Some concerns	Low	Some concerns
Nicolae et al [37], 2020	Some concerns	Some concerns	High	Low	Low	High
Repici et al [38], 2020	Some concerns	Low	Low	Low	Low	Some concerns
Su et al [39], 2020	Low	Low	Low	Low	Low	Low
Wang et al [40], 2020	Low	Low	Low	High	Low	High
Wang et al [41], 2020	Low	Low	Low	Some concerns	Low	Some concerns
Wijnberge et al [42], 2020	Some concerns	Low	Low	Low	Low	Some concerns
Weisinger et al [43], 2021	Some concerns	Low	High	High	Low	High
Blomberg et al [44], 2021	Low	Low	High	Low	Low	High
Browning et al [45], 2021	Low	Low	Low	Low	Low	Low
Jayakumar et al [46], 2021	Some concerns	Low	Low	High	Low	High
Kamba et al [47], 2021	Some concerns	High	High	Some concerns	Low	High
Luo et al [48], 2021	Some concerns	Low	Low	Low	Low	Some concerns
Rafferty et al [49], 2021	High	Low	Low	High	Low	High
Repici et al [50], 2021	Some concerns	Some concerns	High	High	Low	High
Strömblad et al [51], 2021	Low	Low	Low	Low	Low	Low
Wu et al [52], 2021	Low	Low	Low	Low	Low	Low
Yao et al [53], 2021	Some concerns	High	High	High	Low	High
Brown et al [54], 2021	Low	Low	Low	High	Low	High

Discussion

Principal Findings

Despite the plethora of claims for the benefits of AI in enhancing clinical outcomes, there is a paucity of robust evidence. In this systematic review, we identified only a handful of RCTs comparing AI-assisted tools with standard-of-care management in various medical conditions. Among these RCTs, two-thirds demonstrated improved primary or secondary end points compared with the standard-of-care management. However, not all of these end points are clinically relevant, that is, leading to a change in the management plan, improving the treatment results, shortening or avoiding hospital admissions, or reducing mortality. Although we acknowledge that our definition of a *clinically relevant end point* may be relatively narrow, we believe that the absence of such end points in the RCTs shows a clear deficit in the available evidence.

As expected, most of these studies came from economically advanced, industrialized countries, which constituted two-thirds of the RCTs included in this systematic review. China, as a single nation, accounted for one-third of the RCTs. China's research in this area is empowered by immense amount of resources invested in AI or machine learning (ML), internet, its vast patient population, and the availability of a nationwide electronic health record for hospitalized patients [55]. The geographical distribution of these studies is important as AI or ML relies on data fed to the system. Differences in genomic, metagenomic, and even environmental factors may influence disease patterns and the presentation of diseases. Therefore, it is desirable to develop an AI or ML tool based on data collected from different ethnic groups and tested in individual regions to prove its efficacy. There is only one AI-assisted tool, EndoScreener, which uses different ethnic groups in its development and validation. It was originally developed and trained using a different data set of endoscopic images including an open-source database of endoscopic images from Spain. Subsequently, the tool was validated in 4 prospective RCTs in China and had been recently validated in a multicenter RCT in the United States, proving its effectiveness. Future studies should focus on validation of AI-assisted tools across different ethnic groups and patient populations to ensure generalizability.

More biosignal-based AI-assisted tools have been studied than clinical data–based tools in RCTs. The most widely used were endoscopic images detecting adenoma during colonoscopy. The adenoma detection algorithm appears to be easier for cross-compatibility because of the distinct difference in appearance between adenoma (or polyp) and normal mucosa [56]. There were a total of 5 different AI-assisted adenoma or polyp detection systems tested in 9 separate RCTs, all of which successfully assisted endoscopists to detect more adenomas or polyps during colonoscopy. However, only one study successfully showed that the AI-assisted adenoma detection system could improve adenoma detection of all sizes. Other studies could only show improvement in diminutive adenoma (<5 mm) or small adenoma (<10 mm) detection [39]. Advanced adenoma or colorectal cancer detection was not improved by the AI-assisted adenoma detection system used in these studies. The US Multi-Society Task Force on Colorectal Cancer [57] suggested that patients with 1 to 2 nonadvanced adenoma sized <10 mm are at low risk and could have their surveillance colonoscopy in 7 to 10 years. The value of improvement in diminutive or small adenoma detection is uncertain. Among all studies, there was only one reported long-term outcome, that is, in-hospital mortality. Future studies should emphasize on the impact of AI-assisted tools on the long-term clinical end points.

Classical prediction models are typically clinical risk scores derived from regression-based statistical models, which could be considered an ML model that has been modified for clinical use. Ideally, RCTs should be designed with a control arm (usual clinical care), a "standard-of-care" clinical risk score arm, and a novel AI-assisted tool arm. However, given the expense and effort of a clinical trial, the SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence) extension guidelines clearly state the importance of pre-existing evidence for AI intervention, with evidence that the AI-assisted tool produces better performance compared with the standard of care [58]. Although none of the RCTs found in this systematic review used a multi-arm design, there have been well-designed studies where more "complex" ML approaches have outperformed regression-derived clinical risk scores on external validation [7,59,60].

Conclusions and Recommendations

The findings of this systematic review are by no means to discourage the use of AI in medicine. AI or ML can detect signals in an immense data pool to develop algorithms for clinical decision. Unlike humans, AI or ML can process enormous quantities of data, perform consistently, and constantly improve its performance by learning from new data. However, for AI or ML to be implemented in daily clinical practice, assisting clinicians in making important decisions, proof-of-concept evidence is not sufficient. AI-assisted tools must demonstrate unequivocal improvement in clinically relevant outcomes in properly designed randomized controlled clinical trials in which AI-assisted management is compared with standard-of-care practice. Researchers should not only focus on demonstrating the robustness of the AI algorithm in concept studies but also on translating from code to bedside by conducting RCTs in real-life clinical settings. From our systematic review, automated polyp detection is the most widely implemented AI technology in clinical practice, which sets a good example of the pathway from algorithm development to the implementation of AI technology in real-life clinical practice. Another obstacle to the implementation of AI or ML in daily clinical practice is the regulation of these technologies [8]. To grant approval from regulatory bodies, scientific evidence is required to support the safety and effectiveness of an AI-assisted tool in clinical practice. The framework for AI health care product development highlighted that RCTs are often recommended to provide strong evidence to validate the clinical efficacy and safety of an AI-assisted tool in real-world settings [61]. More RCTs of AI-assisted tools integrated into clinical practice are required to advance the role of AI or ML in medicine. We should also test how machine intelligence and human intelligence can work together on personalized management of patients.

Acknowledgments

Author DS is the co-corresponding author of this manuscript and can be reached by email at dennis.shung@yale.edu.

Data Availability

The full search strategy is provided in Multimedia Appendix 1. Additional data are available on request.

Authors' Contributions

JJYS developed the concept of the systematic review. TYTL, MFKC, YLM, KML, and JJYS contributed to the development of the study protocol. YLM and KML refined the search strategy and searched for articles. TYTL and MFKC performed study selection, data extraction, data verification, and critical appraisal. JJYS gave an independent view in case of any discrepancies. TYTL drafted the manuscript. DS and JJYS critically revised the manuscript for intellectually important content. All authors had full access to all the data presented in this manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy.
[DOCX File , 15 KB - jmir_v24i8e37188_app1.docx]

Multimedia Appendix 2
Study design and artificial intelligence (AI)–assisted tools characteristics.
[DOCX File , 28 KB - jmir_v24i8e37188_app2.docx]

References

1. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. AI Magazine 2006 Dec 15;27(4):12 [FREE Full text] [doi: 10.1609/aimag.v27i4.1904]
2. Hamamoto R. Application of artificial intelligence for medical research. Biomolecules 2021 Jan 12;11(1):90 [FREE Full text] [doi: 10.3390/biom11010090] [Medline: 33445802]
3. Shung DL, Sung JJ. Challenges of developing artificial intelligence-assisted tools for clinical medicine. J Gastroenterol Hepatol 2021 Feb;36(2):295-298. [doi: 10.1111/jgh.15378] [Medline: 33624889]
4. Sung JJ, Stewart CL, Freedman B. Artificial intelligence in health care: preparing for the fifth Industrial Revolution. Med J Aust 2020 Sep;213(6):253-5.e1. [doi: 10.5694/mja2.50755] [Medline: 32892395]
5. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RV, Imaging and Informatics in Retinopathy of Prematurity (i-ROP) Research Consortium. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol 2018 Jul 01;136(7):803-810 [FREE Full text] [doi: 10.1001/jamaophthalmol.2018.1934] [Medline: 29801159]
6. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017 May;48(5):1416-1419 [FREE Full text] [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
7. Shung DL, Au B, Taylor RA, Tay JK, Laursen SB, Stanley AJ, et al. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. Gastroenterology 2020 Jan;158(1):160-167 [FREE Full text] [doi: 10.1053/j.gastro.2019.09.009] [Medline: 31562847]
8. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020 Sep 11;3:118 [FREE Full text] [doi: 10.1038/s41746-020-00324-0] [Medline: 32984550]
9. Hassan C, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc 2021 Jan;93(1):77-85.e6. [doi: 10.1016/j.gie.2020.06.059] [Medline: 32598963]
10. Batchu S, Liu F, Amireh A, Waller J, Umair M. A review of applications of machine learning in mammography and future challenges. Oncology 2021;99(8):483-490 [FREE Full text] [doi: 10.1159/000515698] [Medline: 34023831]

11. Yeo M, Tahayori B, Kok HK, Maingard J, Kutaiba N, Russell J, et al. Review of deep learning algorithms for the automatic detection of intracranial hemorrhages on computed tomography head imaging. J Neurointerv Surg 2021 Apr;13(4):369-378. [doi: 10.1136/neurintsurg-2020-017099] [Medline: 33479036]
12. Zhu T, Li K, Herrero P, Georgiou P. Deep learning for diabetes: a systematic review. IEEE J Biomed Health Inform 2021 Jul;25(7):2744-2757. [doi: 10.1109/JBHI.2020.3040225] [Medline: 33232247]
13. 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]
14. Sackett D. Evidence-based medicine. Lancet 1995 Oct 28;346(8983):1171. [Medline: 7475646]
15. 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]
16. 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]
17. El Solh A, Akinnusi M, Patel A, Bhat A, TenBrock R. Predicting optimal CPAP by neural network reduces titration failure: a randomized study. Sleep Breath 2009 Nov;13(4):325-330. [doi: 10.1007/s11325-009-0247-5] [Medline: 19259717]
18. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res 2017 Nov 9;4(1):e000234 [FREE Full text] [doi: 10.1136/bmjresp-2017-000234] [Medline: 29435343]
19. Gracey B, Jones CA, Cho D, Conner S, Greene E. Improving medication adherence by better targeting interventions using artificial intelligence - a randomized control study. Value Health 2018 May;21:S76. [doi: 10.1016/j.jval.2018.04.532]
20. Liu X. A0069 Evaluating the impact of an integrated computer-based decision support with person-centered analytics for the management of hypertension. J Hypertension 2018 Oct;36:e2. [doi: 10.1097/01.hjh.0000547994.70348.44]
21. Vennalaganti PR, Kaul V, Wang KK, Falk GW, Shaheen NJ, Infantolino A, et al. Increased detection of Barrett's esophagus-associated neoplasia using wide-area trans-epithelial sampling: a multicenter, prospective, randomized trial. Gastrointest Endosc 2018 Feb;87(2):348-355. [doi: 10.1016/j.gie.2017.07.039] [Medline: 28757316]
22. Biester T, Nir J, Remus K, Farfel A, Muller I, Biester S, et al. DREAM5: an open-label, randomized, cross-over study to evaluate the safety and efficacy of day and night closed-loop control by comparing the MD-Logic automated insulin delivery system to sensor augmented pump therapy in patients with type 1 diabetes at home. Diabetes Obes Metab 2019 Apr;21(4):822-828. [doi: 10.1111/dom.13585] [Medline: 30478937]
23. Pouska J, Cerveny V, Zatloukal J, Kletecka J, Benes J. The use of HPI (Hypotension probability indicator) during major intracranial surgery; preliminary results of a prospective randomized trial. Intensive Care Med Exp 2019;7(Suppl 3):001525.
24. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. EClinicalMedicine 2019 Mar 17;9:52-59 [FREE Full text] [doi: 10.1016/j.eclinm.2019.03.001] [Medline: 31143882]
25. Kamdar M, Centi AJ, Agboola S, Fischer N, Rinaldi S, Strand JJ, et al. A randomized controlled trial of a novel artificial intelligence-based smartphone application to optimize the management of cancer-related pain. J Clin Oncol 2019 May 20;37(15_suppl):11514. [doi: 10.1200/jco.2019.37.15_suppl.11514]
26. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, et al. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. JAMA Netw Open 2020 Mar 02;3(3):e200255 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0255] [Medline: 32119093]
27. Voss C, Schwartz J, Daniels J, Kline A, Haber N, Washington P, et al. Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial. JAMA Pediatr 2019 May 01;173(5):446-454 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.0285] [Medline: 30907929]
28. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019 Oct;68(10):1813-1819 [FREE Full text] [doi: 10.1136/gutjnl-2018-317500] [Medline: 30814121]
29. Wu L, Zhang J, Zhou W, An P, Shen L, Liu J, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut 2019 Dec;68(12):2161-2169 [FREE Full text] [doi: 10.1136/gutjnl-2018-317366] [Medline: 30858305]
30. Pavel AM, Rennie JM, de Vries LS, Blennow M, Foran A, Shah DK, et al. A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial. Lancet Child Adolesc Health 2020 Oct;4(10):740-749 [FREE Full text] [doi: 10.1016/S2352-4642(20)30239-X] [Medline: 32861271]
31. Alfonsi JE, Choi EE, Arshad T, Sammott SA, Pais V, Nguyen C, et al. Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized control trial. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e22074 [FREE Full text] [doi: 10.2196/22074] [Medline: 33112249]
32. Auloge P, Cazzato RL, Ramamurthy N, de Marini P, Rousseau C, Garnon J, et al. Augmented reality and artificial intelligence-based navigation during percutaneous vertebroplasty: a pilot randomised clinical trial. Eur Spine J 2020 Jul;29(7):1580-1589. [doi: 10.1007/s00586-019-06054-6] [Medline: 31270676]

33. Avari P, Leal Y, Herrero P, Wos M, Jugnee N, Arnoriaga-Rodríguez M, et al. Safety and feasibility of the PEPPER adaptive bolus advisor and safety system: a randomized control study. Diabetes Technol Ther 2021 Mar;23(3):175-186. [doi: 10.1089/dia.2020.0301] [Medline: 33048581]
34. Chen D, Wu L, Li Y, Zhang J, Liu J, Huang L, et al. Comparing blind spots of unsedated ultrafine, sedated, and unsedated conventional gastroscopy with and without artificial intelligence: a prospective, single-blind, 3-parallel-group, randomized, single-center trial. Gastrointest Endosc 2020 Feb;91(2):332-9.e3. [doi: 10.1016/j.gie.2019.09.016] [Medline: 31541626]
35. Gong D, Wu L, Zhang J, Mu G, Shen L, Liu J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. Lancet Gastroenterol Hepatol 2020 Apr;5(4):352-361. [doi: 10.1016/S2468-1253(19)30413-3] [Medline: 31981518]
36. Liu P, Wang P, Glissen Brown JR, Berzin TM, Zhou G, Liu W, et al. The single-monitor trial: an embedded CADe system increased adenoma detection during colonoscopy: a prospective randomized study. Therap Adv Gastroenterol 2020;13:1756284820979165 [FREE Full text] [doi: 10.1177/1756284820979165] [Medline: 33403003]
37. Nicolae A, Semple M, Lu L, Smith M, Chung H, Loblaw A, et al. Conventional vs machine learning-based treatment planning in prostate brachytherapy: results of a Phase I randomized controlled trial. Brachytherapy 2020;19(4):470-476 [FREE Full text] [doi: 10.1016/j.brachy.2020.03.004] [Medline: 32317241]
38. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology 2020 Aug;159(2):512-20.e7. [doi: 10.1053/j.gastro.2020.04.062] [Medline: 32371116]
39. Su JR, Li Z, Shao XJ, Ji CR, Ji R, Zhou RC, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointest Endosc 2020 Feb;91(2):415-24.e4. [doi: 10.1016/j.gie.2019.08.026] [Medline: 31454493]
40. Wang P, Liu P, Glissen Brown JR, Berzin TM, Zhou G, Lei S, et al. Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study. Gastroenterology 2020 Oct;159(4):1252-61.e5. [doi: 10.1053/j.gastro.2020.06.023] [Medline: 32562721]
41. Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. Lancet Gastroenterol Hepatol 2020 Apr;5(4):343-351. [doi: 10.1016/S2468-1253(19)30411-X] [Medline: 31981517]
42. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. JAMA 2020 Mar 17;323(11):1052-1060 [FREE Full text] [doi: 10.1001/jama.2020.0592] [Medline: 32065827]
43. Weisinger BS, Bornstein NM, Shohami E, Segal Y, Alter A, Lifshitz A, et al. Abstract P194: Artificial intelligence-powered non-invasive and frequency-tuned electromagnetic field therapy improves upper extremity motor function in sub-acute stroke patients: a pilot randomized controlled trial. Stroke 2021 Mar;52(Suppl_1):194. [doi: 10.1161/str.52.suppl_1.p194]
44. Blomberg SN, Christensen HC, Lippert F, Ersbøll AK, Torp-Petersen C, Sayre MR, et al. Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: a randomized clinical trial. JAMA Netw Open 2021 Jan 04;4(1):e2032320 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.32320] [Medline: 33404620]
45. Browning M, Bilderbeck AC, Dias R, Dourish CT, Kingslake J, Deckert J, et al. The clinical effectiveness of using a predictive algorithm to guide antidepressant treatment in primary care (PReDicT): an open-label, randomised controlled trial. Neuropsychopharmacology 2021 Jun;46(7):1307-1314 [FREE Full text] [doi: 10.1038/s41386-021-00981-z] [Medline: 33637837]
46. Jayakumar P, Moore MG, Furlough KA, Uhler LM, Andrawis JP, Koenig KM, et al. Comparison of an artificial intelligence-enabled patient decision aid vs educational material on decision quality, shared decision-making, patient experience, and functional outcomes in adults with knee osteoarthritis: a randomized clinical trial. JAMA Netw Open 2021 Feb 01;4(2):e2037107 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.37107] [Medline: 33599773]
47. Kamba S, Tamai N, Horiuchi H, Matsui H, Kobayashi M, Ego M, et al. ID: 3519580 A multicentre randomized controlled trial to verify the reducibility of adenoma miss rate of colonoscopy assisted with artificial intelligence based software. Gastrointest Endosc 2021 Jun 1;93(6):AB195. [doi: 10.1016/j.gie.2021.03.439]
48. Luo Y, Zhang Y, Liu M, Lai Y, Liu P, Wang Z, et al. Artificial intelligence-assisted colonoscopy for detection of colon polyps: a prospective, randomized cohort study. J Gastrointest Surg 2021 Aug;25(8):2011-2018 [FREE Full text] [doi: 10.1007/s11605-020-04802-4] [Medline: 32968933]
49. Rafferty AJ, Hall R, Johnston CS. A novel mobile app (Heali) for disease treatment in participants with irritable bowel syndrome: randomized controlled pilot trial. J Med Internet Res 2021 Mar 02;23(3):e24134 [FREE Full text] [doi: 10.2196/24134] [Medline: 33650977]
50. Repici A, Spadaccini M, Antonelli G, Correale L, Maselli R, Galtieri PA, et al. Artificial intelligence and colonoscopy experience: lessons from two randomised trials. Gut 2022 Apr;71(4):757-765. [doi: 10.1136/gutjnl-2021-324471] [Medline: 34187845]

51. Strömblad CT, Baxter-King RG, Meisami A, Yee SJ, Levine MR, Ostrovsky A, et al. Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources: a randomized clinical trial. JAMA Surg 2021 Apr 01;156(4):315-321 [FREE Full text] [doi: 10.1001/jamasurg.2020.6361] [Medline: 33502448]
52. Wu L, He X, Liu M, Xie H, An P, Zhang J, et al. Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric cancer: a randomized controlled trial. Endoscopy 2021 Dec;53(12):1199-1207. [doi: 10.1055/a-1350-5583] [Medline: 33429441]
53. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. Nat Med 2021 May;27(5):815-819. [doi: 10.1038/s41591-021-01335-4] [Medline: 33958795]
54. Glissen Brown JR, Mansour NM, Wang P, Chuchuca MA, Minchenberg SB, Chandnani M, et al. Deep learning computer-aided polyp detection reduces adenoma miss rate: a United States multi-center randomized tandem colonoscopy study (CADeT-CS trial). Clin Gastroenterol Hepatol 2022 Jul;20(7):1499-507.e4 [FREE Full text] [doi: 10.1016/j.cgh.2021.09.009] [Medline: 34530161]
55. Zhang L, Long J, Jiang W, Shi Y, He X, Zhou Z, et al. Trends in chronic kidney disease in China. N Engl J Med 2016 Sep 01;375(9):905-906. [doi: 10.1056/NEJMc1602469] [Medline: 27579659]
56. Samarasena JB. Guns, germs, and steel…and artificial intelligence. Gastrointest Endosc 2021 Jan;93(1):99-101. [doi: 10.1016/j.gie.2020.07.021] [Medline: 33353643]
57. Gupta S, Lieberman D, Anderson JC, Burke CA, Dominitz JA, Kaltenbach T, et al. Recommendations for follow-up after colonoscopy and polypectomy: a consensus update by the US multi-society task force on colorectal cancer. Am J Gastroenterol 2020 Mar;115(3):415-434 [FREE Full text] [doi: 10.14309/ajg.0000000000000544] [Medline: 32039982]
58. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ, SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health 2020 Oct;2(10):e549-e560 [FREE Full text] [doi: 10.1016/S2589-7500(20)30219-3] [Medline: 33328049]
59. Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M. Machine learning outperforms ACC / AHA CVD risk calculator in MESA. J Am Heart Assoc 2018 Nov 20;7(22):e009476 [FREE Full text] [doi: 10.1161/JAHA.118.009476] [Medline: 30571498]
60. Brown SM, Jones J, Kuttler KG, Keddington RK, Allen TL, Haug P. Prospective evaluation of an automated method to identify patients with severe sepsis or septic shock in the emergency department. BMC Emerg Med 2016 Aug 22;16(1):31 [FREE Full text] [doi: 10.1186/s12873-016-0095-0] [Medline: 27549755]
61. Higgins D, Madai VI. From bit to bedside: a practical framework for artificial intelligence product development in healthcare. Adv Intell Syst 2020 Jul 02;2(10):2000052. [doi: 10.1002/aisy.202000052]

Abbreviations

AI: artificial intelligence
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SPIRIT-AI: Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence

Edited by G Eysenbach; submitted 10.02.22; peer-reviewed by T Jamieson, B Puladi, SY Shin; comments to author 19.05.22; revised version received 13.06.22; accepted 29.07.22; published 25.08.22.

Please cite as:
Lam TYT, Cheung MFK, Munro YL, Lim KM, Shung D, Sung JJY
Randomized Controlled Trials of Artificial Intelligence in Clinical Practice: Systematic Review
J Med Internet Res 2022;24(8):e37188
URL: https://www.jmir.org/2022/8/e37188
doi:10.2196/37188
PMID:35904087

Review

Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks

Norah L Crossnohere[1,2], PhD; Mohamed Elsaid[1], PhD; Jonathan Paskett[1], BSc; Seuli Bose-Brill[2], MD; John F P Bridges[1], PhD

[1]Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, OH, United States

[2]Division of General Internal Medicine, Department of Internal Medicine, The Ohio State University College of Medicine, Columbus, OH, United States

Corresponding Author:
Norah L Crossnohere, PhD
Department of Biomedical Informatics
The Ohio State University College of Medicine
1800 Cannon Drive
Columbus, OH, 43210
United States
Phone: 1 3476286314
Email: norah.crossnohere@osumc.edu

Abstract

Background: Artificial intelligence (AI) is rapidly expanding in medicine despite a lack of consensus on its application and evaluation.

Objective: We sought to identify current frameworks guiding the application and evaluation of AI for predictive analytics in medicine and to describe the content of these frameworks. We also assessed what stages along the AI translational spectrum (ie, AI development, reporting, evaluation, implementation, and surveillance) the content of each framework has been discussed.

Methods: We performed a literature review of frameworks regarding the oversight of AI in medicine. The search included key topics such as “artificial intelligence,” “machine learning,” “guidance as topic,” and “translational science,” and spanned the time period 2014-2022. Documents were included if they provided generalizable guidance regarding the use or evaluation of AI in medicine. Included frameworks are summarized descriptively and were subjected to content analysis. A novel evaluation matrix was developed and applied to appraise the frameworks’ coverage of content areas across translational stages.

Results: Fourteen frameworks are featured in the review, including six frameworks that provide descriptive guidance and eight that provide reporting checklists for medical applications of AI. Content analysis revealed five considerations related to the oversight of AI in medicine across frameworks: transparency, reproducibility, ethics, effectiveness, and engagement. All frameworks include discussions regarding transparency, reproducibility, ethics, and effectiveness, while only half of the frameworks discuss engagement. The evaluation matrix revealed that frameworks were most likely to report AI considerations for the translational stage of development and were least likely to report considerations for the translational stage of surveillance.

Conclusions: Existing frameworks for the application and evaluation of AI in medicine notably offer less input on the role of engagement in oversight and regarding the translational stage of surveillance. Identifying and optimizing strategies for engagement are essential to ensure that AI can meaningfully benefit patients and other end users.

(J Med Internet Res 2022;24(8):e36823) doi:10.2196/36823

KEYWORDS

artificial intelligence; translational science; translational research; ethics; engagement; reproducibility; transparency; effectiveness; medicine; health care; AI

Introduction

Artificial intelligence (AI) allows computers to accomplish tasks that normally require the use of human intelligence. Creating AI, or an AI computer system, begins when developers feed the system existing data and allow it to “learn.” This learning experience enables AI to understand, infer, communicate, and make decisions similar to, or better than, humans [1,2]. The use of AI in medicine is an area of rapid growth, with worldwide spending on health care AI technologies estimated to reach US $45 billion by 2026 [3]. AI is used across numerous medical specialties, and can be applied to inform medical decision-making in numerous ways, such as through expediting and reducing the costs of drug discovery [4]; offering insight that aids clinicians in diagnosing, prognosing, or optimizing treatment plans at the point of care; and automating medical administration activities such as appointment reminders [5].

Numerous concerns have been raised regarding a lack of oversight for the rapid development and expansion of AI in medicine. Commentators have drawn attention to the potential weaknesses and limitations of AI in medicine, including challenges spanning ethical, legal, regulatory, methodological, and technical domains [6]. These perspectives have highlighted pitfalls such as implicit bias, reproducibility, and clinical validity [7-9]. There is further concern that the methods for development and approaches for evaluation of AI are not as robust and rigorous as those of other medical interventions [10]. Although several best practices for the design, implementation, and evaluation of AI can be informed by the biostatistical and data science literature, such guidelines are not sufficient to address all concerns related to AI in medicine [11].

Translational science is the study of how to turn concepts, observations, or theories into actions and interventions by following defined stages of research and development. This is done to improve the health of individuals and society [12]. The stages of the translation for typical diagnostics and therapeutics often follow a traditional pathway from ideation to community implementation and social benefit [13]. Very clear, albeit complex, translation pathways exist for diagnostics and therapeutics, and are enforced by regulatory, funding, and ethical review. For AI, the translational pathway is less well-defined and overseen, but generally includes stages such as development, design, validation, reporting, implementation, and scaling [14]. Nevertheless, questions remain regarding how to adapt translational oversight mechanisms for AI in medicine [15].

Developing robust guidance for the oversight of AI along its translational pathway is essential to facilitating its clinical impact [16]. Several professional organizations have developed frameworks to address concepts specific to the development, reporting, and validation of AI in medicine [2,16-20]. These frameworks are focused primarily on informing the technological developers of AI (such as by offering guidance on how to promote transparency in the design and reporting of AI algorithms), rather than informing the clinical application of AI [2,20]. Regulatory oversight of AI is also in nascent stages. Guidance on how to critically evaluate actual applications of AI in medicine are currently in development by the US Food and Drug Administration (FDA) [21]. The European Commission has led a multidisciplinary initiative to increase the trustworthiness of AI [22,23], and the European Medicines Agency has identified the regulation of AI as a strategic priority [24].

Identifying considerations for the oversight of AI across the translational spectrum is essential to increasing the utility of AI in medicine. In this study, we explored and characterized existing frameworks regarding the oversight of AI in medicine. We then identified specific considerations raised in these frameworks and mapped them to different stages of the translational process for AI.

Methods

Identification of Frameworks

We performed a literature review to identify guidance on the use of predictive analytic AI in medicine. The search spanned the PubMed, Web of Science, and Embase databases, and also included a grey literature search of Google. Key terms for searching included "artificial intelligence," "machine learning," "guidance as topic," and "translational science." Documents were included if they provided generalized guidance (ie, were a framework) on applying or evaluating AI in medicine. Documents that described specific AI applications without offering overarching guidance on the use of AI were excluded. The reference lists of included frameworks were screened for additional relevant sources. Frameworks were not restricted to the use of AI in any specific condition or medical setting. The time period of the review was January 2014 to May 2022; 2014 was selected as the cut-off point, as this was the year when regulatory agencies in the United States and Europe began using the authorization designation of "software as a medical device," which includes regulation over AI.

Data Abstraction, Coding, and Analysis

A structured abstraction process was used to collect general information about each framework, including title, author/affiliation, year, summary, and intended audience. Frameworks were analyzed using content analysis, which is an approach for exploring themes or patterns from textual documents [25]. Content analysis of text-based sources can be either qualitative, where theory or themes are identified, or quantitative, wherein numeric information is derived [26]. We employed both approaches in this study. We first used qualitative content analysis to identify the different topics ("domains") discussed by frameworks. Codes for these domains were not developed a priori but were rather identified inductively through a reading of the frameworks. Frameworks were evaluated to assess whether they discussed each domain in relation to each of the translational stages [27]. Stages of AI translation were predefined to reflect the full AI product lifecycle, including development, validation, reporting, implementation, and surveillance. We used evaluation matrix methodologies [28-30] to depict how many frameworks described the domains identified through content analysis.

Data were visualized using several approaches. First, we used spider plots to visualize, for each individual framework, how many stages of translation were discussed in relation to each of the five domains. Second, we applied a heatmap to depict the number of frameworks discussing a given domain across each translational stage. The heatmap cross-walked the domains across the five stages of translation.

Results

Overview of the Frameworks

A total of 14 documents were included in the review, which are summarized in Table 1. One framework was published in 2016 (*Guidelines for Developing and Reporting* [31]) and all others were published from 2019 to 2020. Several of the frameworks were developed through pathways with professional organizations (*AI in Health Care* [32], *CONSORT-AI* [20], *SPIRIT-AI* [2], *DECIDE-AI* [33]). All frameworks were published as journal articles, and *AI in Healthcare* was published as both a journal article [7] and a White Paper [32]; since the journal article was a synopsis of the White Paper, the latter was used as the primary document of reference for this review. The frameworks explored in this review were generally consensus- rather than evidence-based. All but three frameworks [19,34,35] identified greater than one intended audience, and typical audiences included AI developers, investigators, clinicians, patients, and policymakers. Frameworks provided either general guidance on the use of AI in medicine, typically in narrative prose (herein referred to as "descriptive frameworks") [19,32,34-37] or guidance specifically on the reporting of AI studies in medicine, typically in checklist style (herein referred to as "reporting frameworks") [2,17,20,31,33,38-40].

Table 1. Summary of frameworks for the use of artificial intelligence (AI) in medicine.

Frameworks	Summary	Audience
Descriptive frameworks		
AI in Healthcare, Matheny et al [32][a]	Describes general challenges and opportunities associated with the use of AI in medicine	AI developers, clinicians, patients, policymakers
Clinician Checklist, Scott et al [34]	Describes recommendations on evaluating the suitability of AI applications for clinical settings	Clinicians
Ethical Considerations, Char et al [36]	Describes a roadmap for considering ethical aspects of AI with health care applications	AI developers, investigators, clinicians, policymakers
Evaluating AI, Park et al [37]	Describes an evaluation framework for the application of AI in medicine	Investigators, health care organizations
Users' Guide, Liu et al [19]	Describes an approach for assessing published literature using AI for medical diagnoses	Clinicians
Reporting and Implementing Interventions, Bates et al [35]	Describes barriers to the implementation of AI in medicine and provides solutions to address them	Health care organizations
Reporting frameworks		
20 Critical Questions, Vollmer et al [17]	Proposes 20 questions for evaluating the development and use of AI in research (20 reporting items)	Investigators, clinicians, patients, policymakers
Comprehensive Checklist, Cabitza and Campagner [38]	Proposes a comprehensive checklist for the self-assessment and evaluation of medical papers (30 reporting items)	Investigators, editors and peer reviewers
CONSORT[b]*-AI*, Liu et al [20][a]	Provides reporting guidelines for clinical trials evaluating interventions with an AI component (25 core and 15 AI-specific reporting items)	AI developers, investigators
CAIR[c]*Checklist*, Olczak et al [39]	Provides guidelines and an associated checklist for the reporting of AI research to clinicians (15 reporting items)	Investigators, developers, clinicians
DECIDE-AI, Vasey et al [33][a]	Provides reporting guidelines for evaluations of early-stage clinical decision support systems developed using AI (10 generic and 17 AI-specific reporting items)	Investigators, clinicians, patients, policymakers
Guidelines for Developing and Reporting, Luo et al [31]	Provides guidelines for applying and reporting AI model specifications/results in biomedical research (12 reporting items)	AI developers, investigators
MINIMAR[d], Hernandez-Boussard et al [40]	Provides minimum reporting standards for AI in health care (16 reporting items)	AI developers, investigators
SPIRIT[e]*-AI*, Rivera et al [2][a]	Provides guidelines for clinical trials protocols evaluating interventions with an AI component (25 core and 15 AI-specific reporting items)	AI developers, investigators

[a] Publication associated with a professional organization; AI in Healthcare=National Academy of Medicine; CONSORT-AI=CONSORT Group; DECIDE-AI=DECIDE-AI Expert Group; SPIRIT-AI=SPIRIT Group.

[b] CONSORT: Consolidated Standards of Reporting Trials.

[c] CAIR: Clinical AI Research.

[d] MINIMAR: Minimum Information for Medical AI Reporting.

[e] SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials.

Descriptive Frameworks

AI in Health Care

Matheny and colleagues [32] synthesized current knowledge related to the accountable development, application, and maintenance of AI in health care. This narrative describes existing and upcoming Al solutions, and underscores current challenges, limitations, and best practices for AI development, implementation, and maintenance.

Clinician Checklist

Scott and colleagues [34] proposed a checklist to evaluate the potential impact on clinical decision-making and patient outcomes of emerging machine-learning algorithms. Targeted toward clinicians, the checklist has been tailored for nonexperts, and provides a brief background of relevant machine-learning concepts and examples. The checklist addresses issues such as validity, utility, feasibility, safety, and ethical use.

Ethical Considerations

Char and colleagues [36] outlined a systematic approach for addressing ethical concerns surrounding machine-learning health care applications, and highlighted the need for interdisciplinary collaboration of diverse stakeholders. Evaluation and oversight tasks are described at each stage of the machine-learning pipeline from conception to implementation. Key questions and ethical considerations address common concerns found through a literature search as well as considerations that have received less attention.

Evaluating AI

Park and colleagues [37] highlighted the need for real-word evaluations of AI applications in health care. They present the phases of clinical trials for drugs and medical devices along with how AI applications could be evaluated in a similar manner. For each phase (including discovery and invention, technical performance and safety, efficacy and side effects, therapeutic efficacy, and safety and effectiveness), they propose appropriate study designs and methods for AI evaluation.

Users' Guide

Liu and colleagues [19] presented a users' guide to inform primarily clinicians about the major principles of machine learning. They describe the need for effective machine-learning model validation, review basic machine learning concepts, and provide recommendations on effective ways to implement machine-learning models in clinical medicine.

Reporting and Implementing Interventions

After presenting clinical examples of beneficial AI use, Bates and colleagues [35] discuss three major bottlenecks slowing the adoption of AI and machine-learning technologies in health care: methodological issues in evaluating AI-based interventions, the need for standards in reporting, and institution hurdles. They also highlight the role of FDA regulation and consider the need for rapid innovation in AI development.

Reporting Frameworks

20 Critical Questions

Vollmer and colleagues [17] provided a set of 20 questions focused on improving the transparency, replicability, ethics, and effectiveness of AI methods in health care. Statutory regulators and members of national advisory bodies and academic organizations, mostly from the United Kingdom and United States, collaboratively developed the questions.

Comprehensive Checklist

Cabitza and Campagner [38] proposed an extensive 30-item checklist to assess the quality of medical machine-learning studies. The checklist has been formatted both for authors to evaluate their own contributions and for reviewers to indicate where revisions may be necessary, and is organized in six phases: problem understanding, data understanding, data preparation, modeling, validation, and deployment.

CONSORT-AI

Liu and colleagues [20] extended the CONSORT (Consolidated Standards for Reporting Trials) framework to include additional considerations for the reporting of AI trials. The primary purpose of the extension is to facilitate the transparent reporting of interventional trials using AI, and the reporting checklist also provides some guidance for the development and critical appraisal of AI intervention studies.

CAIR Checklist

Olczak and colleagues [39] proposed a checklist for reporting medical AI research to clinicians and other stakeholders. They describe common performance and outcome measures that clinicians should be familiar with, and incorporate guidance about which metrics should be presented at each stage of a manuscript into the checklist. They also address ethical considerations that arise from AI use in health care.

DECIDE-AI

Vasey and colleagues [33] presented reporting guidelines for early-stage clinical trials of AI decision-support systems. The checklist focuses on four key aspects: proof of clinical utility, safety, the evaluation of human factors, and preparation for larger trials. This checklist was developed through a consensus process involving 151 experts and 20 stakeholder groups.

Guidelines for Developing and Reporting

Luo and colleagues [31] generated a set of guidelines on reporting machine-learning predictive models in biomedical research. The objective of these guidelines is to provide best practices for AI in biomedical research. This framework includes a list of minimum reporting items to be included in research manuscripts and a set of recommendations for optimal use of predictive models.

MINIMAR

Hernandez-Boussard and colleagues [40] proposed a list of minimum information that should be reported for all medical AI technologies. This list is intended to promote broader discussion and help inform extensions to other checklists. The four essential components in their guidelines include study population and setting, patient demographic characteristics, model architecture, and model evaluation.

SPIRIT-AI

Rivera and colleagues [2] presented reporting guidelines to evaluate clinical trial protocols involving interventions with an AI component. The purpose of the guidelines is to promote transparency and comprehensiveness for clinical trials with AI interventions. The guidelines were developed as AI extensions to the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) and CONSORT guidelines.

Content Domains

Overview of Domains

We identified five domains through the content analysis, including transparency, reproducibility, ethics, effectiveness, and engagement. These domains are described in turn below. Table 2 depicts each framework's coverage of content domains across translational stages. Figure 1 depicts the coverage of each individual framework and Figure 2 presents the aggregate coverage of frameworks as a heatmap.

Table 2. Coverage of frameworks across content domains and translational stages.

Domain and stage	Descriptive frameworks						Reporting frameworks							
	AI[a] in health care	Clinician Checklist	Ethical Considerations	Evaluating AI	Users' Guide	Reporting and Implementing Interventions	20 Critical Questions	Comprehensive Checklist	CONSORT[b]-AI	CAIR[c] Checklist	DECIDE-AI	Guidelines for Developing and Reporting	MINIMAR[d]	SPIRIT[e]-AI
Transparency														
Development	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Validation		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Reporting	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Implementation		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Surveillance	✓			✓			✓	✓						
Reproducibility														
Development			✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Validation	✓	✓	✓	✓	✓	✓	✓				✓		✓	
Reporting		✓				✓	✓	✓	✓	✓	✓	✓	✓	✓
Implementation		✓			✓	✓	✓	✓	✓		✓	✓		✓
Surveillance							✓							
Ethics														
Development	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓
Validation	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓
Reporting	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓
Implementation	✓	✓	✓	✓	✓	✓	✓	✓		✓				✓
Surveillance			✓	✓										
Effectiveness														
Development	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Validation		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Reporting	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Implementation	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		
Surveillance	✓		✓	✓	✓		✓	✓						
Engagement														
Development	✓		✓				✓				✓			
Validation														
Reporting			✓								✓			✓
Implementation							✓	✓						
Surveillance	✓		✓											

[a] AI: artificial intelligence.

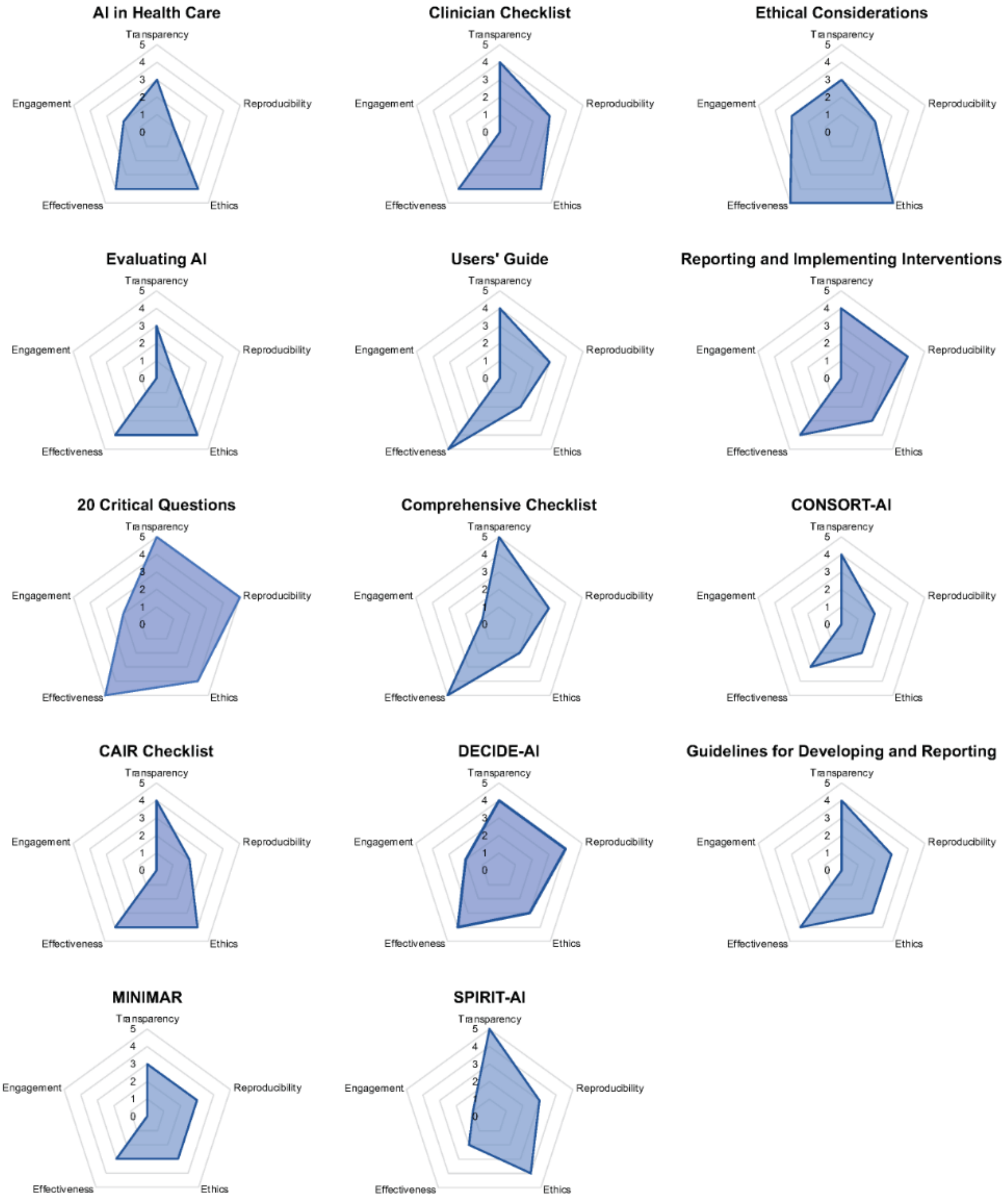
[b] CONSORT: Consolidated Standards for Reporting Trials.

[c] CAIR: Clinical AI Research.

[d] MINIMAR: Minimum Information for Medical AI Reporting.

[e] SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials.

Figure 1. Coverage of frameworks across content domains. AI: artificial intelligence; CAIR: Clinical AI Research; CONSORT: Consolidated Standards of Reporting Trials; MINIMAR: Minimum Information for Medical AI Reporting; SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials.



AI in Health Care

Clinician Checklist

Ethical Considerations

Evaluating AI

Users' Guide

Reporting and Implementing Interventions

20 Critical Questions

Comprehensive Checklist

CONSORT-AI

CAIR Checklist

DECIDE-AI

Guidelines for Developing and Reporting

MINIMAR

SPIRIT-AI

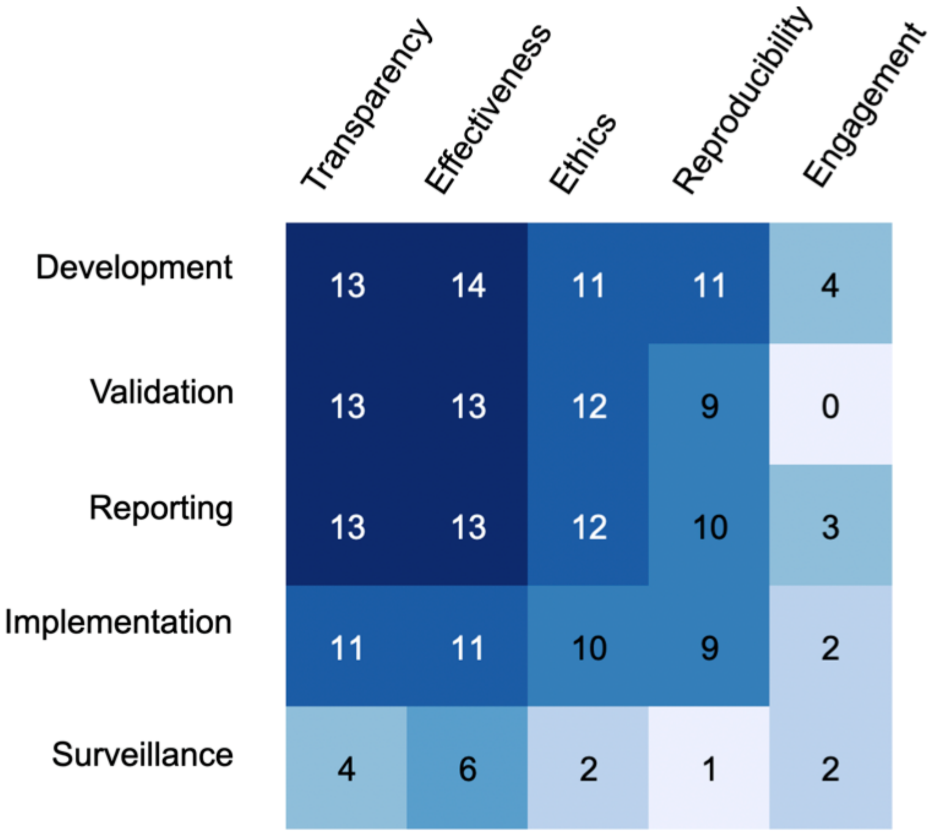


Figure 2. Heatmap of the frameworks' coverage across the five stages of translation. Darker boxes indicate areas where more frameworks offered guidance, whereas lighter boxes indicate areas where fewer frameworks offered guidance.

	Transparency	Effectiveness	Ethics	Reproducibility	Engagement
Development	13	14	11	11	4
Validation	13	13	12	9	0
Reporting	13	13	12	10	3
Implementation	11	11	10	9	2
Surveillance	4	6	2	1	2

Transparency

Transparency describes how openly and thoroughly information is disclosed to the public and the scientific community [41]. Transparency allows for independent evaluation of an AI algorithm's predictive power [42]. Involving stakeholders to help identify errors and bias in development or implementation also requires transparency [43]. Health care providers need transparency to interpret and justify medical decisions that result from AI use.

All but one framework (*Evaluating AI)* provided input on transparency with regard to the development and reporting of AI. Only four frameworks (*AI in Health Care*, *Comprehensive Checklist*, *Evaluating AI*, *20 Critical Questions)* commented on transparency with regard to surveillance. Two frameworks (*20 Critical Questions*, *Comprehensive Checklist*) commented on transparency in regard to all five stages of translation. The number of translational stages considered for transparency ranged from 3 to 5, with an average score of 3.9 across all frameworks. On average, descriptive frameworks discussed transparency in regard to fewer stages of translation than reporting frameworks (3.5 vs 4.1).

Reproducibility

Reproducibility describes how likely it is that others could develop or apply an AI tool with similar results. Reproducibility is a basic tenet of good scientific practice [44]. The ability to reproduce AI models is key to external validation [45]. Reproducibility accounts for burdens such as costs and high computational needs. Reproducibility in implementation and surveillance is necessary to improve the widespread, equitable use of AI.

All frameworks commented on reproducibility. Only one (*20 Critical Questions)* commented on reproducibility in regard to all five stages of translation, and this was also the only framework to comment on reproducibility in regard to the surveillance of AI. Most frameworks described reproducibility in relation to the validation, reporting, and implementation of AI. Scores for reproducibility ranged from 1 to 5 with a mean score of 2.9 across all frameworks. On average, descriptive frameworks discussed reproducibility in regard to fewer stages of translation than reporting frameworks (2.3 vs 3.3).

Ethics

Ethics considers values such as benevolence, fairness, respect for autonomy, and privacy. Such values are essential to avoiding harm and ensuring societal benefit in AI use [46]. Ethical practice for the use of AI in medicine relies on collaboration with ethicists, social scientists, and regulators. Racial, gender, and insurance provider biases are the largest ethical concerns with AI use [47].

Only one framework commented on ethical considerations across all stages of translation (*Ethical Considerations)*. Four frameworks (*AI in Health Care*, *20 Critical Questions*,*CAIR Checklist*, and *SPIRIT-AI)* addressed ethical considerations for development, validation, reporting, and implementation, and one tool addressed ethical considerations for development, validation, implementation, and surveillance (*Evaluating AI*). Scores for ethics ranged from 2 to 5 with a mean score of 3.4. On average, descriptive frameworks discussed ethics in regard to more stages of translation than reporting frameworks (3.7 vs 3.2).

Effectiveness

Effectiveness describes the success and efficiency of models and methods when they are applied in a given context. Effectiveness is concerned with matters such as data quality and model fit during the development of AI models [48]. External validation helps ensure effective discrimination and calibration to prevent overfitting [49]. Measures of effectiveness should be clearly and consistently reported [20,48]. There is a lack of appropriate benchmarks and standards of care to accurately measure the clinical benefit of many AI models [50]. Strategies are needed to continually measure effectiveness after implementation [17].

Four frameworks (*Ethical Considerations, Users' Guide, 20 Critical Questions, Comprehensive Checklist*) commented on effectiveness across all translational stages. All frameworks reported on effectiveness as a consideration for the development of medical AI. All but one framework (*AI in Healthcare*) reported on effectiveness during validation. Six frameworks commented on effectiveness as a consideration for surveillance (*AI in Health Care, Ethical Considerations, Evaluating AI, Users' Guide, 20 Critical Questions, Comprehensive Checklist*). Scores for effectiveness ranged from 3 to 5 with a mean score of 4.1. On average, descriptive frameworks discussed ethics in regard to more stages of translation than reporting frameworks (4.3 vs 3.9).

Engagement

Engagement explores to what extent the opinions and values of patients and other end users or stakeholders are collected and accounted for in decision-making. The degree of engagement can range from consultation (lowest level) to partnership and shared leadership [17]. In health research, using engagement approaches has been demonstrated to increase study enrollment, improve data quality, and improve the relevance of research design and conduct [51]. Patient engagement can also improve the quality and efficiency of health care, and reduce costs [52].

No frameworks considered engagement across all five stages. Engagement was discussed in relation to development by four frameworks (*AI in Health Care, Ethical Considerations, 20 Critical Questions, DECIDE-AI*) and in relation to reporting by three frameworks (*Ethical Considerations, DECIDE-AI, SPIRIT-AI*). No frameworks explored engagement in the validation stage of translation. Scores for engagement ranged from 0 to 3 with a mean of 0.8, which did not differ across descriptive and reporting frameworks.

Discussion

Principal Findings

Frameworks for applying and evaluating AI in medicine are rapidly emerging and address important considerations for the oversight of AI, such as those regarding transparency, reproducibility, ethics, and effectiveness. Providing guidance on integrating stakeholder engagement to inform AI is not a current strength of frameworks. Frameworks included in this review were the least likely to provide guidance on using engagement to inform the translation of AI in comparison to other considerations. The relative paucity of guidance on engagement reflects the larger AI landscape, which does not actively engage diverse end users in the translation of AI. For many stakeholders, AI remains a black box [53,54].

More than half of the frameworks provided reporting guidance on the use of AI in medicine. Additionally, nearly all frameworks in this review were published in 2019 or later. Given the rapid expansion of the field, it is essential to assess the consistency of recommendations across reporting frameworks to build shared understanding.

A near-miss in this review was the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) Statement [55], which provides reporting guidelines for studies using prediction models for diagnosis or prognosis. As this framework is often used to evaluate AI models, we did evaluate its content and found that it offered comments on transparency, reproducibility, and effectiveness in the translational stages of development, validation, and reporting. It also provided considerations for ethics in the validation of models, but not in other translational stages. It did not pose any guidance on the use of engagement. A TRIPOD-AI [18] extension is forthcoming, which is engaging diverse stakeholders in its development. We hope that the guidelines themselves will recommend the use of end-user engagement.

The content domains and stages of translation that we have considered are far from exhaustive, and there are many other features and specific stages of AI development, application, and evaluation that are worthy of discussion. For instance, as the scope of AI in medicine expands, it will require broadened evaluation. For instance, there have been few economic evaluations of AI tools in medicine, which may be a barrier to their implementation [56]. Another form of evaluation might include the use of randomized controlled trials to assess the efficacy of tools in clinical contexts. Another consideration is regarding conflicts of interest, and it will be important to establish approaches to evaluate and mitigate potential conflicts of interest.

None of the frameworks included in this review used an explicit translational science lens to provide explicit guidance across the AI life cycle. Having resources that detail considerations for AI application and evaluation at each stage of the translational process would be helpful for those seeking to develop AI with meaningful medical applications. Resources that could be helpful would include patient/community-centered educational resources about the value of AI, a framework to optimize the patient-centered translation of AI predictive analytics into clinical decision-making, and critical appraisal tools for use in comparing different applications of AI to inform medical decision-making.

There was a paucity of guidance regarding the surveillance of AI in medicine. Although some research has described the use of AI to inform primarily public health surveillance [57,58], little work—even outside of the frameworks included in this review—has provided specific guidance on how to surveil the use of AI with medical applications. Existing recommendations for the surveillance of pharmaceutical and other medical interventions might be applicable to AI, but tailored recommendations will also be needed. It is likely that surveillance will need to be an ongoing process to provide up-to-date information on how AI tools perform in light of new clinical information and research, and to recalibrate AI tools to incorporate this knowledge into clinical predictions [59].

The goal of the framework evaluations was not intended to reflect the *quality* of the frameworks but rather to indicate the *coverage* of AI guidance either at the individual framework level (Figure 1) or across the literature (Figure 2). These evaluations could be used as a quick reference for clinicians, developers, patients, and others to identify which framework(s) may provide the most relevant recommendations to their specific AI application. For instance, *CONSORT-AI* was specifically developed as a checklist to inform the reporting of AI research. Although it had the lowest overall score, it provided recommendations for reporting relevant to four out of the five considerations raised in this review.

The field of AI in medicine could stand to learn from the clearer methodological standards and best practices currently existent in established fields such as patient-centered outcomes research (PCOR) [51,60]. PCOR works to advance the quality and relevance of evidence about how to prevent, diagnose, treat, monitor, and manage health care; this evidence helps patients, caregivers, clinicians, policymakers, and other health care stakeholders make better decisions. The translation of AI in medicine lacks the user-centeredness that is central to PCOR [61]. At a minimum, AI for use in medicine should be developed by multidisciplinary teams, where stakeholders from relevant fields (eg, bioinformatics, specific medical specialties, patient experience) offer their expertise to inform the development of a given AI application. Ideally, more integrated transdisciplinary approaches, wherein stakeholders from relevant fields collectively create shared knowledge that transcends their individual disciplines, would be used to develop AI. Using a transdisciplinary approach has the potential to create AI that is technically robust, provides clinically relevant information, and can be easily integrated into the clinical workflow to inform patient and clinician decision-making.

Conclusion

There is a growing literature offering input on the oversight of AI in medicine, with more guidance from regulatory bodies such as the US FDA forthcoming. Although existing frameworks provide general coverage of considerations for the oversight of AI in medicine, they fall short in their ability to offer input on the use of engagement in the development of AI, as well as in providing recommendations for the specific translational stage of surveilling AI. Frameworks should emphasize engaging patients, clinicians, and other end users in the development, use, and evaluation of AI in medicine.

Acknowledgments

The project described was supported by Award Number UL1TR002733 from the National Center For Advancing Translational Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center For Advancing Translational Sciences or the National Institutes of Health.

Authors' Contributions

NLC, SBB, and JFPB participated in study design. NLC, ME, JP, and JFPB participated in data collection, analysis, and in identification of data. All authors participated in writing of the report. All authors have reviewed and approved submission of this article.

Conflicts of Interest

None declared.

References

1. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017 Apr;69S:S36-S40. [doi: 10.1016/j.metabol.2017.01.011] [Medline: 28126242]
2. Rivera SC, Liu X, Chan A, Denniston AK, Calvert MJ, SPIRIT-AICONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ 2020 Sep 09;370:m3210 [FREE Full text] [doi: 10.1136/bmj.m3210] [Medline: 32907797]
3. Artificial Intelligence in Healthcare Market by Offering, Technology, Application, End User and Geography - Global Forecast to 2027. ReportLinker. 2021 Oct. URL: https://tinyurl.com/4dh7bdn7 [accessed 2021-07-14]
4. Chan HS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing drug discovery via artificial intelligence. Trends Pharmacol Sci 2019 Aug;40(8):592-604. [doi: 10.1016/j.tips.2019.06.004] [Medline: 31320117]
5. Amisha PM, Malik P, Pathania M, Rathaur V. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019 Jul;8(7):2328-2331 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_440_19] [Medline: 31463251]
6. Amarasingham R, Audet A, Bates D, Glenn Cohen I, Entwistle M, Escobar GJ, et al. Consensus Statement on Electronic Health Predictive Analytics: a guiding framework to address challenges. EGEMS 2016;4(1):1163 [FREE Full text] [doi: 10.13063/2327-9214.1163] [Medline: 27141516]

7. Matheny ME, Whicher D, Thadaney Israni S. Artificial intelligence in health care: a report from the National Academy of Medicine. JAMA 2020 Feb 11;323(6):509-510. [doi: 10.1001/jama.2019.21579] [Medline: 31845963]
8. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4] [Medline: 31044724]
9. Obermeyer Z, Nissan R, Stern M, Eaneff S, Bembeneck E, Mullainathan S. Algorithmic Bias Playbook. Chicago Booth The Center for Applied Artificial Intelligence. 2021. URL: https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias [accessed 2022-08-05]
10. Dhindsa K, Bhandari M, Sonnadara R. What's holding up the big data revolution in healthcare? BMJ 2018 Dec 28;363:k5357. [doi: 10.1136/bmj.k5357] [Medline: 30593447]
11. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015 Jan 06;162(1):W1-W73 [FREE Full text] [doi: 10.7326/M14-0698] [Medline: 25560730]
12. Wehling M. Principles of translational science in medicine: from bench to bedside. Cambridge, MA: Academic Press; 2015.
13. Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? Genet Med 2007 Oct;9(10):665-674 [FREE Full text] [doi: 10.1097/GIM.0b013e31815699d0] [Medline: 18073579]
14. Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, et al. A path for translation of machine learning products into healthcare delivery. EMJ Innov 2020 Jan 27;10:00172 [FREE Full text] [doi: 10.33590/emjinnov/19-00172]
15. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. NPJ Digit Med 2019 Jul 26;2(1):69. [doi: 10.1038/s41746-019-0148-3] [Medline: 31372505]
16. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct 10;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
17. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020 Mar 20;368:l6927 [FREE Full text] [doi: 10.1136/bmj.l6927] [Medline: 32198138]
18. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021 Jul 09;11(7):e048008 [FREE Full text] [doi: 10.1136/bmjopen-2020-048008] [Medline: 34244270]
19. Liu Y, Chen PC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. JAMA 2019 Nov 12;322(18):1806-1816. [doi: 10.1001/jama.2019.16489] [Medline: 31714992]
20. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020 Sep 09;370:m3164 [FREE Full text] [doi: 10.1136/bmj.m3164] [Medline: 32909959]
21. Artificial Intelligence and Machine Learning in Software as a Medical Device. US Food and Drug Administration. 2021. URL: https://tinyurl.com/2beyvzfp [accessed 2022-08-05]
22. White Paper on Artificial Intelligence: a European approach to excellence and trust. European Commission. 2020 Feb 19. URL: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en [accessed 2022-08-05]
23. Minssen T, Gerke S, Aboy M, Price N, Cohen G. Regulatory responses to medical machine learning. J Law Biosci 2020;7(1):lsaa002 [FREE Full text] [doi: 10.1093/jlb/lsaa002] [Medline: 34221415]
24. Regulatory Science Strategy to 2025. European Medicines Agency. 2020. URL: https://www.ema.europa.eu/en/about-us/how-we-work/regulatory-science-strategy#regulatory-science-strategy-to-2025-section [accessed 2022-08-05]
25. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov 01;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
26. Devi N. Understanding the qualitative and quantitative methods in the context of content analysis. 2009 Presented at: Qualitative and Quantitative Methods in Libraries International Conference; 2009; Chania, Crete, Greece.
27. Bordier M, Delavenne C, Nguyen DTT, Goutard FL, Hendrikx P. One Health Surveillance: a matrix to evaluate multisectoral collaboration. Front Vet Sci 2019 Apr 24;6:109. [doi: 10.3389/fvets.2019.00109] [Medline: 31106210]
28. Andrews E, Berghofer K, Long J, Prescott A, Caboral-Stevens M. Satisfaction with the use of telehealth during COVID-19: An integrative review. Int J Nurs Stud Adv 2020 Nov;2:100008 [FREE Full text] [doi: 10.1016/j.ijnsa.2020.100008] [Medline: 33083791]
29. Bryant-Lukosius D, Spichiger E, Martin J, Stoll H, Kellerhals SD, Fliedner M, et al. Framework for evaluating the impact of advanced practice nursing roles. J Nurs Scholarsh 2016 Mar 11;48(2):201-209. [doi: 10.1111/jnu.12199] [Medline: 26869323]

30. Oliveira R, Ignacio C, de Moraes Neto AHA, de Lima Barata MM. Evaluation matrix for health promotion programs in socially vulnerable territories. Cien Saude Colet 2017 Dec;22(12):3915-3932 [FREE Full text] [doi: 10.1590/1413-812320172212.24912017] [Medline: 29267709]
31. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
32. Matheny M, Israni S, Ahmed M, Whicher D. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: National Academy of Medicine; 2019.
33. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, DECIDE-AI expert group. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 2022 May 18;377:e070904 [FREE Full text] [doi: 10.1136/bmj-2022-070904] [Medline: 35584845]
34. Scott I, Carter S, Coiera E. Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health Care Inform 2021 Feb 05;28(1):e100251 [FREE Full text] [doi: 10.1136/bmjhci-2020-100251] [Medline: 33547086]
35. Bates DW, Auerbach A, Schulam P, Wright A, Saria S. Reporting and implementing interventions involving machine learning and artificial intelligence. Ann Intern Med 2020 Jun 02;172(11 Suppl):S137-S144 [FREE Full text] [doi: 10.7326/M19-0872] [Medline: 32479180]
36. Char DS, Abràmoff MD, Feudtner C. Identifying ethical considerations for machine learning healthcare applications. Am J Bioeth 2020 Nov 26;20(11):7-17 [FREE Full text] [doi: 10.1080/15265161.2020.1819469] [Medline: 33103967]
37. Park Y, Jackson G, Foreman M, Gruen D, Hu J, Das A. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open 2020 Oct;3(3):326-331 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa033] [Medline: 33215066]
38. Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies. Int J Med Inform 2021 Sep;153:104510. [doi: 10.1016/j.ijmedinf.2021.104510] [Medline: 34108105]
39. Olczak J, Pavlopoulos J, Prijs J, Ijpma FFA, Doornberg JN, Lundström C, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. Acta Orthop 2021 Oct 14;92(5):513-525 [FREE Full text] [doi: 10.1080/17453674.2021.1918389] [Medline: 33988081]
40. Hernandez-Boussard T, Bozkurt S, Ioannidis J, Shah N. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc 2020 Dec 09;27(12):2011-2015 [FREE Full text] [doi: 10.1093/jamia/ocaa088] [Medline: 32594179]
41. 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]
42. Benda N, Das L, Abramson E, Blackburn K, Thoman A, Kaushal R, et al. "How did you get to this number?" Stakeholder needs for implementing predictive analytics: a pre-implementation qualitative study. J Am Med Inform Assoc 2020 May 01;27(5):709-716 [FREE Full text] [doi: 10.1093/jamia/ocaa021] [Medline: 32159774]
43. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan 7;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
44. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Massive Analysis Quality Control (MAQC) Society Board of Directors, Waldron L, et al. Transparency and reproducibility in artificial intelligence. Nature 2020 Oct 14;586(7829):E14-E16 [FREE Full text] [doi: 10.1038/s41586-020-2766-y] [Medline: 33057217]
45. Reps J, Schuemie M, Suchard M, Ryan P, Rijnbeek P. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc 2018 Aug 01;25(8):969-975 [FREE Full text] [doi: 10.1093/jamia/ocy032] [Medline: 29718407]
46. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep 19;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6] [Medline: 31427808]
47. Chen I, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? AMA J Ethics 2019 Feb 01;21(2):E167-E179 [FREE Full text] [doi: 10.1001/amajethics.2019.167] [Medline: 30794127]
48. Chen PC, Liu Y, Peng L. How to develop machine learning models for healthcare. Nat Mater 2019 May 18;18(5):410-414. [doi: 10.1038/s41563-019-0345-0] [Medline: 31000806]
49. Van Calster B, Wynants L, Timmerman D, Steyerberg E, Collins G. Predictive analytics in health care: how can we know it works? J Am Med Inform Assoc 2019 Dec 01;26(12):1651-1654 [FREE Full text] [doi: 10.1093/jamia/ocz130] [Medline: 31373357]
50. Parikh RB, Kakad M, Bates DW. Integrating predictive analytics into high-value care: the dawn of precision delivery. JAMA 2016 Feb 16;315(7):651-652. [doi: 10.1001/jama.2015.19417] [Medline: 26881365]
51. Sheridan S, Schrandt S, Forsythe L, Hilliard TS, Paez KA, Advisory Panel on Patient Engagement (2013 inaugural panel). The PCORI Engagement Rubric: promising practices for partnering in research. Ann Fam Med 2017 Mar 13;15(2):165-170 [FREE Full text] [doi: 10.1370/afm.2042] [Medline: 28289118]

52. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. Health Aff 2013 Feb;32(2):223-231. [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]
53. Brill SB, Moss KO, Prater L. Transformation of the doctor-patient relationship: big data, accountable care, and predictive health analytics. HEC Forum 2019 Dec 17;31(4):261-282. [doi: 10.1007/s10730-019-09377-5] [Medline: 31209679]
54. Bjerring JC, Busch J. Artificial intelligence and patient-centered decision-making. Philos Technol 2020 Jan 08;34(2):349-371. [doi: 10.1007/s13347-019-00391-6]
55. Collins G, Reitsma J, Altman D, Moons K. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015 Feb;102(3):148-158. [doi: 10.1002/bjs.9736] [Medline: 25627261]
56. Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic review of health economic evaluations focused on artificial intelligence in healthcare: The Tortoise and the Cheetah. Value Health 2022 Mar;25(3):340-349 [FREE Full text] [doi: 10.1016/j.jval.2021.11.1362] [Medline: 35227444]
57. Thiébaut R, Cossin S, Section Editors for the IMIA Yearbook Section on Public HealthEpidemiology Informatics. Artificial Intelligence for Surveillance in Public Health. Yearb Med Inform 2019 Aug 16;28(1):232-234 [FREE Full text] [doi: 10.1055/s-0039-1677939] [Medline: 31419837]
58. Neill DB. New directions in artificial intelligence for public health surveillance. IEEE Intell Syst 2012 Jan;27(1):56-59. [doi: 10.1109/mis.2012.18]
59. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearb Med Inform 2019 Aug 25;28(1):128-134 [FREE Full text] [doi: 10.1055/s-0039-1677903] [Medline: 31022752]
60. Esmail LC, Barasky R, Mittman BS, Hickam DH. Improving comparative effectiveness research of complex health interventions: standards from the Patient-Centered Outcomes Research Institute (PCORI). J Gen Intern Med 2020 Nov 26;35(Suppl 2):875-881 [FREE Full text] [doi: 10.1007/s11606-020-06093-6] [Medline: 33107006]
61. Shaoibi A, Neelon B, Lenert LA. Shared decision making: from decision science to data science. Med Decis Making 2020 Feb 06;40(3):254-265. [doi: 10.1177/0272989x20903267]

Abbreviations

AI: artificial intelligence
CONSORT: Consolidated Standards for Reporting Trials
FDA: Food and Drug Administration
PCOR: patient-centered outcomes research
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

Edited by G Eysenbach; submitted 26.01.22; peer-reviewed by D Smithard, I Scott; comments to author 10.03.22; revised version received 02.06.22; accepted 14.07.22; published 25.08.22.

Please cite as:
Crossnohere NL, Elsaid M, Paskett J, Bose-Brill S, Bridges JFP
Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks
J Med Internet Res 2022;24(8):e36823
URL: https://www.jmir.org/2022/8/e36823
doi:10.2196/36823
PMID:36006692

Review

The Measurement of Dose and Response for Smoking Behavior Change Interventions in the Digital Age: Systematic Review

Megumi Ichimiya[1], MSPH; Raquel Gerard[1], BA; Sarah Mills[2], BA; Alexa Brodsky[1], BA; Jennifer Cantrell[3], DrPH; W Douglas Evans[1], PhD

[1]Department of Prevention and Community Health, Milken Institute School of Public Health, The George Washington University, Washington, DC, United States

[2]Schroeder Institute, Truth Initiative, Washington, DC, United States

[3]Department of Social and Behavioral Sciences, School of Global Public Health, New York University, New York, NY, United States

Corresponding Author:
Megumi Ichimiya, MSPH
Department of Prevention and Community Health
Milken Institute School of Public Health
The George Washington University
950 New Hampshire Ave, NW
Washington, DC, 20052
United States
Phone: 1 202 994 7400
Email: imegumi7@gwu.edu

Abstract

Background: There is little consensus regarding effective digital health interventions for diverse populations, which is due in part to the difficulty of quantifying the impact of various media and content and the lack of consensus on evaluating dosage and outcomes. In particular, digital smoking behavior change intervention is an area where consistency of measurement has been a challenge because of emerging products and rapid policy changes. This study reviewed the contents and outcomes of digital smoking interventions and the consistency of reporting to inform future research.

Objective: This study aims to systematically review digital smoking behavior change interventions and evaluate the consistency in measuring and reporting intervention contents, channels, and dose and response outcomes.

Methods: PubMed, Embase, Scopus, PsycINFO, and PAIS databases were used to search the literature between January and May 2021. General and journal-based searches were combined. All records were imported into Covidence systematic review software (Veritas Health Innovation) and duplicates were removed. Titles and abstracts were screened by 4 trained reviewers to identify eligible full-text literature. The data synthesis scheme was designed based on the concept that exposure to digital interventions can be divided into intended doses that were planned by the intervention and enacted doses that were completed by participants. The intended dose comprised the frequency and length of the interventions, and the enacted dose was assessed as the engagement. Response measures were assessed for behaviors, intentions, and psychosocial outcomes. Measurements of the dose-response relationship were reviewed for all studies.

Results: A total of 2916 articles were identified through a database search. Of these 2916 articles, the title and abstract review yielded 324 (11.11%) articles for possible eligibility, and 19 (0.65%) articles on digital smoking behavior change interventions were ultimately included for data extraction and synthesis. The analysis revealed a lack of prevention studies (0/19, 0%) and dose-response studies (3/19, 16%). Of the 19 studies, 6 (32%) reported multiple behavioral measures, and 5 (23%) reported multiple psychosocial measures as outcomes. For dosage measures, 37% (7/19) of studies used frequency of exposure, and 21% (4/19) of studies mentioned the length of exposure. The assessment of clarity of reporting revealed that the duration of intervention and data collection tended to be reported vaguely in the literature.

Conclusions: This review revealed a lack of studies assessing the effects of digital media interventions on smoking outcomes. Data synthesis showed that measurement and reporting were inconsistent across studies, illustrating current challenges in this field. Although most studies focused on reporting outcomes, the measurement of exposure, including intended and enacted doses, was unclear in a large proportion of studies. Clear and consistent reporting of both outcomes and exposures is needed to develop further evidence in intervention research on digital smoking behavior change.

(J Med Internet Res 2022;24(8):e38470) doi:10.2196/38470

KEYWORDS

digital health; digital media; social media; behavior change interventions; smoking; vaping; dose-response; telehealth; mobile health; mHealth; mobile phone

Introduction

Background

There are few published data on exposure to and evaluations of digital behavior change interventions, and to date, there is little consensus regarding effective interventions for diverse populations. This is because of the difficulty of quantifying digital health interventions that use various media and content and the lack of consensus on how to evaluate and report dose and outcomes. However, digital health is developing rapidly, and understanding the latest evidence is critical for research in this field.

Digital media is central to and an integral part of modern life; however, the study of its effects on health behavior is just beginning [1]. As reported by the World Bank, 45% of the world's population or 3.5 billion people use social media. Worldwide, the average user spends approximately 3 hours of their day on social media [2]. Given the widespread exposure to digital technologies such as social media, it is increasingly important to understand how digital media affects individual health decision-making and behavior, as well as social networks and communities. These facts make it critically important to understand how digital media influences behavior. Social and behavioral scientists who study digital media must learn how to design and evaluate effective behavior change interventions, the evidence-based approaches that are effective, and how digital media affects targeted outcomes.

Efforts to understand the practice of digital media interventions have historically been made by public health scientists. Hu [3] reviewed 348 journal articles and structured the subject, health topics, technologies, and methods used for digital interventions between 2008 and 2012. Abad et al [4] conducted a scoping review on digital public health surveillance and revealed that only 0.8% of the related studies between 2005 and 2020 deployed a digital health surveillance system that can be used for monitoring and targeted interventions, despite its impact on the study methodology and public health actions. A recent systematic review by Seiler et al [1] found relatively few rigorous studies on the effectiveness of digital media–related behavior change campaigns and interventions. This review also found that the reporting of design, measures, data collection, and other methods needs to be improved and systematized. Recommendations for improvements included clarification of what is meant by *dose* and *dose-response*; how and with what intensity interventions are delivered; and measurement of outcomes, including attitudes, beliefs, social norms, and health behavior. This review also reiterated a previous finding that evidence for behavior change using digital interventions stems primarily from studies conducted in high-income countries [5-9].

Although digital media has been used for a variety of public health programs, one of the most rapidly changing areas is digital smoking behavior change interventions. Owing to the emerging products, devices, and policy changes, it has been more difficult to systematically quantify and evaluate the effectiveness of interventions in this area. Despite the importance of using consistent measures and evaluation methods to understand the impacts of interventions, little effort has been made to understand the common measures and methods used to evaluate the effectiveness of these interventions. There is a need for an assessment of the measures used in this area to inform future research to accurately evaluate the outcomes across a variety of products and devices, as well as rapid policy and market change. To the authors' knowledge, no previous systematic review has focused on the detailed measurements used in digital smoking interventions.

Although tobacco use has declined overall [10], it remains at unacceptable levels, exacting personal and social costs, particularly among young adults [11,12]. Tobacco is the leading preventable cause of death in the United States [13]. In the United States, 18.2% of young adults aged 18 to 24 years reported current use of tobacco products, and 10.4% reported being current cigarette smokers [14]. Although the age at which smokers initiate cigarette use has been increasing over time, almost all cigarette use initiation occurs before the age of 26 years [12], making young adults a critical target for prevention efforts. Given the widespread use of digital media among the young adult population [15], digital interventions may be promising.

Objective

To address these gaps in digital smoking behavior change interventions, this review aimed to (1) systematically review and codify the measures used for digital health interventions in tobacco and nicotine use research, (2) evaluate the quality of reporting of dose and response, and (3) identify areas for improvement in the field.

Methods

This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The protocol was registered in PROSPERO (International Prospective Register of Systematic Reviews; ID CRD42021285655).

Study Search

The PubMed, Embase, Scopus, PsycINFO, and PAIS databases were used to search the literature. The search terms included "digital intervention," "health promotion," "health education," "health communication," "digital technology," "social media," "social marketing," "health," "measures," "methods," "frequency," "impression," and "reach." General and journal-based searches were combined to find literature that was directly related to the scope of the study. Specific journals that were searched included *The Lancet Digital Health*, *Journal of Medical Internet Research*, *Digital Health*, *Social Networks*, *npj Digital Medicine*, *Digital Medicine*, *Digital Biomarkers*, *Frontiers in Digital Health*, *Communication Methods and Measures*, *Health Technology Assessment*, *BMC Medical Research Methodology*, *Computers in Human Behavior*, *Computers in Biology and Medicine*, *Journal of Health Communication*, *Journal of Communication in Healthcare*, *Health Communication*, *Health Communication Science Digest*, *Health Education and Behavior*, *Digital Medicine*, *International Journal of Digital Healthcare*, *Journal of Health and Social Behavior*, *American Journal of Health Behavior*, *Journal of Behavioral Health*, *The Journal of Behavioral Health Services and Research*, *Health Behavior Research*, *American Journal of Health Promotion*, *Health Promotion Practice*, *Journal of Prevention and Health Promotion*, *Health Promotion International*, *International Journal of Health Promotion and Education*, and *Journal of Health Promotion and Behavior*. The literature was searched between January and May 2021, followed by a series of monthly searches to identify additional studies. Search strategies and terms were developed in collaboration with librarians at George Washington University.

All records searched through the database were imported into Covidence systematic review software (Veritas Health Innovation) and duplicates were removed.

Screening

Titles and abstracts were screened by 4 trained reviewers to identify eligible full-text literature. Each study was screened by 2 reviewers, and disagreements were resolved through discussion between the reviewers. The inclusion criteria were (1) publication after 2000, (2) full text available in English, (3) peer-reviewed original journal articles, (4) health-related topics, (5) at least one behavior change intervention defined, (6) use of the internet or mobile-based platform for mass or targeted communication, (7) use of digital devices, and (8) measurement of original data related to behavior or psychosocial measures.

Behavior change interventions were defined as planned programs that had stated objectives related to behavior change, target populations, and targeted messages in text, audio, video, graphics, or other distributed forms in a one-to-many format. Studies that assessed at least one behavioral or psychosocial measure defined in the social behavioral theories were included. Studies that only measured engagement were excluded. Digital media is defined as an internet-based platform for mass and targeted communications, including social media, apps, websites, software, blogging, and one-to-one chat platforms used for mass and targeted communications (eg, WhatsApp). Video games, emails, radio, and television were excluded from digital media. Studies that used digital media as a channel for one-to-one communication such as conversations between health care providers and patients were also excluded.

After literature was screened using the abovementioned criteria, articles related to smoking were extracted for this review. The terms used for this process included “tobacco,” “smoking,” “cigarette,” “vaping,” “vape,” “e-cig,” “ENDS,” “nicotine,” “hookah,” “JUUL,” “cigar,” “e-liquid,” “flavor,” “smokeless,” “smoker,” and “vaper.” The literature search was repeated replacing the term “health” with “smoking” and “vaping” on the database, and we checked for the coverage of the articles on the abovementioned topics.

Data Extraction and Synthesis

The data extraction and synthesis scheme was designed to identify the format of the digital media intervention, the measurement of each component that assessed dose and response in the intervention, and the study design used in the research. Dose measurement items were developed based on the concept that exposure to digital health interventions can be divided into 2 parts [16]. The first part is the intended dose, which refers to planned exposure by the intervention side. The other part is an enacted dose that corresponds to a portion of the intended dose that is actually completed by the participant. The dose measurement items were designed so that doses can be expressed as the frequency of the intervention multiplied by the length of intervention component and amount of engagement, which offers supplemental information about active involvement by participants. In this review, the intended dose comprises the frequency and length of the intervention, and the enacted dose comprises engagement. Engagement was defined as the interaction between the intervention content and the participants, such as views, clicks, likes, comments, and shares. Response equals the outcomes of the intervention, including behavior, intention, and other psychosocial factors that were previously confirmed to be connected in social behavioral change theories. A dose-response relationship was defined as the association between different levels of doses (exposure) and responses (outcomes), and its application was assessed in this review.

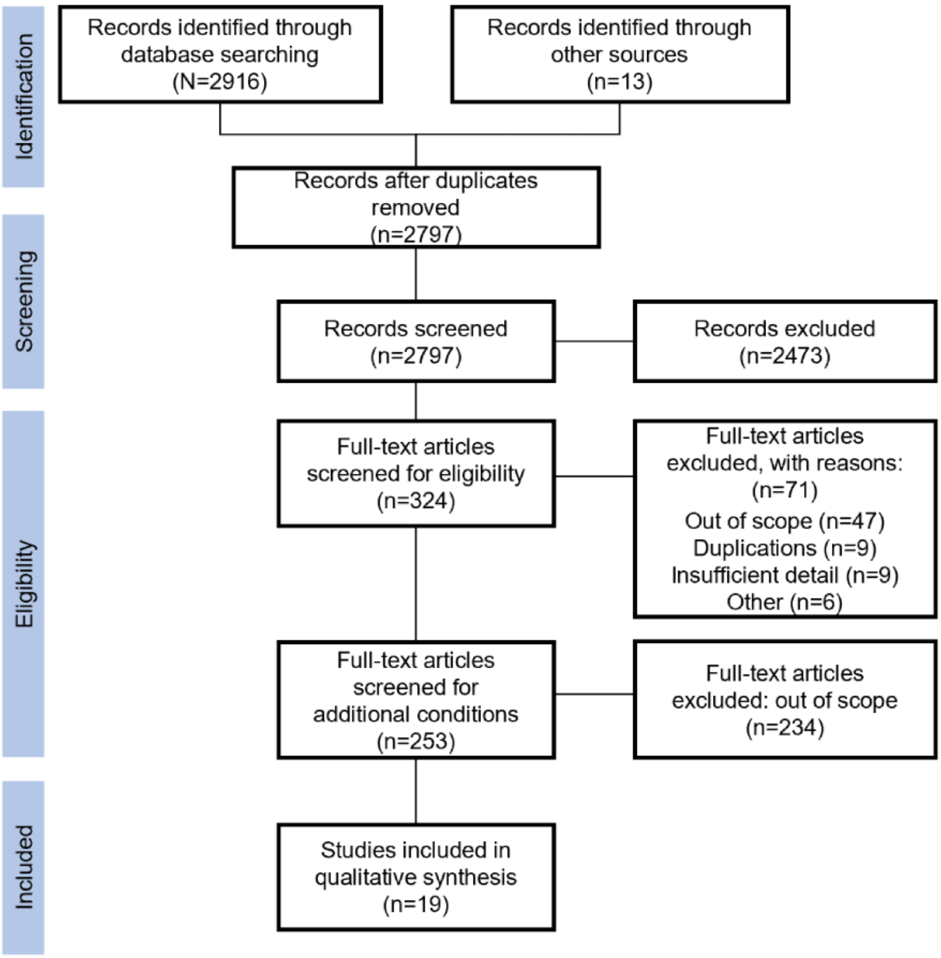
The codebook for data extraction and synthesis was developed in a Microsoft Excel format and piloted using 20 randomly selected articles on digital health behavior change interventions. The extracted data included basic information about the study, types of digital media and devices used, modes of intervention, measures used for intervention exposure, outcomes, engagement, study designs, model applications, cost and funding information, and source of bias. Data were converted into standardized forms where necessary and checked for clarity of reporting. The codebook had 3 types of input formats, including categorical options with data validation; hence, coders could only type prespecified responses, dichotomous options with data validation that indicated yes or no, and free answers where coders could leave notes. Coders were the same as the reviewers in the screening phase, who had 4 weeks of training using this codebook. The coding for this review started after the interrater reliability met 80% agreement in the training. Disagreements were resolved by discussion. The specific items extracted and synthesized are summarized in the *Results* section.

Results

Overview

A total of 2916 articles were identified through a database search, of which 2797 (95.92%) records were screened for titles and abstracts after duplicates were removed. Title and abstract review of the 2797 articles yielded 324 (11.58%) articles for possible eligibility, and 253 (9.05%) articles were screened for the subject matter. Ultimately, of the 253 articles, 19 (7.5%) articles on digital smoking behavior change interventions were included for data extraction and synthesis. The flow of literature screening is presented in Figure 1.

Figure 1. Flowchart of literature screening.



Digital Interventions

A summary of digital smoking behavior change interventions is presented in Table 1. Of the 19 studies included, 13 (68%) focused on smoking cessation and some also included smoking reduction as a secondary outcome that led to smoking cessation. Several articles focused on promoting a social movement for rejecting tobacco and reducing the influence of peer smoking, and others had more general topics such as promoting healthy lifestyles, tobacco-free lifestyles, and antitobacco norms.

Of the 19 studies, 10 (53%) studies used multiple digital media platforms. Websites, apps, and social media were the most frequently used channels. Among the studies that used single digital media, Facebook (3/19, 16%), apps (3/19, 16%), websites (1/19, 5%), YouTube (1/19, 5%), and software (1/19, 5%) were used. Multiple devices, including laptops, tablets, and smartphones, were used in 53% (10/19) of the studies. Of the 19 studies, 2 (11%) used smartphones, and 1 (5%) study each used tablets, cell phones, and a special device developed for the intervention, respectively. Approximately 21% (4/19) of studies did not explicitly report the devices used.

Of the 19 studies, 9 (53%) combined multiple modes of communication for the intervention. SMS text messages, images, and videos were typically used. Studies focusing on a single mode used text messages (3/19, 16%), videos (1/19, 5%), or images (1/19, 5%). Approximately 16% (3/19) of the studies were unclear about the modes of communication.

Table 1. Format of digital smoking interventions.

Source	Topic	Digital media	Device	Mode	Theoretical model	Conceptual model
Baskerville et al [17], Canada	Smoking cessation	Website, app, Facebook, and YouTube	Laptop and smartphone	Video, image, and articles	No	No
Davis et al [18], United States	Smoking cessation	YouTube and television	Desktop or laptop, mobile device, and television	Digital video advertisements and television advertisements	No	No
Goldenhersch et al [19], Argentina	Smoking cessation	App and cardboard headset device	Smartphone	Video, audio, and group chat	No	No
Guillory et al [20], United States	Tobacco-free lifestyle	Digital and social media (not reported specifically), radio, print, out of home, and local events	Laptop and smartphone	Unclear	Yes	No
Hair et al [21], United States	Antitobacco social movement	Social media, website, YouTube, and television	Laptop, tablet, and smartphone	Unclear	No	No
Kenfield et al [22], United States	Smoking as part of multiple themes	Website, Fitbit One, and SMS text messaging	Laptop, smartphone, and FitBit	SMS text message	No	No
Kim et al [23], United States	Smoking cessation	Website (video, mobile, and search advertisements that direct the campaign website) and display	Laptop, tablet, and smartphone	Video and image	No	No
Kim et al [24], United States	Smoking cessation and reduction	Facebook	Laptop and smartphone	Video-, text- and, image-based materials	No	No
An et al [25], United States	Smoking as part of multiple themes	Websites	Laptop and phone	SMS text message, image, video, and phone call	Yes	No
Marler et al [26], United States	Smoking cessation and reduction	Apps	Other (The Pivot Breath Sensor, a mobile Pivot app)	SMS text message	Yes	No
Masaki et al [27], Japan	Smoking cessation	App, a connected cloud system, a paired mobile exhaled carbon monoxide checker device, and a web-based PC software	Smartphone and laptop	Digital diary, videos, chatbot, and biomedical recording	No	No
Namkoong et al [28], United States	Antitobacco social movement	Facebook	Not reported	Videos, text, and pictures	Yes	Yes
Onezi et al [29], Saudi Arabia	Smoking cessation (smoking relapse prevention)	Twitter and WhatsApp	Not reported	Other (social media–based support groups)	No	No
Romer et al [30], United States	Reduction of influence of peer smoking	YouTube	Not reported	Video	No	No
Thrul and Ramo [31], United States	Smoking cessation	Facebook	Laptop, tablet, and smartphone	Unclear	Yes	No
Tsoh et al [32], United States	Smoking cessation	Facebook	Tablet	Assessment, video messages, 1-page summary printout, or email	No	No
Webb et al [33], United Kingdom	Smoking cessation	App	Smartphone	Video, audio, quizzes, and quit coach through digital chat	No	No
Bary-Weisberg et al [34], Israel	Smoking cessation	Website and SMS text messaging	Cell phone (for SMS text messaging)	SMS text message	No	No

Source	Topic	Digital media	Device	Mode	Theoretical model	Conceptual model
Burford et al [35], Australia	Smoking cessation	Other (The APRIL Face Aging software)	Not reported	Image	Yes	No

Approximately 68% (13/19) of the articles stated that specific theoretical models applied to interventions, whereas 32% (6/19) did not report any models. These models were defined as previously published theoretical models related to social, cognitive, and behavioral factors. Specific models mentioned in the literature include the social cognitive theory [36], transtheoretical models [37], theory of reasoned action, d theory of planned behavior, self-determination theory [38,39], and social branding framework. This study also reviewed the application of conceptual models defined as frameworks designed for specific interventions in the literature. Approximately 95% (18/19) of the studies did not mention any conceptual model specific to the interventions.

Measures of Dose and Response

The measures used to assess the dose (exposures), response (outcomes), and form of measurement are summarized in Table 2. For response measures, 32% (6/19) of studies reported multiple behavioral outcomes. Measures included smoking status, smoking reduction, abstinence, quit attempts, successful quitting, information search related to smoking cessation, campaign-related topics, and the use of cessation aids. Of the 19 studies, 5 (23%) used multiple psychosocial measures, and 4 (21%) reported a single psychosocial measure. Approximately 37% (7/19) of studies did not report any psychosocial measures. Intention (4/19, 21%), self-efficacy (4/19, 21%), awareness (3/19, 16%), norms (2/19, 11%), and stage of change (2/19, 11%) were mainly used.

For dose measurements, the frequency of exposure, length of exposure, and engagement were assessed. This was designed under the assumption that the amount of dose can be expressed as the sum of the intended dose and the enacted dose; the intended dose equals the frequency of exposure multiplied by the length of exposure planned by the intervention side, and the enacted dose equals the amount of engagement that was actively received by the targeted. Approximately 37% (7/19) of studies explicitly reported the frequency of exposure, whereas 63% (12/19) of studies were unclear on that point. Of the 19 studies, 3 (16%) studies reported the frequency per week, 3 (16%) other studies reported the frequency per day, and 2 (11%) studies reported that the frequency varied for each week. The length of exposure was explicitly reported in 21% (4/19) of studies, whereas 79% (15/19) of studies were unclear about this. All 4 studies that mentioned the length of exposure used videos for the intervention. Engagement was measured in 58% (11/19) of studies. Of the 19 studies, 8 (42%) reported multiple measures, and 3 (16%) reported a single measure. Visits, clicks, log-ins, and views were measured most frequently across different modes of the intervention. Social media interventions reported subscriptions, likes, comments, and postings of content. Finally, 42% (8/19) of studies did not report on engagement.

The dose-response relationship was assessed in 16% (3/19) of studies using different levels of exposure. One of the studies compared the effects of intervention between a standard-dose group and a higher-dose group. Another study examined the interaction between the time of exposure and the treatment arm. One of the studies examined the relationship between participants' levels of active engagement and the targeted behavior. Outcome assessment was self-reported in 68% (13/19) of studies and a combination of self-report and objective measurements in 21% (4/19) of studies. One of the studies applied only objective measurements using tracking software, and another study used an aggregated self-report measure that assessed population-level awareness.

Table 2. Measures used to assess dose and response.

Source	Response (outcome)		Dose (exposure)			Dose-response	Outcome report
	Behavioral	Psychosocial	Frequency of exposure	Length of exposure	Engagement		
Baskerville et al [17], Canada	Smoking cessation and use of cessation aid	Intention to quit	No	No	Visit, installation of the app, and posting content	No	Self-reported
Davis et al [18], United States	Not measured	Advertisement awareness	No (gross rating points used); depends on the size of the market	Yes (30-second advertisements)	Not measured	Yes	Self-reported
Goldenhersch et al [19], Argentina	Abstinence	Readiness to quit	Yes (every day)	Yes (10-minute videos)	Not measured	No	Self-reported
Guillory et al [20], United States	Not measured	Awareness and receptivity	No	No	Not measured	Yes (time×treatment interaction)	Self-reported
Hair et al [21], United States	Current cigarette use	Advertisement awareness and intentions	No	No	View	No	Other; aggregated self-reported advertisement recall across people grouped by time (weeks) to form a measure of advertisement awareness
Kenfield et al [22], United States	Smoking	Not measured	Yes (4-5 SMS text messages each week)	No	Activity data from Fitbit, response to SMS text messages, website log-in, and page view	No	Self-reported
Kim et al [23], United States	Campaign-related topics search	Not measured	No	No	Visit, impressions, and clicks	No	Measured objectively
Kim et al [24], United States	Smoking reduction	Antismoking attitudes, readiness to quit, motivation to quit, self-efficacy beliefs, and perceived social support	Yes (different each week)	No	Likes and comments	Yes	Self-reported
An et al [25], United States	30-day abstinence from cigarette smoking	Not measured	Yes (weekly)	No	Not measured	No	Self-reported
Marler et al [26], United States [26]	Quit attempts, cigarettes per day reduction, and abstinence	Stage of change, desire to quit, readiness to quit, confidence to quit, difficulty to quit, and goals	Yes (>4 times use a day); up to twice weekly SMS text messages	No	Not measured	No	Both
Masaki et al [27], Japan	Smoking cessation	Not measured	No	No	Not measured	No	Both
Namkoong et al [28], United States	Smoking-related information seeking	Attitude, descriptive norms, subjective norms, behavioral control, and behavioral intention	Yes (every day)	No	Likes and comments	No	Self-reported

Source	Response (outcome)		Dose (exposure)			Dose-response	Outcome report
	Behavioral	Psychosocial	Frequency of exposure	Length of exposure	Engagement		
Onezi et al [29], Saudi Arabia	Smoking cessation and smoking frequency	Not measured	No	No	Subscription to a social media support group	No	Self-reported
Romer et al [30], United States	Not measured	Smoking norms, mortality beliefs, and smoking attitudes	No	Yes (approximately 4-5 seconds and display of messages)	Not measured	No	Self-reported
Thrul and Ramo [31], United States	Purposeful 24-hour smoking quit attempt	Not measured	No	No	Not measured	No	Self-reported
Tsoh et al [32], United States	Smoking abstinence, 24-hour quit attempts, and quit methods	Not measured	No	Yes (videos ranged from 8 to 65 seconds, averaging 29 seconds in length); patients watched 14 to 22 video segments depending on their responses	Session completion and patient-provider discussion	No	Both
Webb et al [33], United Kingdom	Smoking status, self-reported 7-day point prevalence abstinence at 4 weeks after the quit date, 14-day point prevalence abstinence, and any additional quit attempts after the quit date	Attitudes and perceptions of smoking; self-reported changes in confidence levels, knowledge, attitudes, and perceptions related to smoking cessation; and changes in Smoking Abstinence Self-efficacy questionnaire	No	No	App opens, stage progression, number of messages sent, check-ins, and diary entries	No	Self-reported
Bary-Weisberg et al [34], Israel	Smoking status	Self-efficacy	Yes (different each week)	No	Keywords sent on text	No	Self-reported
Burford et al [35], Australia	Successful quitting and quit attempts	Progression along the transtheoretical stages of change model and self-perceptions and attitudes toward smoking behavior	No	No	Not measured	No	Both

Quality of Reporting

The clarity of reporting was assessed for media role, dose and response measurement, and funding sources. The findings on clarity of reporting, study design, and bias are summarized in Table 3.

The role of media was clearly reported in all the studies (19/19, 100%). Behavior was clearly measured in all 15 studies that assessed behavioral outcomes. Among the 12 studies that assessed psychosocial outcomes, 11 (92%) reported them clearly, whereas 1 (8%) was unclear about its measurement.

Table 3. Clarity of measures and reporting.

Source	Clarity of media role	Response (outcomes)		Dose (exposure)			Study design	Bias	Funding source
		Behavior report	Psychosocial report	Intervention duration	Data collection duration	Engagement report			
Baskerville et al [17], Canada	Yes	Yes	Yes	Unclear	2 to <3 months	Yes	Quasi-experimental study	Generalizability and self-report	Yes
Davis et al [18], United States	Yes	N/A[a]	Yes	Unclear	Unclear	N/A	Nonexperimental study	Randomization and self-report	Yes
Goldenhersch et al [19], Argentina	Yes	Yes	Yes	2 weeks to <1 month	2 to <3 months	Yes	Experimental study	Short period of data collection	Yes
Guillory et al [20], United States	Yes	N/A	Yes	Unclear	>1 year	N/A	Quasi-experimental study	Aided awareness report (recall bias)	Yes
Hair et al [21], United States	Yes	Yes	Yes	Unclear	>1 year	No	Nonexperimental study	Representativeness	Yes
Kenfield et al [22], United States	Yes	Yes	N/A	2 to <3 months	2 to <3 months	Yes	Experimental study	Representativeness	Yes
Kim et al [23], United States	Yes	Yes	N/A	2 to <3 months	2 to <3 months	Yes	Nonexperimental study	Generalizability	Yes
Kim et al [24], United States	Yes	Yes	Yes	2 weeks to <1 month	1 to <2 months	Yes	Nonexperimental study	Self-report, small sample size, and representativeness	Yes
An et al [25], United States	Yes	Yes	N/A	1 to <2 months	2 to <3 months	N/A	Experimental study	Self-report (recall bias)	Yes
Marler et al [26], United States	Yes	Yes	Yes	2 to <3 months	2 to <3 months	N/A	Nonexperimental study	Generalizability	Yes
Masaki et al [27], Japan	Yes	Yes	N/A	4 to <6 months	6 months to <1 year	N/A	Experimental study	Representativeness	Yes
Namkoong et al [28], United States	Yes	N/A	Yes	2 weeks to <1 month	2 weeks to <1 month	No	Quasi-experimental study	Representativeness	No
Onezi et al [29], Saudi Arabia	Yes	Yes	N/A	Not reported	Not reported	Yes	Nonexperimental study	Generalizability and cross-sectional	Yes
Romer et al [30], United States	Yes	N/A	Yes	Not reported	Not reported	N/A	Quasi-experimental study	Generalizability and control setting	Yes
Thrul and Ramo [31], United States	Yes	Yes	N/A	1 to <2 months	6 months to <1 year	N/A	Nonexperimental study	Self-report (recall bias), representativeness, and low test power	Yes
Tsoh et al [32], United States	Yes	Yes	N/A	Other	2 to <3 months	Yes	Nonexperimental study	No control and self-report	Yes
Webb et al [33], United Kingdom	Yes	Yes	Yes	2 weeks to <1 month	6 months to <1 year	Yes	Experimental study	Generalizability and biased sample	Yes
Bary-Weisberg et al [34], Israel	Yes	Yes	Yes	4 to <6 months	Unclear	Yes	Nonexperimental study	Representativeness	Yes
Burford et al [35], Australia	Yes	Yes	No	Not reported	4 to <6 months	N/A	Experimental study	No blinding	No

[a] N/A: not applicable.

The duration of intervention was unclear in 21% (4/19) of studies and was not reported in 16% (3/19) of studies. Of the 19 studies, 12 (63%) reported this explicitly. The intervention duration ranged from 10 minutes to 6 months. One of the studies (1/12, 8%) reported it in the range as it varied among participants because of the customization function of the digital intervention. Of the 19 studies, the intervention duration was 2 weeks to a month in 4 (33%) studies, 1 to 2 months in 2 (17%) studies, 2 to 3 months in 3 (25%) studies, and 4 to 6 months in 2 (17%) studies. The duration of data collection was unclear in 11% (2/19) of studies and was not reported in 11% (2/19) of studies. It ranged from 2 to 3 months in 47% (7/15) of studies, 6 months to a year in 20% (3/15) of studies, and >1 year in 13% (2/15) of studies. Approximately 7% (1/15) of studies each collected data for 2 weeks to a month, 1 to 2 months, and 4 to 6 months, respectively. The reporting of engagement was clear in 82% (9/11) of the studies that assessed engagement.

Of the 19 studies, 6 (32%) applied experimental study designs, 4 (21%) used quasi-experimental designs, and 9 (47%) used nonexperimental study designs. Quasi-experimental designs were defined as studies with a control group that did not involve random assignment. Of the 6 experimental studies, 3 (50%) reported representativeness and a biased sample as a potential source of bias. Quasi-experimental studies reported that self-reporting and aided recall were threats to bias (2/4, 50%). Lack of randomization, self-reporting, small sample size, and lower statistical power were frequently reported, in addition to generalizability and representativeness among nonexperimental studies. Of the 19 studies, 17 (89%) reported a source of funding, and 2 (11%) did not report this information.

Discussion

Principal Findings

This study identified the literature on digital behavior change interventions related to smoking and addressed the current practice of measuring and reporting intervention contents, channels, and dose and response outcomes. Data synthesis showed that both measurement and reporting were inconsistent across studies, illustrating the current challenges in this field of research.

This review revealed a lack of preventive studies on tobacco and nicotine use. Among 324 digital behavior change intervention papers, only 19 (5.9%) papers were relevant to smoking, and none centered on preventing its initiation. Although a literature search detected numerous prevention research papers, this review only included papers that assessed behavioral and psychosocial outcomes and did not include studies that focused on perceptions and engagement. This lack of research may reflect the fact that prevention studies for digital interventions remain at an early stage of identifying the basic conditions to make interventions engaging and effective rather than measuring the effectiveness of such interventions in achieving behavior change. This may also be because of a lack of funding in prevention research, as well as challenges in long-term follow-up to detect differences in smoking initiation rates.

Another gap in the literature on digital health identified in this review was the dose-response relationship. Although the effectiveness of interventions depends on the amount of exposure, only 16% of the literature has assessed the outcomes across different levels of doses.

The biggest issue was the inconsistency and vagueness of the reporting. Most studies paid attention to reporting outcomes and were relatively clear on their measurements in the methods sections. However, reporting of the amount of intervention offered to participants and the actual engagement was often unclear and not explicitly mentioned in a large proportion of studies. Insufficient details on the amount of exposure make it more difficult to compare outcomes across studies and conduct meta-analyses, which hinders the provision of evidence on effective digital interventions.

These findings were consistent with previous research on digital health interventions. Hu et al [3] found that only 10.6% of papers on digital health focused on health promotion and interventions, and only 2.6% centered on substance use, including tobacco use. Abad et al [4] pointed out that only 6.8% of the literature on digital health focused on smoking. Another review by Seiler et al [1] demonstrated that there have been relatively few rigorous studies on the effectiveness of digital behavior change interventions, and the reporting of design, measures, data collection, and other methods need to be improved [5].

The amount of exposure to an intervention can be divided into the intended dose planned by the intervention side and the enacted dose that was completed by the participant side. Dose can be expressed as the frequency multiplied by the length of the intended intervention dosage, and the amount of engagement can offer information about the enacted dosage. Clarifying these elements in the literature will lead to the promotion of comparable reporting and advancement of the evidence development of digital health interventions.

The types of behaviors used to report the outcomes were mixed. It can be divided into (1) targeted behavior (ie, smoking cessation), (2) surrogate behavior (ie, smoking reduction), and (3) behavior related to the improvement of the likelihood of conducting targeted behaviors (ie, use of cessation aid and smoking cessation information search). Although the measurement of these outcomes was predominantly self-reported, some studies combined objective measurements. The methods of objective measurement included (1) biochemical devices (eg, exhaled carbon monoxide tracker), (2) automatic digital tracking (eg, for information searches), (3) electronic health records, and (4) group-level psychological measures (eg, group-level awareness of campaigns). Leveraging these emerging methods will enhance the validity and reliability of measurements and advance evidence-based digital health interventions.

Strengths

To the best of our knowledge, this is the first study to review the dose and response measurements of digital smoking interventions. Dose measures were divided into the intended dose from the intervention side and the enacted dose that was actually completed by the participant side. The intended dose was separately assessed for frequency and length. Response measures were organized into behavioral and psychosocial outcomes, and measurement methods were assessed to determine their validity. The proportion of dose-response relationship studies was identified to determine the stage of current research in the field. The literature search combined general and journal-specific searches and yielded 2916 studies that showed high coverage.

Research gaps exist in assessing digital behavior change interventions with clear dose and response measurements and the dose-response relationship between the levels of intervention exposure and outcomes. More prospective studies are needed to examine the relationship between higher and lower dosages of interventions on smoking outcomes. For example, a well-designed dose-response relationship study on vaping outcomes among a specific population will identify the effective amount of digital intervention to prevent the initiation and decrease the amount of e-cigarette use. This will provide evidence for identifying the effective amount of intervention and offer grounds for conducting well-designed meta-analyses that synthesize evidence for digital behavior change interventions. This paper contributes to building a base for these studies.

Limitations

This review had several limitations. First, the number of studies included in this analysis was small. This was mainly because of the lack of digital behavior change intervention studies related to smoking and the measurement of behavioral and psychosocial outcomes in the current phase in the area. Second, the risk of bias assessment was omitted from the review. This was because of the nature of the mixed study designs in the included studies as there is little risk of bias assessment tools intended to apply to a variety of study designs. Instead, the quality of the papers was assessed by reviewing and evaluating the study design, theoretical model application, conceptual model application, cost, funding source, and bias source. In addition, the clarity of the media role and every measure of dose and response were assessed as part of the review for each study. When digital smoking behavior change interventions become more common and more literature can be included in the future, these limitations can be minimized.

Conclusions

There are challenges in every emerging area of scientific research. Digital smoking interventions are a new and growing area of research involving continuous and rapid advancements in technologies for health intervention delivery, implementation, and dissemination. This characteristic implies that it is essential to use consistent, standardized methods to evaluate the outcomes to accurately understand the efficacy and effectiveness of the interventions in rapid change, particularly in this area. For future improvement in this realm, this study assessed the consistency of measures used to quantify the dose and response in digital smoking interventions. These results suggest that both dose and response measures are often not clearly defined and are inconsistent across studies. This can imply that synthesizing the evidence on digital smoking interventions has been a challenge, and past studies could have been based on vague reporting and outcome evaluations. This is particularly an issue as interventions to date could be based on biased findings.

This study details and structures the composition of evaluation measures and assesses their use in digital smoking intervention studies. To the best of our knowledge, this is the first study to review dose and response measurements of digital smoking interventions. Although it has limitations such as the small number of papers included and the limited risk of bias assessment, the study successfully (1) reviewed and codified measures used for digital health interventions in tobacco and nicotine use research, (2) evaluated the quality of reporting of dose and response, and (3) identified areas for improvement in the field. More prospective studies that examine the clarity and consistency of measures among a larger number of studies are needed to develop further grounds for evidence-based digital smoking interventions. There is also a need for a clear and consistent reporting scheme for digital health interventions to accurately evaluate outcomes and conduct well-designed meta-analyses. Provision of clear and consistent reporting of both outcomes and exposures is needed to develop further evidence in this field that leads to protecting the lives and health of the public.

Acknowledgments

This study was supported by the National Cancer Institute (R01CA253013; principal investigator: WDE).

Authors' Contributions

MI and WDE conceived the study. MI designed the study protocol. WDE and JC contributed to the study design and provided expertise in digital media studies on tobacco control. MI, RG, SM, and AB performed the analyses. MI prepared the manuscript. All authors contributed to the final approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Seiler J, Libby TE, Jackson E, Lingappa J, Evans W. Social media-based interventions for health behavior change in low- and middle-income countries: systematic review. J Med Internet Res 2022 Apr 14;24(4):e31889 [FREE Full text] [doi: 10.2196/31889] [Medline: 35436220]
2. Sanchez-Paramow C, Legovini A. Using social media to change norms and behaviors at scale. World Bank Blogs. 2021 Jan 12. URL: https://blogs.worldbank.org/voices/using-social-media-change-norms-and-behaviors-scale [accessed 2022-03-28]
3. Hu Y. Health communication research in the digital age: a systematic review. J Commun Healthc 2015 Dec 18;8(4):260-288. [doi: 10.1080/17538068.2015.1107308]
4. Shakeri Hossein Abad Z, Kline A, Sultana M, Noaeen M, Nurmambetova E, Lucini F, et al. Digital public health surveillance: a systematic scoping review. NPJ Digit Med 2021 Mar 03;4(1):41 [FREE Full text] [doi: 10.1038/s41746-021-00407-6] [Medline: 33658681]
5. Elaheebocus SM, Weal M, Morrison L, Yardley L. Peer-based social media features in behavior change interventions: systematic review. J Med Internet Res 2018 Feb 22;20(2):e20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]
6. Hsu MS, Rouf A, Allman-Farinelli M. Effectiveness and behavioral mechanisms of social media interventions for positive nutrition behaviors in adolescents: a systematic review. J Adolesc Health 2018 Nov;63(5):531-545. [doi: 10.1016/j.jadohealth.2018.06.009] [Medline: 30197198]
7. Chau M, Burgermaster M, Mamykina L. The use of social media in nutrition interventions for adolescents and young adults-a systematic review. Int J Med Inform 2018 Dec;120:77-91 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.10.001] [Medline: 30409348]
8. Brigden A, Anderson E, Linney C, Morris R, Parslow R, Serafimova T, et al. Digital behavior change interventions for younger children with chronic health conditions: systematic review. J Med Internet Res 2020 Jul 31;22(7):e16924 [FREE Full text] [doi: 10.2196/16924] [Medline: 32735227]
9. Chan L, O'Hara B, Phongsavan P, Bauman A, Freeman B. Review of evaluation metrics used in digital and traditional tobacco control campaigns. J Med Internet Res 2020 Aug 11;22(8):e17432 [FREE Full text] [doi: 10.2196/17432] [Medline: 32348272]
10. Miech R, Johnston L, Bachman J, O?Malley P, Schulenberg J, Patrick M. Monitoring the future: a continuing study of American youth (12th-Grade Survey), 2020. Inter-university Consortium for Political and Social Research. 2021 Oct 26. URL: https://doi.org/10.3886/ICPSR38156.v1 [accessed 2022-05-15]
11. Truth initiative homepage. Truth Initiative. URL: http://truthinitiative.org/ [accessed 2022-05-15]
12. Preventing Tobacco Use Among Youth and Young Adults A Report of the Surgeon General. Pittsburgh, PA: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General; 2012.
13. Health effects of cigarette smoking. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/index.htm#cancer [accessed 2022-05-15]
14. Cornelius ME, Wang TW, Jamal A, Loretan CG, Neff LJ. Tobacco product use among adults - United States, 2019. MMWR Morb Mortal Wkly Rep 2020 Nov 20;69(46):1736-1742 [FREE Full text] [doi: 10.15585/mmwr.mm6946a4] [Medline: 33211681]
15. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2022-05-15]
16. McVay MA, Bennett GG, Steinberg D, Voils CI. Dose-response research in digital health interventions: concepts, considerations, and challenges. Health Psychol 2019 Dec;38(12):1168-1174 [FREE Full text] [doi: 10.1037/hea0000805] [Medline: 31580127]
17. Baskerville NB, Azagba S, Norman C, McKeown K, Brown KS. Effect of a digital social media campaign on young adult smoking cessation. Nicotine Tob Res 2016 Mar;18(3):351-360. [doi: 10.1093/ntr/ntv119] [Medline: 26045252]
18. Davis KC, Shafer PR, Rodes R, Kim A, Hansen H, Patel D, et al. Does digital video advertising increase population-level reach of multimedia campaigns? Evidence from the 2013 tips from former smokers campaign. J Med Internet Res 2016 Sep 14;18(9):e235 [FREE Full text] [doi: 10.2196/jmir.5683] [Medline: 27627853]
19. Goldenhersch E, Thrul J, Ungaretti J, Rosencovich N, Waitman C, Ceberio MR. Virtual reality smartphone-based intervention for smoking cessation: pilot randomized controlled trial on initial clinical efficacy and adherence. J Med Internet Res 2020 Jul 29;22(7):e17571 [FREE Full text] [doi: 10.2196/17571] [Medline: 32723722]
20. Guillory J, Henes A, Farrelly MC, Fiacco L, Alam I, Curry L, et al. Awareness of and receptivity to the fresh empire tobacco public education campaign among hip hop youth. J Adolesc Health 2020 Mar;66(3):301-307. [doi: 10.1016/j.jadohealth.2019.09.005] [Medline: 31704108]
21. Hair EC, Niederdeppe J, Rath JM, Bennett M, Romberg A, Pitzer L, et al. Using aggregate temporal variation in ad awareness to assess the effects of the truth® campaign on youth and young adult smoking behavior. J Health Commun 2020 Mar 03;25(3):223-231. [doi: 10.1080/10810730.2020.1733144] [Medline: 32129727]
22. Kenfield SA, Van Blarigan EL, Ameli N, Lavaki E, Cedars B, Paciorek AT, et al. Feasibility, acceptability, and behavioral outcomes from a technology-enhanced behavioral change intervention (prostate 8): a pilot randomized controlled trial in men with prostate cancer. Eur Urol 2019 Jun;75(6):950-958. [doi: 10.1016/j.eururo.2018.12.040] [Medline: 30638635]

23. Kim A, Hansen H, Duke J, Davis K, Alexander R, Rowland A, et al. Does digital ad exposure influence information-seeking behavior online? Evidence from the 2012 tips from former smokers national tobacco prevention campaign. J Med Internet Res 2016 Mar 16;18(3):e64 [FREE Full text] [doi: 10.2196/jmir.4299] [Medline: 26983849]
24. Kim SJ, Marsch LA, Brunette MF, Dallery J. Harnessing Facebook for smoking reduction and cessation interventions: Facebook user engagement and social support predict smoking reduction. J Med Internet Res 2017 May 23;19(5):e168 [FREE Full text] [doi: 10.2196/jmir.6681] [Medline: 28536096]
25. An LC, Demers MR, Kirch MA, Considine-Dunn S, Nair V, Dasgupta K, et al. A randomized trial of an avatar-hosted multiple behavior change intervention for young adult smokers. J Natl Cancer Inst Monogr 2013 Dec;2013(47):209-215 [FREE Full text] [doi: 10.1093/jncimonographs/lgt021] [Medline: 24395994]
26. Marler JD, Fujii CA, Wong KS, Galanko JA, Balbierz DJ, Utley DS. Assessment of a personal interactive carbon monoxide breath sensor in people who smoke cigarettes: single-arm cohort study. J Med Internet Res 2020 Oct 02;22(10):e22811 [FREE Full text] [doi: 10.2196/22811] [Medline: 32894829]
27. Masaki K, Tateno H, Nomura A, Muto T, Suzuki S, Satake K, et al. A randomized controlled trial of a smoking cessation smartphone application with a carbon monoxide checker. NPJ Digit Med 2020;3:35 [FREE Full text] [doi: 10.1038/s41746-020-0243-5] [Medline: 32195370]
28. Namkoong K, Nah S, Record RA, Van Stee SK. Communication, reasoning, and planned behaviors: unveiling the effect of interactive communication in an anti-smoking social media campaign. Health Commun 2017 Jan;32(1):41-50. [doi: 10.1080/10410236.2015.1099501] [Medline: 27119592]
29. Onezi HA, Khalifa M, El-Metwally A, Househ M. The impact of social media-based support groups on smoking relapse prevention in Saudi Arabia. Comput Methods Programs Biomed 2018 Jun;159:135-143. [doi: 10.1016/j.cmpb.2018.03.005] [Medline: 29650308]
30. Romer D, Jamieson PE, Jamieson KH, Jones C, Sherr S. Counteracting the influence of peer smoking on YouTube. J Health Commun 2017 Apr;22(4):337-345 [FREE Full text] [doi: 10.1080/10810730.2017.1290164] [Medline: 28306481]
31. Thrul J, Ramo DE. Cessation strategies young adult smokers use after participating in a Facebook intervention. Subst Use Misuse 2017 Jan 28;52(2):259-264 [FREE Full text] [doi: 10.1080/10826084.2016.1223690] [Medline: 27759475]
32. Tsoh JY, Quach T, Duong TB, Sa Nan Park E, Wong C, Huang SM, et al. Interactive mobile doctor (iMD) to promote patient-provider discussion on tobacco use among Asian American patients in primary care: a pilot study. J Am Board Fam Med 2018 Nov 09;31(6):869-880 [FREE Full text] [doi: 10.3122/jabfm.2018.06.180018] [Medline: 30413543]
33. Webb J, Peerbux S, Smittenaar P, Siddiqui S, Sherwani Y, Ahmed M, et al. Preliminary outcomes of a digital therapeutic intervention for smoking cessation in adult smokers: randomized controlled trial. JMIR Ment Health 2020 Oct 06;7(10):e22833 [FREE Full text] [doi: 10.2196/22833] [Medline: 33021488]
34. Bary-Weisberg D, Meltser M, Oberman M, Pato Benari A, Bar-Zeev Y, Shalev S, et al. Feasibility of a text-messaging smoking cessation program for soldiers in Israel. BMC Public Health 2019 Jun 26;19(1):715 [FREE Full text] [doi: 10.1186/s12889-019-6958-z] [Medline: 31238914]
35. Burford O, Jiwa M, Carter O, Parsons R, Hendrie D. Internet-based photoaging within Australian pharmacies to promote smoking cessation: randomized controlled trial. J Med Internet Res 2013 Mar 26;15(3):e64 [FREE Full text] [doi: 10.2196/jmir.2337] [Medline: 23531984]
36. 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]
37. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change: applications to addictive behaviors. Am Psychol 1992;47(9):1102-1114. [doi: 10.1037/0003-066x.47.9.1102]
38. 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]
39. Fishbein M. A theory of reasoned action: some applications and implications. Nebr Symp Motiv 1980;27:65-116. [Medline: 7242751]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews

Edited by R Kukafka; submitted 04.04.22; peer-reviewed by H Mehdizadeh, N Chekin; comments to author 03.05.22; revised version received 27.06.22; accepted 25.07.22; published 25.08.22.

Please cite as:
Ichimiya M, Gerard R, Mills S, Brodsky A, Cantrell J, Evans WD
The Measurement of Dose and Response for Smoking Behavior Change Interventions in the Digital Age: Systematic Review
J Med Internet Res 2022;24(8):e38470
URL: https://www.jmir.org/2022/8/e38470
doi:10.2196/38470
PMID:36006682

Review

Planning and Reporting Effective Web-Based RAND/UCLA Appropriateness Method Panels: Literature Review and Preliminary Recommendations

Jordan B Sparks[1], MPH; Mandi L Klamerus[1], MPH; Tanner J Caverly[1,2,3,4], MD, MPH; Sarah E Skurla[1], MPH; Timothy P Hofer[1,2,3], MD, MSc; Eve A Kerr[1,2,3], MD, MPH; Steven J Bernstein[1,2,3], MD, MPH; Laura J Damschroder[1,5], MS, MPH

[1]VA Center for Clinical Management Research, Ann Arbor, MI, United States
[2]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States
[3]Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, United States
[4]Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI, United States
[5]VA Quality Enhancement Research Initiative (QUERI) Personalizing Options through Veteran Engagement (PROVE) Program, Ann Arbor, MI, United States

Corresponding Author:
Laura J Damschroder, MS, MPH
VA Center for Clinical Management Research
VA Ann Arbor Healthcare System
2215 Fuller Rd (152)
Ann Arbor, MI, 48105
United States
Phone: 1 734 845 3603
Email: laura.damschroder@va.gov

Abstract

Background: The RAND/UCLA Appropriateness Method (RAM), a variant of the Delphi Method, was developed to synthesize existing evidence and elicit the clinical judgement of medical experts on the appropriate treatment of specific clinical presentations. Technological advances now allow researchers to conduct expert panels on the internet, offering a cost-effective and convenient alternative to the traditional RAM. For example, the Department of Veterans Affairs recently used a web-based RAM to validate clinical recommendations for de-intensifying routine primary care services. A substantial literature describes and tests various aspects of the traditional RAM in health research; yet we know comparatively less about how researchers implement web-based expert panels.

Objective: The objectives of this study are twofold: (1) to understand how the web-based RAM process is currently used and reported in health research and (2) to provide preliminary reporting guidance for researchers to improve the transparency and reproducibility of reporting practices.

Methods: The PubMed database was searched to identify studies published between 2009 and 2019 that used a web-based RAM to measure the appropriateness of medical care. Methodological data from each article were abstracted. The following categories were assessed: composition and characteristics of the web-based expert panels, characteristics of panel procedures, results, and panel satisfaction and engagement.

Results: Of the 12 studies meeting the eligibility criteria and reviewed, only 42% (5/12) implemented the full RAM process with the remaining studies opting for a partial approach. Among those studies reporting, the median number of participants at first rating was 42. While 92% (11/12) of studies involved clinicians, 50% (6/12) involved multiple stakeholder types. Our review revealed that the studies failed to report on critical aspects of the RAM process. For example, no studies reported response rates with the denominator of previous rounds, 42% (5/12) did not provide panelists with feedback between rating periods, 50% (6/12) either did not have or did not report on the panel discussion period, and 25% (3/12) did not report on quality measures to assess aspects of the panel process (eg, satisfaction with the process).

Conclusions: Conducting web-based RAM panels will continue to be an appealing option for researchers seeking a safe, efficient, and democratic process of expert agreement. Our literature review uncovered inconsistent reporting frameworks and insufficient detail to evaluate study outcomes. We provide preliminary recommendations for reporting that are both timely and important for producing replicable, high-quality findings. The need for reporting standards is especially critical given that more people may prefer to participate in web-based rather than in-person panels due to the ongoing COVID-19 pandemic.

(J Med Internet Res 2022;24(8):e33898) doi:10.2196/33898

KEYWORDS

quality indicators; health care; web-based; virtual; RAND/UCLA appropriateness method; research design; de-implementation; digital health; health research; virtual health research; health technology; researchers; medical professionals

Introduction

The RAND/UCLA Appropriateness Method (RAM), a variant of the Delphi Method, was developed to synthesize existing evidence and the clinical judgement of a panel of medical experts. The goal of this method is to produce recommendations for appropriate treatment of specific clinical presentations, given current best evidence [1]. This method has been widely used to develop care recommendations and performance measures that define quality of care [2-6]; it provides a transparent and systematic approach that can garner trust and acceptance among physicians, other clinicians, patients, payers, and health systems [7].

The RAM classically involves engaging credible experts to evaluate specific clinical presentations in a 2-round rating process. In the initial round, experts independently rate each clinical scenario. During the second round, panelists participate in a 1 to 2-day in-person session where they have an opportunity to review and discuss each other’s first round ratings, revise the initial list of scenarios, and individually rerate each clinical indication. Indications are categorized as “appropriate,” “uncertain,” or “inappropriate” based on panelists’ median score and level of disagreement [1]. Compared to the standard Delphi Method, the RAM does not require panelists to reach group consensus after multiple rating rounds [1,8].

A difficulty in convening appropriate experts in person is their often-limited time and capacity to participate. Thus, there is a need to identify best practices for conducting expert panels via the internet not only to lower barriers to experts’ participation but also to reduce the costs involved with implementing traditional in-person RAMs. While the use of RAMs with a web-based component in health research was increasing prior to COVID-19, the pandemic has greatly accelerated the need for web-based alternatives with improved technology and end-user familiarity with these tools.

While there is a substantial body of literature describing and testing various aspects of the traditional in-person or hybrid RAMs, few studies report using a completely web-based RAM, and even fewer provide detailed descriptions on how the expert panels were conducted. There is often little information or guidance for designing approaches to meet the goals of specific studies. Boulkedid et al [9] published a systematic review of 80 articles published through 2009, finding that 63% used a “modified” Delphi Method but lacked enough detail to replicate or judge the quality of modified approaches to developing recommendations for quality health care indicators. Moreover, measures of process quality, such as consistent panelist engagement, are rarely reported. Because best practices for conducting virtual RAMs are unclear and reporting is inconsistent, we conducted a literature review to develop preliminary recommendations for implementing and reporting virtual RAMs.

Methods

Literature Search and Data Abstraction

In March 2019, we searched PubMed to identify studies published from 2009 to 2019 that reported using a virtual RAM to measure the appropriateness of medical care. The following search terms were used to identify relevant articles: “RAND/UCLA Appropriateness Method” OR “RAND Appropriateness Method” OR “Modified RAND” OR “RAND AND panel” AND “online OR e-delphi OR web OR virtual.” The full search strategy can be found in Multimedia Appendix 1. Two reviewers (JS and LD) screened each article and developed a list of inclusion and exclusion criteria, which are described in Multimedia Appendix 2. To be included, articles must have used the RAM to measure the appropriateness of medical care or focused on the development of clinical practice guidelines or performance measures. Moreover, the expert panel ratings must have been completed on the internet. Web-based ratings could include a teleconference component. Articles were included even if they did not report a rerate session or discussion period among panelists. Non-English articles and articles published prior to 2009 were excluded. Studies with goals not aimed toward providers (eg, improving support for patient caregivers) were also excluded. Additionally, articles were excluded if they were reviews or summaries of the literature. Relevant panel process data (ie, first author, year published, title, mode of administration, topic, and objectives) from each article included in our review were abstracted in a predefined matrix (Multimedia Appendix 3 [2,10-21]). Team members (JS, MK, and SS) independently abstracted the same sample of articles twice to (1) ensure that the basic data collected have been correctly entered in the spreadsheet and (2) verify that the selection criteria have been appropriately applied.

Subsequently, we expanded the Delphi reporting categories recommended by Boulkedid et al [9], which formed the basis of our article abstraction template. Specifically, we organized descriptions with respect to the following categories adapted from Boulkedid et al [9]: (1) composition and characteristics of the web-based RAM expert panels; (2) characteristics of the web-based RAM panel procedures; (3) results; and (4) panel satisfaction and engagement. Additional information abstracted into the matrix included the following: (1) descriptions of how quality indicators were selected; (2) the method used for participant recruitment; (3) whether materials were sent to participants prior to the first expert panel rating; (4) Delphi panel size and composition, as well as the duration of time the panel was convened; (5) purpose of convening the RAM panels; (6) criteria used to rate indications; (7) the web-based system used to host the panels; (8) the number of reported Delphi rounds; (9) description of feedback provided to panelists; and (10) descriptions of participation levels in the discussion rounds. Additionally, we captured information about second or third rounds of ratings and how the final list of indications was selected. Lastly, we captured information about panelist satisfaction and engagement. Four authors (JS, LD, MK, and SS) completed a second, more detailed, data abstraction. The authors independently reviewed each article and completed several rounds of data verification.

Ethics Approval

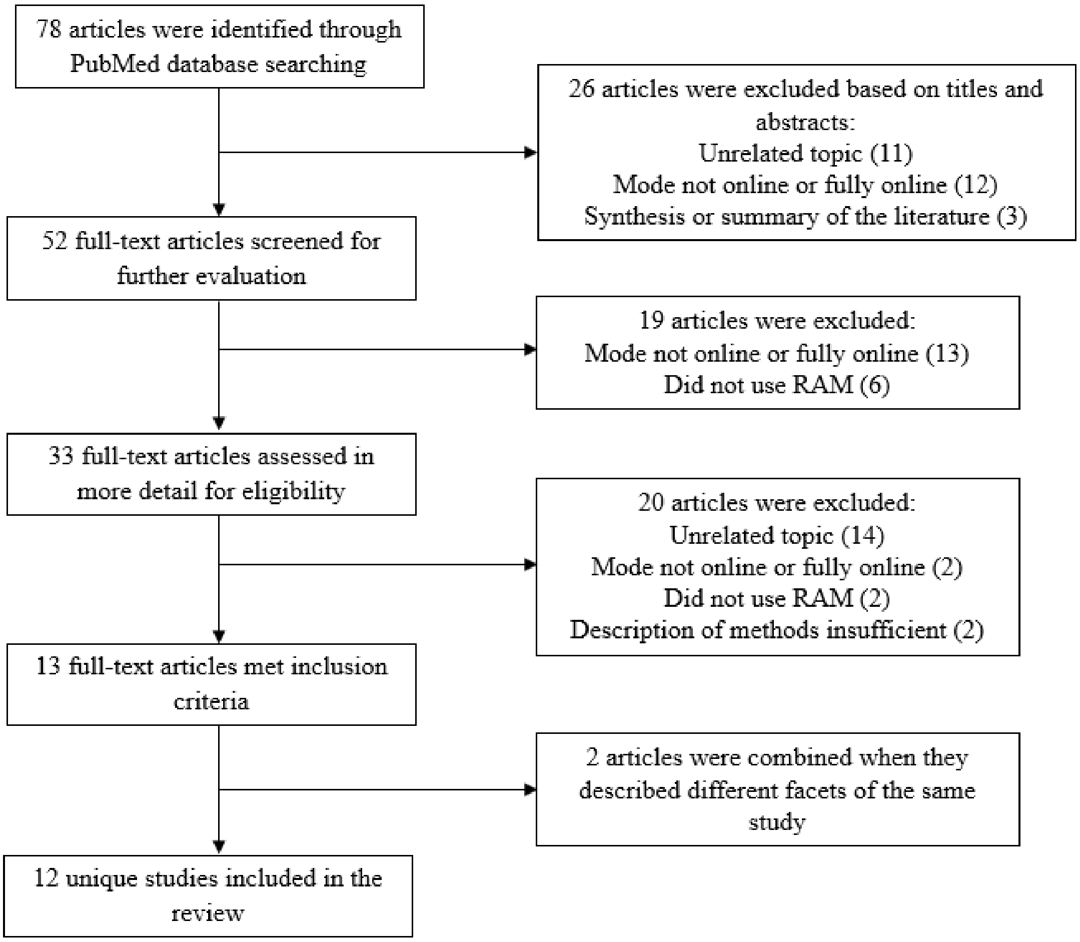
The ASSURES study was approved by the Ann Arbor VA Healthcare System IRB (project ID: 1597260).

Results

Article Selection

We identified 78 articles that reported using an “online” or “virtual” RAM from our narrative review; 26/78 (33%) articles were excluded based on the title or abstract (Figure 1). A full-text review of 52/78 (67%) articles was completed, resulting in the exclusion of an additional 39/52 (75%) articles; 13/52 (25%) articles were included in the review. We combined 2 published articles that met the inclusion criteria but described different facets of the same study, so the final review included 12/52 (23%) unique studies. Subsequent calculations are based on these 12 studies. The studies included in this literature review used completely web-based RAM approaches to accomplish their goals that ranged from developing quality performance measures or indicators to setting clinical practice standards. Throughout the manuscript, we use the term “indication” or “indicator” to standardize the description of statements panelists were asked to approve during the RAM process. Based on our narrative review, we developed foundational reporting recommendations from Boulkedid et al [9].

Figure 1. Article search and selection flowchart. RAM: RAND/UCLA Appropriateness Method.



Recruitment and Composition of Web-Based Expert Panels

Table 1 summarizes the characteristics of the RAM expert panels included in our review [2,10-21]. The methods for identifying and recruiting participants were not reported in 5/12 (42%) studies [13-15,18-20]. Of those who relied on professional networks or a snowball sampling approach, 3 (50%) reported inviting prospective participants via email [10,11,17].

The number of people invited to participate in the panels was reported in 9/12 (75%) of studies, ranging from 20 to 352 people with a median of 50 [2,10-13,16,17,20,21]. Of the 12 studies, 9 (75%) reported the number of panelists who participated in at least the first rating, ranging from 10 to 102 individuals, with a median of 42 [2,10-13,16-21]. In addition, all studies described included the types of stakeholders or experts who participated in the panels. In 11/12 (92%) studies, panelists were clinicians related to the topic studied [2,10-13,16-21]; 5 (42%) also included patients or people living with the condition studied [2,11,12,17,21]. Three-quarters of studies (9/12) had between one and two types of stakeholders. An additional 2 (17%) studies included three types of stakeholders; 1 (8%) study reported four or more types.

Table 1. Characteristics of the RAND/UCLA Appropriateness Method panels and procedure included in literature review (N=12).

Characteristics	Values
Methods for recruiting individuals (7/12, 58%), n (%)	
Professional networks or stakeholders	6 (86)
People invited to participate (9/12, 75%), median (min-max[a])	50 (20-352)
Participated in first rating (9/12, 75%), median (min-max)	42 (10-102)
Type of stakeholders[b], n (%)	
Clinicians[c]	11 (92)
Patients	5 (42)
Other	5 (42)
Stakeholder types per study, n (%)	
1	6 (50)
2	3 (25)
3	2 (17)
≥4	1 (8)
Type of RAND[d] procedure, n (%)	
Partial RAM[e]	7 (58)
Full RAM	5 (42)
Number of rounds (11/12, 92%), median (min-max)	3 (1-3)
Topic, n (%)	
Performance or outcome assessment measures	6 (50)
Assessment criteria	2 (17)
Prescribing indicators	1 (8)
Documentation standards	1 (8)
Antibiotic stewardship	1 (8)
Clinical practice standards	1 (8)
Web-based platform or system used (9/12, 75%), n (%)	
ExpertLens	4 (44)
SurveyMonkey	3 (33)
Canadian Fluid Survey System	1 (11)
REDCap[f]	1 (11)
Methods used to select indicators for the survey[b], n (%)	
Literature review	8 (67)
Stakeholder feedback	6 (50)
Prior surveys	2 (17)
Focus group	1 (8)
Other	1 (8)
Indicators in the first rating, median (min-max)	48 (6-524)
Prepanel materials, n (%)	1 (8)
Duration of consensus process (weeks), median	12
Geographical scope, n (%)	
National	9 (75)
International	3 (25)

Characteristics	Values
Item selection or rating criteria[b], n (%)	
Importance	4 (33)
Validity	4 (33)
Relevance	4 (33)
Feasibility	3 (25)
Other	3 (25)
Likelihood of use	2 (17)
Appropriateness	2 (17)
Number of selection criteria used, n (%)	
1	4 (33)
2	2 (17)
3	1 (8)
4	5 (42)
Feedback provided after first rating (7/12, 58%), n (%)	
Quantitative[g]	4 (57)
Quantitative and qualitative[h]	2 (29)
Other	1 (14)
Discussion process[b] (6/12, 50%), n (%)	
Asynchronous	5 (83)
Anonymous	5 (83)
Moderated	5 (83)
Rating process reported (8/12, 67%), n (%)	
Rating 2	7 (88)
Rating 3	1 (13)
Item selection process, n (%)	
Median score + IPR[i]/IPRAS[j] consensus	8 (67)
Median score + percentage of agreement	1 (8)
Percentage of agreement	1 (8)
Average score	1 (8)
Other	1 (8)
Process assessment; satisfaction, n (%)	3 (25)
Limitations noted, n (%)	11 (92)

[a] Min-max: minimum-maximum.

[b] The total percentages may exceed 100% because some studies used more than one criterion.

[c] Clinicians include people who actively work in health care settings such as hospitals and clinics to deliver care to patients (ie, doctors, nurses, pharmacists, etc).

[d] RAND corporation.

[e] RAM: RAND/UCLA Appropriateness Method.

[f] REDCap: Research Electronic Data Capture.

[g] Quantitative: group median, minimum, and maximum ratings. Feedback may include panelists' own ratings to illustrate position versus group ratings.

[h] Qualitative: abstract of panelists' comments.

[i] IPR: interpercentile range.

[j] IRPAS: interpercentile range adjusted for symmetry.

Characteristics of Web-Based Expert Panel Procedures

Table 1 summarizes the RAM procedural characteristics across the 12 studies; 5 (42%) studies described the use of all the steps specified by the RAM method (ie, full RAM) [2,11,12,16,21]; the remaining 7 (58%) studies reported the use of some but not all the steps (ie, partial RAM) [10,13-15,17-20]. All studies reported between one and three rating rounds, with a median of 3 rounds. In 6/12 (50%) studies, the RAM was used to develop a set of performance indicators (ie, indicators of clinical care quality) [2,10,11,13,16,21], and the remaining studies focused on developing indications for documenting or describing specific conditions (eg, rheumatoid arthritis and dental caries) [12,14,15,17-20]. The type of web-based system used to conduct the expert panel was reported in 9 (75%) studies [2,11-13,16,17,19-21]. Of these, 4 (44%) studies used the ExpertLens platform [2,11,12,16], and 5 (56%) listed other survey software (eg, SurveyMonkey, Research Electronic Data Capture, and the Canadian Fluid Survey System) [13,17,19-21]; the type of web-based system was not reported in 3/12 (25%) studies [10,14,15,18].

The methods used to select indicators for the survey were reported in all 12 studies. The most common method was a literature review (alone or in addition to stakeholder feedback (8/12, 67%) [2,11-13,16,17,20,21]. The number of indicators in the first rating were reported in all 12 studies, and ranged from 6 to 524 items, with a median of 48 items. Only 1/12 (8%) studies reported sending materials to participants prior to the panel sessions and included a document with rationale, methods, and indicator specifications [2]. Duration of the consensus process was reported in 7/12 (58%) studies [2,10,11,13,16,17,21]; median duration was 12 weeks. The geographical scope of expert panel members was reported in all studies; 9 (75%) panels were classified as national [2,10,12,13,16,18-21], and 3 (25%) were classified as international [11,14,15,17].

All studies specified the criteria used to rate each indicator (ie, relevance, importance, feasibility, etc) [2,10-21]. Half of the studies (6/12, 50%) used between one and two types of selection criteria [12,16,17,19-21]; the remaining studies reported using three or more selection criteria [2,10,11,13-15,18]. After the first rating, 5/12 (42%) studies did not report providing feedback of results to panelists [13-15,18-20]. The remaining studies reported providing panelists with frequency distributions, medians, and interquartile ranges for the group, as well as panelists' own responses compared to the group [2,10-12,16,17,21]. Researchers in 1 (8%) study revised a list of clinical indications based on input from panelists in Rating 1 and distributed this information to participants after the first rating [17]. During this time, panelists review and discuss the ratings, focusing on indications with significant disagreement [1]. However, only 6/12 (50%) of the studies reported included a panelist discussion period after the first round of ratings [2,11,12,16,17,21]; the discussions in 5/6 (83%) studies were either asynchronous, anonymous, or moderated web-based discussions [2,11,12,16,21]; 1 (8%) study used a nonanonymous synchronous webinar format [17].

Web-Based Expert Panel Results

Most studies (8/12; 67%) reported a second round of ratings [2,10-12,16,18,20,21], with 1/8 (13%) study indicating a third rating was conducted (Table 1) [18]. All studies provided information about how indications were selected for the final list. In 8/12 (67%) studies, items were selected following the RAM criteria for disagreement (where the calculated interpercentile range is greater than the interpercentile range adjusted for symmetry, with a panel median score between 6 and 9) [1,2,10-12,16,18,20,21]. The methods used to select indicators differed in the 4 (33%) remaining studies. In the first study, indicators were included based on the following two conditions: (1) the median score for each item was between 8 and 9 and (2) at least 70% of the panelists rated an item in the top third of the scale [17]. In the second study, at least three-quarters of panelists had to agree on an item for it to be selected [14,15]. In the third study, if the average agreement was at least 70% or higher across all 4 criteria of preventability for each item, then the indication was selected [13]. In the last study, more than 50% of panelists had to rate an indicator "extremely important" (ie, 9) for it to be selected [19].

Web-Based Expert Panel Process Assessments and Satisfaction

Approaches for reporting process assessments and satisfaction were not included in recommendations from Boulkedid et al [9]. In our review, very few studies (3/12, 25%) reported process assessments (eg, level of engagement) or panelist satisfaction (Table 1) [11,14,15,21]. The 3 (25%) studies that did report an assessment of process quality focused on narrowly defined characteristics of satisfaction [11,14,15,21]. For example, one study reported that panelists would have liked more time to discuss ideas in a conference call [11]. In another study, it was reported that most panelists were satisfied with their degree of anonymity throughout the rating rounds [21]. The third study reported most panelists felt the web-based RAM process was "suitable for achieving consensus" [14,15].

Discussion

Principal Findings

Web-based RAM panels are increasingly used in health research as an effective, efficient, convenient, and acceptable alternative to traditional consensus processes [22,23]. Previous systematic reviews have assessed the implementation and reporting of the in-person or "modified" Delphi method in research settings [9]. Despite the growing prominence of the virtual RAM, there has been no literature review of design and conduct using completely virtual methods. Documentation, however, is vitally important for researchers to replicate RAM procedures and learn and improve the process across studies. It is also important to assess the validity and applicability of the process and to interpret the results of these studies. Our narrative review of the web-based RAM process in health research helps to fill this gap. Our results show that studies generally provide little information about how the web-based RAM was implemented, making it difficult to interpret and compare study results. After summarizing the main findings of our literature review, we suggest preliminary recommendations for ways to improve the implementation and reporting of virtual RAMs.

The first contribution of this study is to illustrate the underreporting of the web-based RAM process in health research. Our narrative review of 12 unique studies revealed that the vast majority provided only brief descriptions of how their virtual RAM process was implemented. For example, half of the studies did not report a discussion period between rating rounds even though this is a standard feature in the RAM panel process. Adequate time for discussion between rating rounds is necessary for reviewing the distribution of rating results and adjusting the list of clinical indications if necessary. Additionally, although all studies reported the number of panelists who participated in each round, no studies reported consistency of participation across rounds; this is important information to assess the quality and nature of recommendations, ideally generated by highly engaged panel members who consistently participated across the rounds.

Without a common framework for reporting results from web-based RAMs, it is difficult to compare the results across studies. Improved intentionality in designing and transparency in reporting would yield improved results for individual study teams while also allowing external researchers to learn, understand, and build on the process that was used to generate a given set of expert recommendations. Thus, we offer preliminary recommendations for ways the broader field could improve the consistency of the implementation of web-based RAND (RAND Corporation) processes and considerations for individual research teams in designing and reporting on their web-based expert panel studies. We hope these recommendations serve as a launching point for continued development to improve the implementation and reporting of web-based RAMs.

Preliminary Recommendation 1: Establish Data Collection and Reporting Standards for Web-Based RAM Panels

In the intervening decades since the RAM was developed, this method was continuously refined through its practical application in a wide variety of research settings. It was not until 2001 that the RAND Corporation issued a specific set of recommendations and guidelines for implementing the RAM [1]. However, the availability of this guidance alone does not ensure consistency.

As the systematic review of the Delphi Method by Boulkedid et al [9] revealed, there is still considerable variation in implementation and reporting among research teams using modified versions of the more established method. Thus, we recommend that a professional organization convene a group of experts (eg, journal editors, practitioners, RAM users, etc) to formulate a parallel set of best practices that mirror those developed for the in-person Delphi and Rand/UCLA Appropriateness expert panels. Research teams should clearly describe the data collected and any methodological modifications made to the standard RAM. Multimedia Appendix 3 can be used as a template to report these changes for a single study. Because of word limitations, it may be necessary for researchers to develop a separate protocol paper or to report details in appendices that accompany published findings. In our own work in which we conducted a technology-based RAND expert panel, we reported most of the suggested data elements in an extensive array of supplemental files [24]. Transparent and comprehensive reporting of web-based methods will promote the reproducibility of web-based RAM processes. Based on our own experience, and drawing on results from our literature review, we offer 2 additional recommendations.

Preliminary Recommendation 2: Establish Measures of Process Quality

We encourage researchers to develop and assess measures of process quality. Process quality can be assessed by asking expert participants to complete a survey at key points throughout the panel process or at its conclusion. Ideally, this would be carried out in the same web-based platform used to host the expert panel. Based on our own experience leading RAND panels, it is feasible to elicit this feedback. This feedback can yield useful quantitative and qualitative data (from numerical ratings and open-text feedback), which researchers can use to refine future rounds and evaluate ongoing processes, as well as using them for future planning purposes [24]. This review found scant reporting of such measures. Table 1 shows the few studies that reported facets of process quality (3/12, 25%), which did so for only narrowly defined assessments, including the need to discuss ideas over a conference call [13], satisfaction with rating anonymity [21], and the suitability of the process for "achieving consensus" [14,15]. Boulkedid et al [9] did not include these types of measures in their systematic review of the Delphi Method. We recommend eliciting and reporting panelist satisfaction as an indicator of process quality. Panel members may rate their satisfaction with all aspects of the RAM process, including the following: (1) background materials provided (if any); (2) process for revising indications; (3) meeting facilitation; (4) the web-based software used; and (5) their likelihood of participating in a similar process again. We also recommend reporting consistency of participation across rounds of the web-based RAM (eg, of the individuals participating in the first round, the percentage who also participated in subsequent rounds, and whether new participants were added across the rounds). Consistency is an important indicator of the depth of commitment and depth of thought as individuals consider and reconsider ratings across multiple rounds.

Preliminary Recommendation 3: Consider Other Viable Technical Platforms With Similar Levels of Functionality

Most studies included in our review used a web-based platform to conduct RAND panels. Although ExpertLens (RAND Corporation) was the most often used web-based platform among the studies in our review, our own experience using Group System's ThinkTank suggests there are other platforms with similar or perhaps expanded levels of functionality that could be considered [24,25]. A web-based hosting platform should have the following functions: (1) allow teams to engage, collaborate with, and capture and organize input from a large number of individuals; (2) allow for sharing, revising, organizing, and analyzing content in real time or asynchronously; and (3) allow teams to export session content (eg, ratings) to formats such as Microsoft Excel for further analysis.

Limitations

Although our review identified gaps in the literature, there are limitations. The PubMed database does not search full text, so we may have missed articles that reference using web-based RAM in the main text, but not in the title or abstract. This may have limited the number of articles included in our review. In addition, our article selection criteria were narrow, which again could have limited the number of articles included in our review. We also acknowledge that the time frame for our literature search stopped just before the start of the COVID-19 pandemic. Future work should include conducting an updated review, perhaps applying our preliminary recommendations to assess reporting in more recent studies. This study is, to our knowledge, the only investigation to formally review and summarize the literature on using completely web-based RAM approaches. We recommend further development of more formal reporting standards for running web-based RAM panels. Web-based approaches have undoubtedly grown during the pandemic as more and more research has moved to internet-based platforms, a trend that is likely to remain. Our preliminary reporting recommendations may encourage other researchers to report these details to increase research transparency and replicability. This methodological transparency is important for building and expanding knowledge of best practices for conducting RAMs virtually.

Conclusion

In conclusion, conducting research virtually has become particularly important within the context of the COVID-19 pandemic due to the prohibitions and safety concerns about in-person group meetings. This shift to web-based workplaces may outlast the pandemic [26]. We also anticipate that the multiple benefits associated with web-based collaboration will make the RAM with a web-based component an appealing and cost-effective option for researchers seeking an efficient process for incorporating expert opinion into developing recommendations. This narrative review reveals underreported yet important characteristics for conducting and reporting on web-based RAM expert panels. Without a common framework for reporting results from web-based RAMs, it is difficult to compare results across studies. In this way, intentionality in designing and transparency in reporting will yield improved results for individual study teams while also allowing external researchers to understand the process that was used to generate a given set of expert recommendations. We highlight preliminary recommendations for conducting and evaluating virtual RAM approaches that will contribute to replicable high-quality findings using web-based RAMs.

Acknowledgments

This work was funded by the Department of Veterans Affairs Health Services Research and Development IIR 15-131. The funding agency had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategy.
[PDF File (Adobe PDF File), 176 KB - jmir_v24i8e33898_app1.pdf]

Multimedia Appendix 2
Criteria for literature search.
[PDF File (Adobe PDF File), 179 KB - jmir_v24i8e33898_app2.pdf]

Multimedia Appendix 3
Summaries of articles included in literature review.
[XLSX File (Microsoft Excel File), 21 KB - jmir_v24i8e33898_app3.xlsx]

References

1. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lázaro P, et al. The RAND/UCLA appropriateness method user's manual. URL: https://www.rand.org/content/dam/rand/pubs/monograph_reports/2011/MR1269.pdf [accessed 2022-07-22]
2. Barber CE, Marshall DA, Mosher DP, Akhavan P, Tucker L, Houghton K, Arthritis Alliance of Canada Performance Measurement Development Panel. Development of System-level Performance Measures for Evaluation of Models of Care for Inflammatory Arthritis in Canada. J Rheumatol 2016 Mar 15;43(3):530-540 [FREE Full text] [doi: 10.3899/jrheum.150839] [Medline: 26773106]

3. MacLean CH, Kerr EA, Qaseem A. Time Out — Charting a Path for Improving Performance Measurement. N Engl J Med 2018 May 10;378(19):1757-1761. [doi: 10.1056/nejmp1802595]
4. Chopra V, Flanders SA, Saint S, Woller SC, O'Grady NP, Safdar N, et al. The Michigan Appropriateness Guide for Intravenous Catheters (MAGIC): Results From a Multispecialty Panel Using the RAND/UCLA Appropriateness Method. Ann Intern Med 2015 Sep 15;163(6_Supplement):S1-S40. [doi: 10.7326/m15-0744]
5. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The Quality of Health Care Delivered to Adults in the United States. N Engl J Med 2003 Jun 26;348(26):2635-2645. [doi: 10.1056/nejmsa022615]
6. Meddings J, Saint S, Fowler KE, Gaies E, Hickner A, Krein SL, et al. The Ann Arbor Criteria for Appropriate Urinary Catheter Use in Hospitalized Medical Patients: Results Obtained by Using the RAND/UCLA Appropriateness Method. Ann Intern Med 2015 May 05;162(9_Supplement):S1-S34. [doi: 10.7326/m14-1304]
7. Owens DK, Qaseem A, Chou R, Shekelle P, Clinical Guidelines Committee of the American College of Physicians. High-value, cost-conscious health care: concepts for clinicians to evaluate the benefits, harms, and costs of medical interventions. Ann Intern Med 2011 Feb 01;154(3):174-180 [FREE Full text] [doi: 10.7326/0003-4819-154-3-201102010-00007] [Medline: 21282697]
8. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984 Sep;74(9):979-983. [doi: 10.2105/ajph.74.9.979] [Medline: 6380323]
9. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 2011 Jun 9;6(6):e20476 [FREE Full text] [doi: 10.1371/journal.pone.0020476] [Medline: 21694759]
10. Aljamal MS, Ashcroft D, Tully MP. Development of indicators to assess the quality of medicines reconciliation at hospital admission: an e-Delphi study. Int J Pharm Pract 2016 May;24(3):209-216. [doi: 10.1111/ijpp.12234] [Medline: 26893010]
11. Barber CE, Marshall DA, Alvarez N, Mancini GJ, Lacaille D, Keeling S, Quality Indicator International Panel. Development of Cardiovascular Quality Indicators for Rheumatoid Arthritis: Results from an International Expert Panel Using a Novel Online Process. J Rheumatol 2015 Sep 15;42(9):1548-1555 [FREE Full text] [doi: 10.3899/jrheum.141603] [Medline: 26178275]
12. Barber CE, Mosher DP, Ahluwalia V, Zummer M, Marshall DA, Choquette D, RACCD Working Group. Development of a Canadian Core Clinical Dataset to Support High-quality Care for Canadian Patients with Rheumatoid Arthritis. J Rheumatol 2017 Dec 01;44(12):1813-1822 [FREE Full text] [doi: 10.3899/jrheum.170421] [Medline: 28966205]
13. Caughey GE, Kalisch Ellett LM, Wong TY. Development of evidence-based Australian medication-related indicators of potentially preventable hospitalisations: a modified RAND appropriateness method. BMJ Open 2014 Apr 28;4(4):e004625 [FREE Full text] [doi: 10.1136/bmjopen-2013-004625] [Medline: 24776711]
14. de Souza AL, van der Sanden WJ, Leal SC, Frencken JE. The Caries Assessment Spectrum and Treatment (CAST) index: face and content validation. International Dental Journal 2012 Oct;62(5):270-276. [doi: 10.1111/j.1875-595x.2012.00121.x]
15. Frencken JE, de Souza AL, van der Sanden WJ, Bronkhorst EM, Leal SC. The Caries Assessment and Treatment (CAST) instrument. Community Dent Oral Epidemiol 2013 Feb 25;41(1):e71-e77. [doi: 10.1111/cdoe.12027] [Medline: 24916680]
16. Khodyakov D, Ochoa A, Olivieri-Mui BL, Bouwmeester C, Zarowitz BJ, Patel M, et al. Screening Tool of Older Person's Prescriptions/Screening Tools to Alert Doctors to Right Treatment Medication Criteria Modified for U.S. Nursing Home Setting. J Am Geriatr Soc 2017 Mar 23;65(3):586-591 [FREE Full text] [doi: 10.1111/jgs.14689] [Medline: 28008599]
17. Monnier A, Eisenstein B, Hulscher M, Gyssens I, DRIVE-AB WP1 group. Towards a global definition of responsible antibiotic use: results of an international multidisciplinary consensus procedure. J Antimicrob Chemother 2018 Jun 01;73(suppl_6):vi3-vi16 [FREE Full text] [doi: 10.1093/jac/dky114] [Medline: 29878216]
18. N'Goran AA, Blaser J, Deruaz-Luyet A, Senn N, Frey P, Haller DM, et al. From chronic conditions to relevance in multimorbidity: a four-step study in family medicine. Fam Pract 2016 Aug 06;33(4):439-444. [doi: 10.1093/fampra/cmw030] [Medline: 27154549]
19. Shih JJ, Whitlock JB, Chimato N, Vargas E, Karceski SC, Frank RD. Epilepsy treatment in adults and adolescents: Expert opinion, 2016. Epilepsy Behav 2017 Apr;69:186-222. [doi: 10.1016/j.yebeh.2016.11.018] [Medline: 28237319]
20. Somner JEA, Ismail R, Froud R, Azuara-Blanco A, King AJ. Consensus generation of a minimum set of outcome measures for auditing glaucoma surgery outcomes-a Delphi exercise. Graefes Arch Clin Exp Ophthalmol 2018 Dec 24;256(12):2407-2411. [doi: 10.1007/s00417-018-4140-4] [Medline: 30251199]
21. Westby M, Marshall D, Jones C. Development of quality indicators for hip and knee arthroplasty rehabilitation. Osteoarthritis Cartilage 2018 Mar;26(3):370-382 [FREE Full text] [doi: 10.1016/j.joca.2017.10.020] [Medline: 29292095]
22. Bowles KH, Holmes JH, Naylor MD, Liberatore M, Nydick R. Expert consensus for discharge referral decisions using online delphi. AMIA Annu Symp Proc 2003:106-109 [FREE Full text] [Medline: 14728143]
23. Khodyakov D, Hempel S, Rubenstein L, Shekelle P, Foy R, Salem-Schatz S, et al. Conducting online expert panels: a feasibility and experimental replicability study. BMC Med Res Methodol 2011 Dec 23;11(1):174 [FREE Full text] [doi: 10.1186/1471-2288-11-174] [Medline: 22196011]
24. Kerr EA, Klamerus ML, Markovitz AA, Sussman JB, Bernstein SJ, Caverly TJ, et al. Identifying Recommendations for Stopping or Scaling Back Unnecessary Routine Services in Primary Care. JAMA Intern Med 2020 Nov 01;180(11):1500-1508. [doi: 10.1001/jamainternmed.2020.4001] [Medline: 32926088]

25. GroupSystems: ThinkTank. URL: https://thinktank.net/ [accessed 2019-09-27]
26. Guyot K, Sawhill IV. Telecommuting will likely continue long after the pandemic. Brookings Institution. URL: https://www.brookings.edu/blog/up-front/2020/04/06/telecommuting-will-likely-continue-long-after-the-pandemic/ [accessed 2022-11-24]

Abbreviations

RAM: RAND/UCLA Appropriateness Method

Edited by G Eysenbach; submitted 28.09.21; peer-reviewed by R Krukowski, D Verran; comments to author 23.02.22; revised version received 28.06.22; accepted 18.07.22; published 26.08.22.

Please cite as:
Sparks JB, Klamerus ML, Caverly TJ, Skurla SE, Hofer TP, Kerr EA, Bernstein SJ, Damschroder LJ
Planning and Reporting Effective Web-Based RAND/UCLA Appropriateness Method Panels: Literature Review and Preliminary Recommendations
J Med Internet Res 2022;24(8):e33898
URL: https://www.jmir.org/2022/8/e33898
doi:10.2196/33898
PMID:36018626

Review

Telehealth for the Longitudinal Management of Chronic Conditions: Systematic Review

Allison A Lewinski[1,2], MPH, PhD; Conor Walsh[1,3], MD; Sharron Rushton[2], MS, DNP; Diana Soliman[4], MD; Scott M Carlson[4], MD; Matthew W Luedke[5,6], MD; David J Halpern[3,7], MPH, MD; Matthew J Crowley[1,4], MHS, MD; Ryan J Shaw[2], RN, PhD; Jason A Sharpe[1], PhD, DPT; Anastasia-Stefania Alexopoulos[1,4], MHS, MBBS; Amir Alishahi Tabriz[8], MPH, MD, PhD; Jessica R Dietch[9], PhD; Diya M Uthappa[10], BS; Soohyun Hwang[11], MPH; Katharine A Ball Ricks[12], MPH, PhD; Sarah Cantrell[13], MLIS; Andrzej S Kosinski[14], PhD; Belinda Ear[1], MPH; Adelaide M Gordon[1], MPH; Jennifer M Gierisch[1,3,15], MPH, PhD; John W Williams Jr[1,3,16], MHS, MD; Karen M Goldstein[1,3], MSPH, MD

[1]Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Health Care System, Durham, NC, United States
[2]School of Nursing, Duke University, Durham, NC, United States
[3]Division of General Internal Medicine, Department of Medicine, Duke University School of Medicine, Durham, NC, United States
[4]Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Duke University School of Medicine, Durham, NC, United States
[5]Department of Neurology, Duke University Medical Center, Durham, NC, United States
[6]Neurodiagnostic Center, Durham Veterans Affairs Medical Center, Durham, NC, United States
[7]Duke Primary Care, Duke University Medical Center, Durham, NC, United States
[8]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, United States
[9]School of Psychological Science, Oregon State University, Corvallis, OR, United States
[10]Doctor of Medicine Program, Duke University School of Medicine, Durham, NC, United States
[11]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, United States
[12]Cecil G Sheps Center for Health Service Research, University of North Carolina, Chapel Hill, NC, United States
[13]Duke University Medical Center Library, Duke University School of Medicine, Durham, NC, United States
[14]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, United States
[15]Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, United States
[16]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Allison A Lewinski, MPH, PhD
Center of Innovation to Accelerate Discovery and Practice Transformation
Durham Veterans Affairs Health Care System
Attention: Health Services Research and Development Center of Innovation (558/152)
508 Fulton Street
Durham, NC, 27705
United States
Phone: 1 919 286 2936
Fax: 1 919 416 5836
Email: allison.lewinski@duke.edu

Abstract

Background: Extensive literature support telehealth as a supplement or adjunct to in-person care for the management of chronic conditions such as congestive heart failure (CHF) and type 2 diabetes mellitus (T2DM). Evidence is needed to support the use of telehealth as an equivalent and equitable replacement for in-person care and to assess potential adverse effects.

Objective: We conducted a systematic review to address the following question: among adults, what is the effect of synchronous telehealth (real-time response among individuals via phone or phone and video) compared with in-person care (or compared with phone, if synchronous video care) for chronic management of CHF, chronic obstructive pulmonary disease, and T2DM on key disease-specific clinical outcomes and health care use?

Methods: We followed systematic review methodologies and searched two databases (MEDLINE and Embase). We included randomized or quasi-experimental studies that evaluated the effect of synchronously delivered telehealth for relevant chronic conditions that occurred over ≥2 encounters and in which some or all in-person care was supplanted by care delivered via phone or video. We assessed the bias using the Cochrane Effective Practice and Organization of Care risk of bias (ROB) tool and the certainty of evidence using the Grading of Recommendations Assessment, Development, and Evaluation. We described the findings narratively and did not conduct meta-analysis owing to the small number of studies and the conceptual heterogeneity of the identified interventions.

Results: We identified 8662 studies, and 129 (1.49%) were reviewed at the full-text stage. In total, 3.9% (5/129) of the articles were retained for data extraction, all of which (5/5, 100%) were randomized controlled trials. The CHF study (1/5, 20%) was found to have high ROB and randomized patients (n=210) to receive quarterly automated asynchronous web-based review and follow-up of telemetry data versus synchronous personal follow-up (in-person vs phone-based) for 1 year. A 3-way comparison across study arms found no significant differences in clinical outcomes. Overall, 80% (4/5) of the studies (n=466) evaluated synchronous care for patients with T2DM (ROB was judged to be low for 2, 50% of studies and high for 2, 50% of studies). In total, 20% (1/5) of the studies were adequately powered to assess the difference in glycosylated hemoglobin level between groups; however, no significant difference was found. Intervention design varied greatly from remote monitoring of blood glucose combined with video versus in-person visits to an endocrinology clinic to a brief, 3-week remote intervention to stabilize uncontrolled diabetes. No articles were identified for chronic obstructive pulmonary disease.

Conclusions: This review found few studies with a variety of designs and interventions that used telehealth as a replacement for in-person care. Future research should consider including observational studies and studies on additional highly prevalent chronic diseases.

(J Med Internet Res 2022;24(8):e37100) doi:10.2196/37100

KEYWORDS

telemedicine; diabetes mellitus, type 2; heart failure; pulmonary disease; chronic obstructive; veterans; delivery of health care; systematic review

Introduction

Background

As a means to mitigate the risk of viral transmission for both patients and clinicians during the COVID-19 pandemic, many health systems have rapidly converted ≥70% of their outpatient visits to telehealth via phone or video delivery [1-5]. To support this shift, the Centers for Medicare and Medicaid Services in the United States issued an emergency ruling to decrease regulatory requirements for telehealth and created payment parity between in-person care and telehealth delivered via phone or video [6]. Increased telehealth use during the COVID-19 pandemic provided health systems, technology companies, and health care providers experience with telehealth at scale and raised the possibility that telehealth could become a standard option in the postpandemic period. However, concerns remain that care delivered via telehealth is potentially low in quality of care, is difficult to incorporate into workflows, and can exacerbate health disparities [7-10]. Specifically, evidence is needed regarding the efficacy of telehealth as a replacement for in-person care when treating patients.

Extensive literature supports telehealth as a supplement or adjunct to in-person care for the management of chronic conditions [11] such as congestive heart failure (CHF) and type 2 diabetes mellitus (T2DM) [12-14]. These 2 highly prevalent chronic diseases are among the most common and costly conditions affecting approximately 13.4% [15] and 10.5% [16] of all adults in the US, respectively. In addition, CHF and T2DM typically require physical assessment to establish disease status and assess the presence and extent of exacerbations. However, the effects of telehealth as a replacement for in-person health care delivery for CHF, T2DM, and other chronic illnesses remain uncertain [10,17,18] Before the COVID-19 pandemic, many patients with chronic medical conditions, such as CHF, chronic obstructive pulmonary disease (COPD), and T2DM, uniformly received in-person evaluation. During the pandemic, these patients often received telehealth to unknown effect. Although telehealth can increase accessibility to health care by lowering barriers to access [19-21], few studies exist to support the use of telehealth as an equivalent and equitable replacement for in-person care, and the potential adverse effects have not been well defined [18]. Assuming that telehealth can readily replace in-person care may be inappropriate, given the scarcity of evidence examining telehealth applied in this way.

A first step to address the question of equivalence of synchronous (real time) telehealth via phone or video as a replacement for in-person care for chronic diseases is a review focused specifically on evidence from the comparative literature. If there is moderate to strong evidence that telehealth is equivalent to in-person care for patients with chronic conditions, its promise should be developed more fully and incorporated as a standard option for delivering longitudinal care. Early during the COVID-19 pandemic, there was the first complete replacement of telehealth with in-person care [1-5]. However, since then, we have started to see the routine substitution of telehealth for in-person care visits across many specialties and contexts. This substitution (meaning only video) is not usually for all care, but rather can often be a replacement for part of in-person care (some phone visits replaced by video). In addition, currently, there are multiple commercial health care providers who provide only telehealth (Teledoc and CallonDoc). It is within this context that we formulated the questions for this review.

Objective

We conducted a systematic review to summarize and report the use of telehealth as a replacement or substitute for in-person care in the context of chronic management of CHF, COPD, and T2DM. The questions guiding this review were the following:

1. Question 1a—Among adults, what is the effect of synchronous (real time) telehealth (phone or phone and video) compared with in-person care (or compared with phone, if synchronous video care) for the chronic management of CHF, COPD, and T2DM on key disease-specific clinical outcomes and health care use (eg, hospital admission, hospital readmission, and emergency room visits)?
2. Question 1b—For each disease (CHF, COPD, and T2DM), does this effect differ by race and ethnicity, gender, age, and rural status?
3. Question 2—What are the adverse effects of synchronous telehealth for the chronic management of CHF, COPD, and T2DM as compared with in-person care (or compared with phone, if synchronous video care) on patients?

Methods

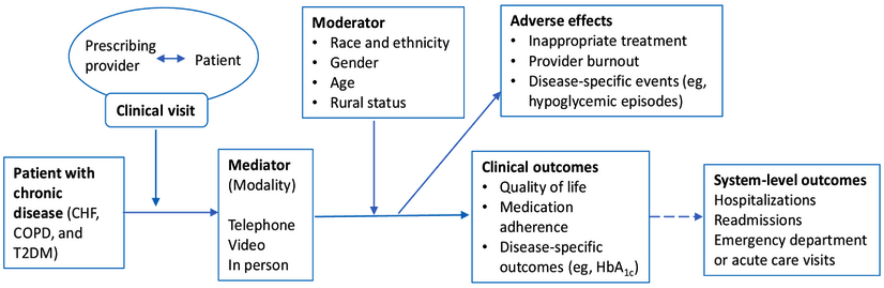
Overview

This systematic review was conducted as part of a Veterans Health Administration (VHA)–funded report [22] in response to a topic proposed by the VHA Office of Rural Health. For this review, similar to completed previous reviews and to meet the goals of the VHA as a learning health care system [23], (1) the partners from the Office of Rural Health were not involved in conducting the review, but informed topic and question development and provided contextual relevance for the study; (2) the partners from the Office of Rural Health were not involved in approving the final write-up of the report; and (3) a technical expert panel guided the conduct of the review and discussion of the findings. We developed and followed an a priori protocol for this review, and there were no significant deviations after registration (PROSPERO [International Prospective Register of Systematic Reviews] registration number CRD42021239756) [24]. Each step was pilot-tested to train and calibrate the study investigators. We adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [25].

Analytic Framework

We developed an analytic framework [26] (Figure 1) that outlined the population, outcomes, mediation effect of the care visit modality, moderation effect of patient characteristics, and any adverse effects. First, we identified clinical activities (medication management, symptom monitoring, and physical examination) for longitudinal follow-up and the ability to complete them via phone, video, or either of these for CHF, COPD, and T2DM. We determined the relevant aspects that should be abstracted from the eligible literature to obtain critical evidence about conducting a telehealth visit in any clinic setting. Then, with this foundation, we determined that the telehealth modality (eg, telephone, video, and in person) mediates the relationship between the clinical visit and prespecified clinical-level and system-level outcomes. The telehealth interventions matched with our operationalized definition of telehealth and included important contextual elements such as delivery mode (telephone, video, and in person), dose (duration and frequency of contact), and clinical context of care provision. In addition, we specified that care delivered via telehealth should be for clinical activities provided by the prescribing clinician such as for evaluation, diagnosis, or medication prescription and not for the provision of self-management education or other support provided adjunctively by a clinical team member other than the prescribing clinician (eg, nurse care manager), because such interventions have been previously evaluated [11].

Figure 1. Analytic framework to guide systematic review activities. CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease; HbA_{1c}: glycosylated hemoglobin; T2DM: type 2 diabetes mellitus.



Search Strategy

In collaboration with an expert medical librarian, we conducted a primary literature search from inception to February 7, 2021, in 2 databases (MEDLINE [via Ovid] and Embase [via Elsevier]). We used database-specific subject headings and keywords to search for relevant titles and abstracts (Multimedia Appendix 1). The search strategies were peer-reviewed by a second expert medical librarian before execution using the Peer Review of Electronic Search Strategies Checklist [27]. In addition, we manually searched previous systematic reviews conducted on this or a related topic for potential inclusion.

Study Selection

Studies identified through our primary search were classified independently by 2 investigators from the study team for relevance to the questions based on the title and abstract from our a priori established eligibility criteria. Study eligibility criteria were organized by population, intervention, comparator, outcome, timing, and setting elements and other criteria such as study design, language, and publication type (Table 1). All studies classified for inclusion by at least one investigator were reviewed at the full-text level. The studies designated for exclusion by one investigator at the title and abstract level were screened by a second investigator. If both investigators agreed on exclusion, the study was excluded. Full-text review included 2 independent reviewers. Conflicts were resolved via discussion. All articles that met the eligibility criteria at the full-text level were included for data abstraction. All results were tracked in an electronic database (EndNote [Clarivate Analytics] for referencing and DistillerSR [Evidence Partners Inc] for data abstraction).

Table 1. Study eligibility.

Study characteristics	Inclusion criteria	Exclusion criteria

Population	• Adults (aged ≥18 years) with the following chronic conditions:
	  • CHF[a]
	  • COPD[b]
	  • T2DM[c]; at least 75% of the sample, if it is a mix of type 1 and type 2
	• Clinicians or clinics providing telehealth for chronic conditions, if relevant to adverse effects associated with CHF, COPD, and T2DM
		• Inpatient populations (eg, tele-ICU[d])
		• Patients receiving care in an ER[e] or tele–urgent care setting
		• Intervention limited only to the management of complications of CHF, COPD, and T2DM such as stroke, retinopathy, neuropathy, and foot ulcers

Intervention	• Synchronous care delivered over ≥2 encounters for the long-term management of relevant chronic conditions in which some or all in-person care is supplanted by telehealth (phone or video) and which is delivered remotely by an independently licensed clinician • May include asynchronous telehealth tools (eg, remote monitoring systems), if in both arms	• Supplemental nurse care management • Telehealth interventions that do not involve synchronous care delivered by a clinician to a patient (eg, 1-way SMS text messages and reminder systems) • Telecardiac or telepulmonary rehabilitation
Comparator	• In-person care without any telehealth delivery or care delivered via telephone, if compared with video	• No comparator

Outcome	• Key clinical outcomes (eg, medication adherence, quality of life, and depression) according to condition:
	  • CHF—for example, NYHA[f] functional classification
	  • COPD—for example, exercise tolerance and dyspnea
	  • T2DM—for example, HbA_{1c}[g] level
	• Clinical use (hospitalization, hospital readmissions, and ER visits or urgent care)
	• Adverse effects (eg, hypoglycemic episodes, inappropriate treatment, and clinician burnout)
		• Outcomes other than those listed in the inclusion criteria

Timing	• No limit	• N/A[h]
Setting	• Any outpatient setting (general medical or specialty care clinic)	• Intervention delivered primarily in hospital inpatient setting (including ER)
Study design	• Studies that meet the EPOC[i] criteria and have prospective data collection, such as the following: • Randomized controlled trials • Nonrandomized trials • Controlled before-after studies • Interrupted time series studies or repeated measures studies	• Not a clinical study (eg, editorial and letter to an editor) • Uncontrolled clinical study • Qualitative studies • Prospective or retrospective observational studies • Clinical guidelines • Measurement or validation studies • Studies that focus on mixed chronic conditions if results for specified conditions are not reported separately
Countries	• OECD[j]	• Non-OECD
Publication types	• Full publication in a peer-reviewed journal	• Letters, editorials, reviews, dissertations, meeting abstracts, and protocols without results

[a] CHF: congestive heart failure.

[b] COPD: chronic obstructive pulmonary disease.

[c] T2DM: type 2 diabetes mellitus.

[d] ICU: intensive care unit.

[e] ER: emergency room.

[f] NYHA: New York Heart Association.

[g] HbA_{1c}: glycosylated hemoglobin.

[h] N/A: not applicable.

[i] EPOC: Effective Practice and Organization of Care.

[j] OECD: Organisation for Economic Co-operation and Development includes Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

Data Extraction and Quality Assessment

Data from published reports were abstracted into a customized DistillerSR database by one reviewer and overread by a second reviewer. Disagreements were resolved by consensus or by obtaining a third reviewer's opinion. Data elements included descriptors to assess applicability, quality elements, intervention details, and outcomes including adverse events. Key characteristics that were abstracted included participant descriptors (eg, race and ethnicity, gender, age, and rural status), intervention characteristics (eg, clinician type and telehealth modality), comparator, and outcomes (eg, glycosylated hemoglobin [HbA_{1c}] level, hospital admission, emergency department visits, and New York Heart Association functional classification). We abstracted all outcomes that were used to evaluate telehealth, but prioritized outcomes identified a priori in collaboration with our partners from the Office of Rural Health and technical expert panel for analysis. Multiple reports from a single study were treated as a single data point, prioritizing results based on the most complete and appropriately analyzed data. When critical data were missing or unclear in the published reports, we requested supplemental data from the study authors. We emailed the authors of 1.6% (2/129) of the studies to obtain additional information and did not receive a reply from any of them. When we did not have sufficient information, we left the field blank.

The investigators who participated in data extraction also completed the quality assessment. Disagreements were resolved by consensus between the 2 investigators or, when needed, by arbitration by a third investigator. For randomized, nonrandomized, and controlled before-after studies, we used the criteria from the Cochrane Effective Practice and Organization of Care (EPOC) risk of bias (ROB) tool [28]. We assigned a summary ROB score (low, unclear, or high) to individual studies. Among the investigators, no ROB disagreements occurred owing to missing results in a synthesis.

The certainty of evidence for each question was assessed using the approach described by Grading of Recommendations Assessment, Development, and Evaluation [29]. We limited the Grading of Recommendations Assessment, Development, and Evaluation ratings to the questions that had at least two included studies. In brief, this approach requires the assessment of four domains: ROB, consistency, directness, and precision. Additional domains to be used when appropriate are coherence, dose-response association, impact of plausible residual confounders, strength of association (magnitude of effect), and publication bias. We considered these domains qualitatively and assigned a summary rating as high, moderate, or low strength of evidence after discussion by a subteam of 5 investigators. In some cases, high, moderate, or low ratings were impossible or imprudent to be provided. In these situations, a grade of *insufficient* was assigned.

Subgroups of Interest

The research questions guided our subgroup analysis. Prespecified potential effect modifiers included study design characteristics (eg, allocation concealment), disease context (CHF, COPD, or T2DM), and intervention type (eg, telehealth modality). Regarding patient-level characteristics of interest (race and ethnicity, gender, age, and rural status), we looked for analyses conducted within the primary literature that sought to identify effect modifications (eg, subgroup analyses and regression model explanatory variables). Manuscripts included in this review did not specify descriptions of gender or sex. For consistency, we use *gender* throughout the *Results* and *Discussion* sections because the interventions examined are more relevant to self-identity and not specific to one's biology at birth. However, we realize that this terminology may not reflect patients who would not have self-identified as such.

Data Synthesis and Analysis

We summarized the primary literature using relevant data abstracted from the eligible studies. Summary tables describe the key characteristics of the primary studies: study design, patient demographics, and details of the intervention and comparator. Owing to conceptual heterogeneity related to the structure, purpose, and delivery of telehealth visits, we did not conduct a meta-analysis, but rather described findings narratively, focusing on identifying patterns in the efficacy and safety of the interventions across conditions and outcome categories.

Continuous outcomes were summarized using the mean patient-level difference (follow-up minus baseline) when the outcome was reported using the same scale. For studies that did not directly report the mean and SD of patient differences, we used the difference in means between the follow-up and baseline. For 20% (1/5) of the studies [30], we computed the SD of the difference based on the reported *P* value for the difference between the 2 arms, assuming the same correlation between follow-up and baseline in each arm. When studies reported only medians and ranges, we translated them into means and SDs [31], and if a study reported only baseline SD, we assumed the same SD at follow-up. Finally, in the absence of other information, we assumed a conservative 0.5 correlation between the follow-up and baseline measures.

Ad hoc Horizon Scan to Identify Relevant Studies in Progress

Given the limited amount of existing literature we identified that addressed our questions, we sought to assess the pool of ongoing studies that would add relevant findings in the near future. To conduct such a scan of the literature on the horizon, we applied our previously developed search terms to the Cochrane Central Register of Controlled Trials. Notably, we did not apply the same rigor to this process as for our primary search process. At least one reviewer screened the studies identified through the horizon scan at the title and abstract level, and all the included studies were verified by a second reviewer.

Results

Overview

The search identified 11,245 studies from the 2 databases (Figure 2). After deduplication, 77.03% (8662/11,245) of the articles underwent the screening process. In total, 0.06% (5/8662) of the studies met the inclusion criteria. Of those 5 studies, 4 (80%) focused on diabetes and 1 (20%) focused on CHF. The details of the included studies are provided in Table 2. We have provided the details of study characteristics (Multimedia Appendix 2), intervention characteristics (Multimedia Appendix 3), all outcomes reported in the included studies (Multimedia Appendix 4), and excluded studies and the reason for exclusion (Multimedia Appendix 5). Common reasons for excluding studies by intervention included telehealth that supplemented rather than replaced in-person care, telehealth interventions delivered by nonprescribing clinicians, and telehealth delivered asynchronously only. In the following sections, we describe the results by chronic disease (CHF, COPD, and T2DM). The certainty of evidence for the included studies is presented in (Table 3).

Figure 2. Literature flowchart. *Search results from MEDLINE (4713) and Embase (3949) were combined. CHF: congestive heart failure; N/A: not applicable; OECD: Organisation for Economic Co-operation and Development.

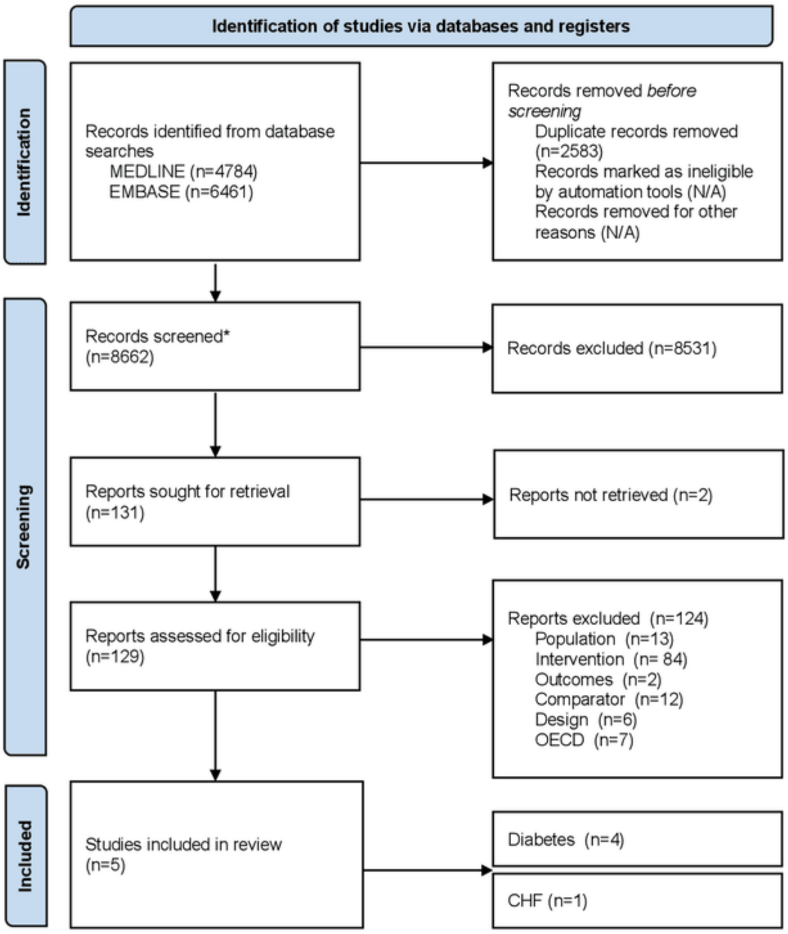


Table 2. Evidence profile of included studies.

Criteria	Study information
Region or location (N=5), n (%)	
United States	2 (40)
Europe	2 (40)
Asia	1 (20)
Disease in focus (N=5), n (%)	
T2DM[a]	4 (80)
CHF[b]	1 (20)
COPD[c]	0 (0)
Patient demographics (N=676)[d]	
Age (years), median	58
Gender, n (%)	
Women	168 (24.9)
Men	508 (75.1)
Race (N=60), n (%)	
White[e]	52(87)
Black[e]	6 (10)
Hispanic[e]	1 (2)
Other[e]	1 (2)
Intervention mode (N=5) , n (%)	
RM[f] and video	1 (20)
Video	2 (40)
RM and telephone	1 (20)
Telephone	1 (20)
Comparisons[g] (N=5), n (%)	
RM and in-person care	2 (40)
Usual in-person care	3 (60)
Outcomes reported (N=5), n (%)	
HbA_{1c}[h] level	4 (80)
NYHA[i] functional classification	1 (20)
Hospitalization	3 (60)
Emergency department visit	2 (40)
Risk of bias—objective (N=5), n (%)	
High	2 (40)
Unclear	1 (20)
Low	2 (40)
Risk of bias—reported by patient (N=5), n (%)	
High	2 (40)
Unclear	1 (20)
Low	1 (20)

Criteria	Study information
N/A[j]	1 (20)

[a] T2DM: type 2 diabetes mellitus.

[b] CHF: congestive heart failure.

[c] COPD: chronic obstructive pulmonary disease.

[d] Of the 5 studies, 1 (20%) study [32] reported 50% (338/676) of the participants.

[e] In total, 80% (4/5) of the studies did not report this information.

[f] RM: remote monitoring.

[g] For this criterion, ≥1 category is possible per study.

[h] HbA_{1c}: glycosylated hemoglobin.

[i] NYHA: New York Heart Association.

[j] N/A: not applicable.

Table 3. Certainty of evidence for included studies of CHF[a] and T2DM[b].

Outcomes	Studies (randomized controlled trials; N=5), n (%)	Patients (N=676), n (%)	Range of effects	*P* value	Certainty of evidence (rationale)
T2DM					
HbA_{1c}[c] level	4 (80)	339 (50.1)	Mean difference of −0.15% to −1.30% in the HbA_{1c} level between the intervention and comparator arms	N/A[d]	Very low certainty that telehealth has an effect on HbA_{1c} level (rated down for serious risk of bias, indirectness, and imprecision)
Hospital admission	2 (40)	285 (42.2)	In total, 0 to 3 admissions in the intervention arm and 0 to 7 admissions in the comparator arm	N/A	Very low certainty that telehealth has an effect on hospital admissions (rated down for serious risk of bias, indirectness, and imprecision)
Emergency department visits	2 (40)	285 (42.2)	In total, 0 emergency department visits in the intervention arm and 0 to 1 visit in the comparator arm	N/A	Very low certainty that telehealth has an effect on emergency department attendance (rated down for serious risk of bias, indirectness, and imprecision)
CHF					
NYHA[e] functional classification	1 (20)	219 (32.4)	Between-group difference	.97	Very low certainty that telehealth has an effect on NYHA functional classification (rated down for serious risk of bias, inconsistency, indirectness, and imprecision)
Hospital admission	1 (20)	219 (32.4)	RM[f] (9.8%), RM and phone (11.3%), and in-person visit (12.7%)	.85	Very low certainty that telehealth has an effect on hospital admission (rated down for serious risk of bias, inconsistency, indirectness, and imprecision)

[a] CHF: congestive heart failure.

[b] T2DM: type 2 diabetes mellitus.

[c] HbA_{1c}: glycosylated hemoglobin.

[d] N/A: not applicable.

[e] NYHA: New York Heart Association.

[f] RM: remote monitoring.

Question 1a (Effect of Telehealth) and 1b (Differences by Special Population)

Findings for CHF

Question 1a: Effect of Telehealth

Overview

We identified only 20% (1/5) of studies that met the inclusion criteria for synchronous telehealth for chronic CHF management [33] and found it to have high ROB. The study was conducted in Germany, enrolled 210 patients, and had a duration of 12 months. The study incorporated phone-based appointments and follow-up in patients with CHF with recent placement of an implanted cardioverter defibrillator or cardiac resynchronization therapy defibrillator. Patients were randomized to receive asynchronous web-based automated review and follow-up of telemetry data alone every 3 months (n=102) or personal physician contact every 3 months in addition to remote monitoring. The personal contact group was further randomized to personal contact via telephone calls (n=53) or personal contact via in-person visits (n=55). In this study [33], the primary outcome was the proportion of patients with worse Packer Heart Failure Clinical Composite Response scores at 13 months compared with scores at 1 month after device placement. The Packer Heart Failure Clinical Composite Response score provides stepwise assessment and incorporates CHF death or hospitalization, change in New York Heart Association class, and self-assessed health status. The secondary outcomes in this study were all-cause mortality, CHF-related hospitalizations, arrhythmias, and changes in reported quality of life. We present the detailed results by outcome: (1) Packer Heart Failure Clinical Composite Response Score, (2) hospitalizations, (3) emergency department visits, and (4) number of contacts and use.

Packer Heart Failure Clinical Composite Response Score

The primary outcome of the study by Hansen et al [33] showed no significant differences in Packer scores in a 3-way comparison between the telemetry arm compared with the personal contact subgroups (remote monitoring and phone call vs remote monitoring and in-person visit; *P*=.97).

Hospitalizations

The authors found no significant differences between the subgroups in any of the outcomes that were measured. Outcomes between study arms included the following: mortality (*P*=.65), CHF-related hospitalization (*P*=.85), detection of supraventricular tachycardia (*P*=.22), detection of ventricular tachycardia (*P*=.75), and reported change in quality of life (*P*=.72).

Emergency Department Visits

The CHF study that was included did not report on emergency department visits [33].

Number of Contacts and Use

The CHF study compared the number of unscheduled follow-up visits conducted either via phone or in person among the telemetry only group, telemetry and phone visit group, and telemetry and in-person visit group [33]. In total, there were 219 unscheduled follow-ups among the 3 groups, involving 83 patients during the course of the study. However, there were no significant differences in the unscheduled follow-up rates among the 3 groups (*P*=.29).

Question 1b: Differences by Special Population

The 20% (1/5) studies that met the inclusion criteria [33] described the age (overall mean 63.8 years, SD was not reported by authors) and gender of their patient population (84.3% were men); however, details regarding race and ethnicity and rural status were not reported. Furthermore, the authors did not perform any subgroup analyses to examine the effect of age or gender on outcomes.

Findings for COPD

No studies that addressed the use of telehealth as a substitute for in-person chronic management of COPD were identified.

Findings for T2DM

Question 1a: Effect of Telehealth

Overview

We identified 80% (4/5) of studies—all of which were randomized controlled trials [30,32,34,35]—that evaluated the provision of synchronous telehealth compared with in-person care for the chronic management of T2DM. Of the 4 studies, 2 (50%) studies were conducted in the United States [30,35], 1 (25%) in South Korea [32], and 1 (25%) in Denmark [34]. Overall, 25% (1/4) of the studies were conducted with patients in the military [35]. Intervention duration varied across studies, from <8 weeks to 52 weeks. Intervention approach varied across all the studies (4/4, 100%) regarding duration and mode of incorporating telehealth into chronic diabetes management. Of the 4 studies, 3 (75%) studies included ≤60 patients [30,34,35] and 1 (25%) study included 338 patients [32]. Of the 4 studies, 3 (75%) studies used technology that facilitated synchronous bidirectional communication between the patient and clinician [32,34,35], and 1 (25%) study relied on telephone and email [30]. In total, 50% (2/4) of the studies included remote monitoring in addition to synchronous telehealth [32,35]. We present the detailed results by outcome: (1) HbA_{1c} level, (2) hospitalizations, (3) emergency department visits, and (4) number of contacts and use.

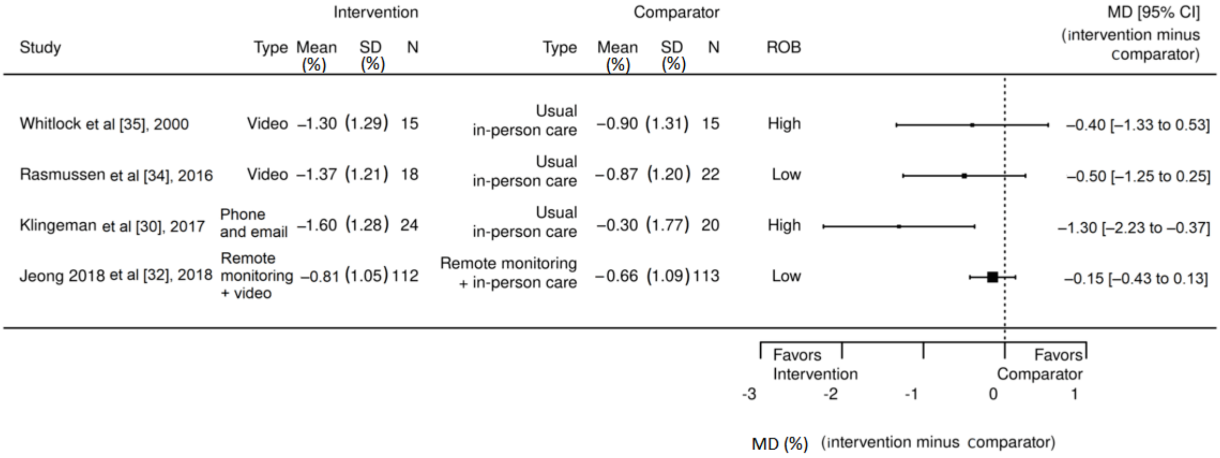
Change in Reduction of HbA_{1c} Level

All the studies (4/4, 100%) compared the change in reduction of HbA_{1c} level from baseline to the end of the study between synchronous telehealth and in-person study arms (Figure 3) [30,32,34,35]. The first study, by Jeong et al [32], was a 24-week 3-arm trial that compared usual care, telemonitoring (remote monitoring with automated clinical decision support with in-person endocrine follow-up appointments), and telemedicine (remote monitoring with automated clinical decision support with video-based endocrine follow-up appointments). Notably, that study was the largest study included and was rated as having low ROB. They enrolled 338 patients, with a baseline mean age of 53 years (SD was not reported by authors). No statistically significant difference was seen at baseline for HbA_{1c} level across groups: usual care (mean 8.39%, SD 1.10%), telemonitoring (mean 8.21%, SD 0.93%), and telemedicine (mean 8.39%, SD

1.10%). Statistically significant difference was seen for within-group reduction in HbA$_{1c}$ level from baseline to 24 weeks for all groups, ranging from −0.66% to −0.81% (P<.001). No statistically significant difference was noted for the extent of HbA$_{1c}$ reduction across groups: usual care versus telemonitoring groups (P=.61), usual care versus telemedicine groups (P=.16), and telemonitoring versus telemedicine groups (P=.34).

The second study, led by Klingeman et al [30], was a 52-week, 2-arm trial consisting of usual endocrine care versus an experimental endocrine clinic group that enrolled 60 patients with T2DM. The setting for the study was an endocrinology clinic at an academic medical center, where patient care was provided by endocrinologists. Patients who were not in the experimental arm received the usual care provided by the clinic's endocrinologists. The specialty clinic model in the experimental group included an endocrinologist and nurse educator who focused on patients with advanced diabetes; contact with the patients in this arm was designed to be variable and patient-specific. Preplanned contacts (via email and phone) were determined at baseline and amended over time, and ad hoc in-person visits occurred if clinically required. Contact was individually tailored based on each patient's outcomes, adverse reactions, and changes in the disease state. The control arm received the usual endocrine care, which included the ability for the patients to contact (via email and phone) clinicians as needed. HbA$_{1c}$ levels were compared between groups at baseline for usual care (mean 8.9%, SD 0.8%) versus specialty clinic model (mean 9.5%, SD 0.9%). In addition, high proportion of patients who were White were enrolled in the intervention group (96.6%) compared with the usual care group (76.8%). Analysis of data at 52 weeks found great reduction in HbA$_{1c}$ level in the specialty clinic model (−1.7%; from 9.6% to 7.9%) as compared with the usual endocrine care (0.3%; from 8.9% to 8.6%), with P=.004. Notably, sensitivity analysis was conducted that dropped data from a patient who was an outlier in the usual care group, with worsened HbA$_{1c}$ values (from 8.3% to 13.5%); however, this did not change the results.

Figure 3. Change in glycosylated hemoglobin levels between intervention and comparator arms across type 2 diabetes mellitus studies. MD: mean difference; ROB: risk of bias [30,32,34,35].



The third study, by Rasmussen et al [34], was a 2-arm trial comparing 3 weeks of brief standard in-person endocrine care versus telemedicine (video-based endocrine care) to stabilize patients with poorly controlled T2DM. They enrolled 40 patients with baseline HbA$_{1c}$ level of 8.1% (range 6.1%-10.7%) in standard care group and 9% (range 7.6%-12%) in the telemedicine group. At 6 months, the HbA$_{1c}$ level ranged from 8.1% to 7.2% in the standard care group and from 9.1% to 7.7% in the telemedicine group. The patients in the telemedicine arm experienced a larger decrease in HbA$_{1c}$ level (14.6%) than those in the standard care arm (10.6%), which was statistically significant (P=.02). Notably, although this study framed its hypothesis as "the treatment by telemedicine at home was similar to standard care," the analysis methods did not use noninferiority analytic approaches.

The fourth study, by Whitlock et al [35], which tested usual care and telemonitoring visits with a case manager and physician, enrolled 28 patients in a 36-week 2-arm trial consisting of a standard of care control versus experimental telemonitoring group. In this study, both groups were referred for multidisciplinary diabetic education classes. The experimental group received weekly telemonitoring via video from a case manager and, then, monthly telemonitoring via video from study physicians. Patients in the standard of care group received routine in-person care from their primary care clinician. Statistically significant within-group difference (P=.05) was noted for the experimental telemonitoring arm, from baseline HbA$_{1c}$ level of 9.5 (range 8.1-12.6) to week-36 HbA$_{1c}$ level of 8.2 (range 5.7-10.2). For the comparator, the mean baseline HbA$_{1c}$ level was 9.5 (range 8.1-11.9) and week-36 HbA$_{1c}$ level was 8.6 (range 7.1-11.9), which was not statistically significant.

Hospitalizations

In total, 50% (2/4) of the studies examined hospitalizations [30,32]. In the study by Jeong et al [32], only 1 patient in the telemonitoring arm experienced a diabetes complication–related hospitalization, and none of the patients in the control or telemedicine arms experienced diabetes-related hospitalizations. In the second study, by Klingeman et al [30], 10% (3/30) of the patients in the experimental arm and 23% (7/30) of the patients in the control arm experienced diabetes-related hospital admission.

Emergency Department Visits

Overall, 50% (2/4) of the studies examined emergency department visits [30,32]. In the first study, by Jeong et al [32], across the 3 study arms, none of the patients experienced diabetes-related visits to the emergency department out of the 338 patients enrolled in the study. In the second study, by Klingeman et al [30], none of the patients in the experimental arm and 1 patient in the control arm experienced a T2DM-related emergency department visit.

Number of Contacts and Use

In total, 75% (3/4) of the studies reported collecting data on number of contacts and use [30,34,35] among patients receiving in-person or telehealth. The study by Klingeman et al [30] reported on (1) diabetes education referrals, (2) diabetes-related visits, (3) use of modality, and (4) number of interactions and HbA_{1c} level. Klingeman et al [30] designed the experimental arm for variable frequency of contact using a specialty clinic model. Preplanned contacts (via email, phone call, or visit) were determined at baseline and amended over time; contact was tailored based on each patient's outcomes, adverse reactions, and changes in the disease state; and the control arm received usual endocrine care. Klingeman et al [30] reported that when diabetes education visits were combined with clinician's diabetes-related visits in the endocrinology clinic, the experimental group had fewer overall visits than the control group. Specifically, the experimental group had 1.5 (SD 0.7) visits and the control group had 3.6 (SD 4) visits over 12 months (*P*<.001). However, the experimental group had significantly more email contacts (mean 11.1, SD 6.4) than the control group with (mean 1.8, SD 3.5; *P*<.001; *note that email communication was a focus in the experimental group*).

The study by Rasmussen et al [34], which tested standard care and video consultation for home treatment of T2DM, reported on (1) number of visits and missed visits and (2) consultation time. The telemedicine group had an average of 4.1 visits, with no missed visits; however, the usual care group had an average of 3.8 visits, with 13% missed visits. Regarding consultation time, the telemedicine group had an average of 18 minutes and the usual care group had an average of 23 minutes. The study by Whitlock et al [35] reported no results on the number of contacts and use, despite describing collecting the number of clinic visits before and during the study in their Methods section.

Question 1b: Differences by Special Population

Only 25% (1/4) of the included studies reported on subgroup analysis [32] by patient characteristics. Jeong et al [32] analyzed two subgroups of a priori interest: gender and age. No statistically significant difference in reduction of HbA_{1c} level was found for men (mean −0.76%, SD 1.11% for telemonitoring vs mean −0.89%, SD 1.12% for telemedicine; *P*=.88) or women (mean −0.46%, SD 1.05% vs mean −0.63%, SD 0.87%; *P*=.16). No statistically significant difference in reduction of HbA_{1c} level was seen among people aged <55 years (mean −0.63%, SD 1.26% for telemonitoring vs mean −0.87%, SD 1.15% for telemedicine; *P*=.21) or among those aged ≥55 years (mean −0.68%, SD 0.88% for telemonitoring vs mean −0.73%, SD 0.93% for telemedicine; *P*=.83). Moreover, Jeong et al [32] reported on additional subgroups of potential interest. Users with high compliance (defined as users with >90% of number of records or data transmitted compared with recommended number of records) had no difference in reduction of HbA_{1c} level when compared with those with low compliance levels across the study arms of interest (mean −0.93%, SD 0.99% for telemonitoring vs mean −1.08%, SD 0.96% for telemedicine; *P*=.47). Similarly, there was no significant difference in the reduction of HbA_{1c} level between patients who had a high school education or less in the telemonitoring (mean −0.65%, SD 0.93%) and telemedicine (mean −0.94%, SD 1.1%) arms (*P*=.26).

Question 2: Adverse Events

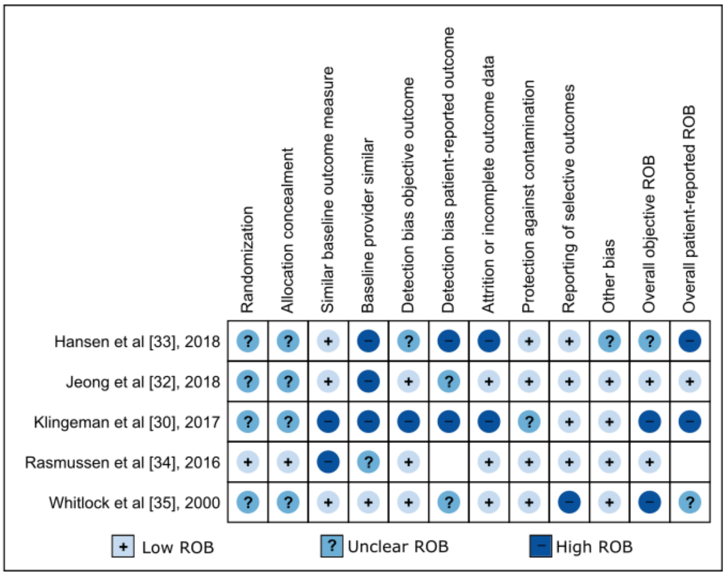
The 20% (1/5) of the studies of CHF, by Hansen et al [33], did not report on adverse events. The 40% (2/5) of studies of T2DM reported adverse events [30,32]. Jeong et al [32] described four groups of adverse events: (1) general events, (2) diabetes-related events, (3) serious events, and (4) biochemical events. Adverse events were noted in the control (n=33 or 29.20%, in-person appointments at 8, 16, and 24 weeks), telemonitoring (n=30 or 26.55%, in-person appointments at 8, 16, and 24 weeks, with remote monitoring of blood glucose data), and telemedicine (n=23 or 20.54%, video visits at 8 and 16 weeks, in-person visits at 24 weeks) arms. Diabetes-related events were noted in the control (n=7 or 6.19%), telemonitoring (n=7 or 6.19%), and telemedicine (n=3 or 2.68%) arms. Serious reported adverse events were noted in the control (n=2 or 1.7%), telemonitoring (n=2 or 1.70%), and telemedicine (n=1 or 0.90%) arms, and it included angina pectoris, rotator cuff syndrome, malignant hepatic neoplasm, skin ulcer, and hematuria [32]. Biochemical parameters for serum alanine aminotransferase (ALT), aspartate aminotransferase, and creatinine levels were measured, and samples were obtained at baseline and 24 weeks [32]. Comparing the relative percentage of patients with worsened laboratory values, ALT was the only parameter that showed significant worsening between the telemedicine and telemonitoring groups. Specifically, none of the participants in the telemonitoring arm and 7 participants in the telemedicine arm (6.7%; *P*=.01) experienced worsening of ALT values. Klingeman et al [30] described two types of adverse events: (1) severe hypoglycemia and (2) foot ulcers. Severe hypoglycemia was noted in the experimental (n=1 or 3.3%) arm, but not in the control (n=0 or 0%) arm. Foot ulcer was noted in the experimental (n=1 or 3.3%) and control (n=3 or 10%) arms.

Quality of Evidence for Included Studies

The 20% (1/5) of the studies of CHF [33] that met our inclusion criteria was rated as having high ROB owing to low numbers of patients enrolled, unclear method for patient randomization, and poor description of both patient dropout and how primary outcomes were assessed. Among the 80% (4/5) of randomized T2DM studies, the ROB (Figure 4) for patient-reported outcomes was judged to be low for 1 (25%) study, unclear for 1 (25%) study, and high for 1 (25%) study and 1 (25%) study did not report this type of outcome [30,32,34,35]. For objective outcomes, ROB was judged to be low for 50% (2/4) of the studies [32,34] and high for 50% (2/4) of the studies [30,35]. Patterns that led to judgments of low ROB (Figure 5) included

(1) noting randomization of study participants, (2) collecting objective outcome data, and (3) generally limited expected impact of bias from patient knowledge of the treatment arm. Patterns that led to high ROB included (1) missing or unclear data on randomization methods, data collection, and analysis; (2) unblinded treatment arm; (3) absence of predetermined intervention assessment patterns in the protocol; (4) unclear primary outcomes; and (5) missing or unclear reporting of patient-reported outcomes.

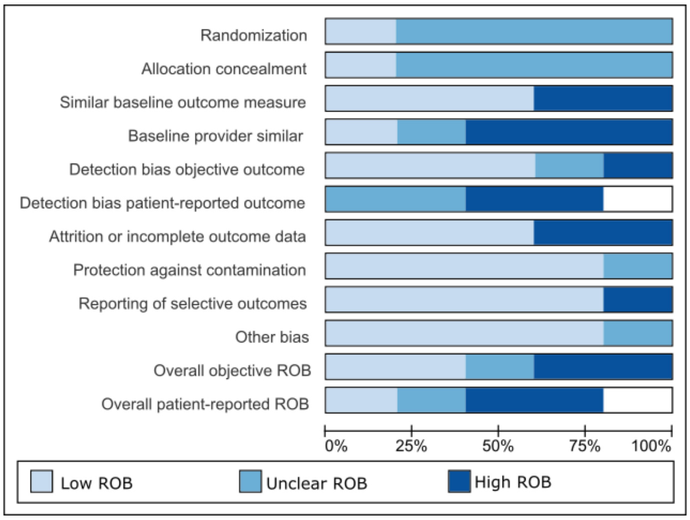
Figure 4. Risk of bias (ROB) assessment for included studies in congestive heart failure and type 2 diabetes mellitus [30,32-35].



	Randomization	Allocation concealment	Similar baseline outcome measure	Baseline provider similar	Detection bias objective outcome	Detection bias patient-reported outcome	Attrition or incomplete outcome data	Protection against contamination	Reporting of selective outcomes	Other bias	Overall objective ROB	Overall patient-reported ROB
Hansen et al [33], 2018	?	?	+	−	?	−	−	+	+	?	?	−
Jeong et al [32], 2018	?	?	+	−	+	?	+	+	+	+	+	+
Klingeman et al [30], 2017	?	?	−	−	−	−	−	?	+	+	−	−
Rasmussen et al [34], 2016	+	+	−	?	+		+	+	+	+	+	
Whitlock et al [35], 2000	?	?	+	+	+	?	+	+	−	+	−	?

+ Low ROB ? Unclear ROB − High ROB

Figure 5. Risk of bias (ROB) assessment across included studies on congestive heart failure and type 2 diabetes mellitus (N=5).



Ad hoc Horizon Scan to Identify Relevant Studies in Progress

This search identified 1787 unique studies. We found only 0.17% (3/1787) of studies [36-38] in our horizon scan that reported on studies without published results (Multimedia Appendix 6) that may potentially meet the inclusion criteria of our systematic review. All of these studies (3/3, 100%) are randomized controlled trials that were designed before the COVID-19 pandemic. Of the 3 studies, 2 (67%) studies focus on T2DM [36,37], whereas the remaining 1 (33%) study is on CHF [38]. Although the noninferiority study [37] will not meet our inclusion criteria as it is conducted in Brazil (a non–Organisation for Economic Co-operation and Development [OECD] country) and the findings may not be applicable to the US population or setting, we mention it here given the low number of studies that otherwise met our inclusion criteria. The other T2DM study [36] is specifically focused on reducing emergency diabetes care for older (aged >50 years) African Americans. The CHF study by Komkov et al [38] has very limited detail available. Although using these registries to identify trials has limitations and there are likely other relevant studies, it appears that there are few trial-based studies currently in the pipeline to inform our questions in this review.

Discussion

This review aimed to summarize and report the use of telehealth as a replacement or substitute for all or a portion of in-person care in the context of chronic management of CHF, COPD, and T2DM.

Principal Findings

We found scant evidence examining chronic disease management delivered through synchronous telehealth compared with in-person delivery for T2DM (4/5, 80%), COPD (0/5, 0%), and CHF (1/5, 20%). This suggests that there is little evidence to help guide practice on when to use telehealth instead of traditional in-person visits while managing these chronic diseases. Our review sought to include studies that used telehealth to replace all or part of in-person care. In other words, some specific in-person visits in the intervention arm were replaced by telehealth visits, whereas the comparator arm maintained all visits as in-person. Note that we consider this to be different from using telehealth as a supplement or add-on to the usual in-person care. However, we did not find any studies that only partially substituted in-person visits. We did not attempt to include studies that used telehealth as an add-on to existing in-person care, as there are already existing high-quality, peer-reviewed publications on this question [11-13]. However, despite the paucity of evidence, telehealth modalities such as video or telephone have increasingly been used to replace in-person clinic visits for managing chronic conditions, particularly during the COVID-19 pandemic [1-5]. Understanding the benefits and risks associated with shifting in-person care to telehealth is critical in shaping how health systems deliver care going forward. Although in-person visits have since increased as more has become known about COVID-19 transmission and prevention practices, telehealth continues to play a much larger role in outpatient care than before the pandemic [39,40].

Comparison With Previous Studies

Evidence indicates that telehealth can be used effectively as an adjunctive or supplemental approach to in-person care. A recent review by Albritton et al [18] examined the impact of video teleconferencing visits on prevention and management of chronic illness. Results from that review indicated that video teleconferencing resulted in similar clinical effectiveness as in-person care for certain diseases [18]. The results from our review differ in indicating clinical effectiveness of telehealth from those of Albritton et al [18] owing to several differences in the review type (systematic vs rapid review), date limitations, search strategies, databases searched, and operationalization of telehealth. Our approach to identify relevant telehealth papers was broad and more comprehensive, which resulted in a large number of articles to review. Of the 7 papers included in the review by Albritton et al [18], only 1 (14%) was not captured in our search. Additional previous reviews have examined various ways of using telehealth modalities in the context of these conditions of interest, but none of them have focused on *replacing* in-person care with telehealth visits [11]. Although we found only 20% (1/5) of the studies on telehealth for chronic management of heart failure as a substitute for in-person care, previous reviews report mixed results for the impact of other supplemental types of telehealth on heart failure outcomes [41-43]. Several recent analyses on the impact of telehealth in T2DM indicated that health outcomes did not worsen because of switching to telehealth compared with those in-person clinic care [14,40,44,45]. However, there is evidence that telehealth as an *adjunctive* strategy to typical in-person care can be associated with a decrease in HbA_{1c} level in patients with both type 1 and type 2 diabetes [46-49].

Importance of Context in Telehealth Implementation

The successful incorporation of telehealth into health care delivery relies upon the fit between the telehealth modality, care delivery context, and disease management approaches [9,19,50]. Presumably, not all areas of health care delivery lend themselves equally well to telehealth, but management of certain chronic diseases (CHF, COPD, and T2DM) may provide good opportunity to replace routine in-person care with telehealth. In our review, we sought to address a critical evidence gap by examining the comparative literature on telehealth as a replacement for in-person care in chronic disease management. Interestingly, our findings came from studies that were conducted in specialty settings, and aspects of the studied telehealth interventions were often incompletely described. However, much of the long-term management of these chronic conditions occurs within the context of primary care settings. As primary care settings likely have different pressures and challenges with telehealth modalities, given the need to address multiple comorbidities during the same visit, the results from our review may not be directly applicable. Thus, we recommend future reviews to examine and provide evidence-based guidance about the effect of telehealth interventions to deliver high-quality care using the right modality for the right patients with the right clinical condition at the right time.

Additional Approaches to Examine Telehealth

A way to determine the effect telehealth is to use noninferiority analytic approaches when hypotheses focus on whether telehealth delivered care is equally effective to in-person care. Our eligibility criteria focused on randomized controlled trials and did not include observational study designs. Randomized controlled trials are the gold standard; however, conducting these trials is time-consuming and resource-intensive. Importantly, findings from randomized controlled trials take years to affect clinical practice, if they are implemented at all. Randomized controlled trials should not be expected to fill all the research gaps in the implementation and adoption of telehealth for chronic disease management. Thus, given the paucity of randomized controlled trials, we strongly recommend that future reviews focused on telehealth include what are likely to be rich and robust, but potentially biased; observational; and alternatively designed studies that emerge during and after the COVID-19 pandemic.

Future Directions

Overall, there are 5 key areas in which future studies on this topic can fill the existing gaps and improve the approach. First, and perhaps most critical, telehealth interventions should be thoroughly described to maximize reproducibility and generalizability in other clinical contexts. Guidance exists on mobile and web-based interventions, which may provide indirect suggestions about key characteristics for telehealth intervention description. Second, there is a need to evaluate how best to integrate telehealth as a replacement for in-person care. Furthermore, there is a need to assess which clinical settings are best suited to the telehealth environment (eg, primary care vs specialty care settings). Approaches to integrating telehealth can be expected to vary across settings with different workflow patterns, clinical resources, and competing clinical demands, which emphasizes the need for solid evidence. Third, outcomes varied across the included studies, and some important outcomes were not addressed by any study (eg, impact on clinical workflow, patient satisfaction with telehealth experience, and subsequent use). Fourth, investigators should be encouraged to consider a priori subgroup evaluations or make individual patient-level data available, so that future reviews can identify patient-level characteristics associated with better outcomes with telehealth. Finally, future studies should also actively solicit and report patient perspectives and feedback on telehealth interventions to better inform intervention design. Such information can guide clinics and health care systems to offer optimal patient-centered telehealth delivery and support efforts to ensure equitable benefits and access to telehealth.

Strengths and Limitations

Our review benefited from being protocol-driven, leveraging input from an expert panel consisting of clinicians and telehealth researchers, identifying disease-specific clinical outcomes, using an analytic framework to guide the understanding of telehealth modalities, and using a detailed approach to categorize and define telehealth components in chronic disease self-management. In addition, our review was based on a clear definition and use of telehealth. Notably, we acknowledge that individual patient characteristics (eg, race and ethnicity, gender, age, and rural status) may moderate the relationship between the modality in which the clinical visit occurs and any clinical-level and system-level outcomes.

Despite these strengths, our approach had some limitations. First, we included only the studies that met the EPOC criteria in this review; however, observational studies may have findings relevant to the provision of synchronous telehealth for chronic illness management. However, we do not believe that this limitation largely affected our findings. Second, we focused this review on 3 of the most prevalent chronic diseases, but there may be appropriately designed studies that targeted other conditions that we did not include. Third, we only included studies conducted in OECD countries, and thus, we may have missed relevant studies conducted in other countries. Fourth, given the small number of studies that we identified, statistical methods to detect publication bias were not conducted. Although it is possible that individual health systems or clinics have conducted quality improvement studies evaluating differences in experiences between synchronous and in-person care—especially during the COVID-19 pandemic—we suspect it to be unlikely that studies meeting EPOC criteria on this intervention have not been published, given the recent emphasis on the role of telehealth. Fifth, we identified few studies overall, and most studies had <100 patients and were assessed as having unclear or high ROB. Intervention core components, intervention fidelity, and impact of intervention on clinical workflow were not reported in any study. In addition, the interactions between clinicians and patients during telehealth episodes were not adequately or explicitly described, and most of our outcomes of interest were not consistently reported across the studies. These omissions limited the interpretation and replication of the evaluated interventions. Sixth, the included telehealth interventions used different telehealth modalities (email, phone, and video) with different hardware, delivered via different numbers of clinical interactions between patients and clinicians, over a wide range of intervention durations, and within different health care systems, which inherently make comparison between them challenging. Finally, the studies included in our review did not specify how they used or defined gender (man, woman, or nonbinary) or sex (male, female, or intersex) in their publications. Information on gender and sex is important to be captured and described for telehealth studies and research. Future studies should consider including observational studies; studies on additional, highly prevalent chronic diseases; studies conducted in non-OECD countries; and studies that do not meet the EPOC criteria especially, as those conducted since the onset of the COVID-19 pandemic may provide useful information.

Conclusions

The COVID-19 pandemic precipitated a rapid shift from in-person to telehealth delivery, without a clear understanding about the impacts of telehealth on important health outcomes. Previous studies have found that telehealth modalities can improve health outcomes through the supplementation of in-person management of certain chronic diseases, particularly with approaches such as remote monitoring and patient education. However, we found that, currently, there is very little evidence on the use of telehealth as a replacement for in-person care for several chronic conditions and that the studies in this area remain insufficient and methodologically inconsistent. In conclusion, our review builds on this existing body of literature by evaluating the comparative literature on the effectiveness of telehealth visits delivered as a substitute for in-person visits for chronic disease management and provides recommendations for future studies in this area.

Acknowledgments

The authors gratefully acknowledge their operational partner, Carolyn Turvey, PhD, for her telehealth expertise and Liz Wing and Sharon Thompson for editorial assistance. In addition, the authors would like to thank the following key partners and technical expert panel members for their feedback during the development and execution of this project: Bryan Batch, MD; Hayden Bosworth, PhD; Marla Clayman, PhD, Master of Public Health; Christopher Mosher, MD; and Scott Sherman, MD, Master of Public Health. This project was funded by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, and Quality Enhancement Research Initiative (grant ESP 09-010). This study also received support from the Durham Center of Innovation to Accelerate Discovery and Practice Transformation (grant CIN 13-410) at the Durham Veterans Affairs Health Care System, Veterans Affairs Health Services Research and Development (grant 18-234; AAL), and Veterans Affairs Office of Academic Affiliations Program (grant AF-3Q-05-2019-C; CW). The content is solely the responsibility of the authors and does not reflect the position or policy of the Duke University, US Department of Veterans Affairs, or US government. SR receives funding from the Health Resources and Services Administration Primary Care Training and Enhancement Program (TOBHP29992), which is not related to this study.

Conflicts of Interest

AAL reports receiving funding from Otsuka and the PhRMA Foundation.

Multimedia Appendix 1
Subject headings and key words used in the search for relevant literature.
[DOCX File , 26 KB - jmir_v24i8e37100_app1.docx]

Multimedia Appendix 2
Study characteristics.
[DOCX File , 28 KB - jmir_v24i8e37100_app2.docx]

Multimedia Appendix 3
Intervention characteristics.
[DOCX File , 28 KB - jmir_v24i8e37100_app3.docx]

Multimedia Appendix 4
All outcomes reported in the included studies.
[DOCX File , 47 KB - jmir_v24i8e37100_app4.docx]

Multimedia Appendix 5
Excluded studies and the reason for exclusion.
[DOCX File , 52 KB - jmir_v24i8e37100_app5.docx]

Multimedia Appendix 6
Ad hoc horizon scan to identify relevant studies in progress.
[DOCX File , 15 KB - jmir_v24i8e37100_app6.docx]

References

1. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
2. Heyworth L, Kirsh S, Zulman D, Ferguson JM, Kizer KW. Expanding access through virtual care: the VA's early experience with Covid-19. NEJM Catalyst. 2020 Jul 1. URL: https://catalyst.nejm.org/doi/full/10.1056/cat.20.0327 [accessed 2021-07-12]
3. Artandi M, Thomas S, Shah NR, Srinivasan M. Rapid system transformation to more than 75% primary care video visits within three weeks at Stanford: response to public safety crisis during a pandemic. NEJM Catalyst. 2020 Apr 21. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0100 [accessed 2021-07-12]
4. Baum A, Kaboli PJ, Schwartz MD. Reduced in-person and increased telehealth outpatient visits during the COVID-19 pandemic. Ann Intern Med 2021 Jan;174(1):129-131 [FREE Full text] [doi: 10.7326/M20-3026] [Medline: 32776780]
5. Mehrotra A, Ray K, Brockmeyer DM, Barnett ML, Bender JA. Rapidly Converting to "Virtual Practices": Outpatient Care in the Era of Covid-19. NEJM Catalyst. 2020 Apr 1. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0091 [accessed 2021-07-13]
6. Trump Administration Makes Sweeping Regulatory Changes to Help U.S. Healthcare System Address COVID-19 Patient Surge. Centers for Medicare & Medicaid Services. 2020 Mar 30. URL: https://www.cms.gov/newsroom/press-releases/trump-administration-makes-sweeping-regulatory-changes-help-us-healthcare-system-address-covid-19 [accessed 2021-07-12]
7. Litchfield I, Shukla D, Greenfield S. Impact of COVID-19 on the digital divide: a rapid review. BMJ Open 2021 Oct 12;11(10):e053440 [FREE Full text] [doi: 10.1136/bmjopen-2021-053440] [Medline: 34642200]
8. Eruchalu CN, Pichardo MS, Bharadwaj M, Rodriguez CB, Rodriguez JA, Bergmark RW, et al. The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York city. J Urban Health 2021 Apr;98(2):183-186 [FREE Full text] [doi: 10.1007/s11524-020-00508-9] [Medline: 33471281]
9. Mohammed HT, Hyseni L, Bui V, Gerritsen B, Fuller K, Sung J, et al. Exploring the use and challenges of implementing virtual visits during COVID-19 in primary care and lessons for sustained use. PLoS One 2021 Jun 24;16(6):e0253665 [FREE Full text] [doi: 10.1371/journal.pone.0253665] [Medline: 34166441]

10. Srinivasan M, Asch S, Vilendrer S, Thomas SC, Bajra R, Barman L, et al. Qualitative assessment of rapid system transformation to primary care video visits at an academic medical center. Ann Intern Med 2020 Oct 06;173(7):527-535 [FREE Full text] [doi: 10.7326/M20-1814] [Medline: 32628536]
11. Totten AM, Womack DM, Eden KB, McDonagh MS, Griffin JC, Grusing S, et al. Telehealth: Mapping the Evidence for Patient Outcomes From Systematic Reviews. Rockville, MD, USA: Agency for Healthcare Research and Quality (US); May 2016.
12. Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, Alberta Kidney Disease Network. Effect of telemedicine on glycated hemoglobin in diabetes: a systematic review and meta-analysis of randomized trials. CMAJ 2017 Mar 06;189(9):E341-E364 [FREE Full text] [doi: 10.1503/cmaj.150885] [Medline: 27799615]
13. Lin MH, Yuan WL, Huang TC, Zhang HF, Mai JT, Wang JF. Clinical effectiveness of telemedicine for chronic heart failure: a systematic review and meta-analysis. J Investig Med 2017 Jun;65(5):899-911. [doi: 10.1136/jim-2016-000199] [Medline: 28330835]
14. Lu AD, Gunzburger E, Glorioso TJ, Smith 2nd WB, Kenney RR, Whooley MA, et al. Impact of longitudinal virtual primary care on diabetes quality of care. J Gen Intern Med 2021 Sep;36(9):2585-2592 [FREE Full text] [doi: 10.1007/s11606-020-06547-x] [Medline: 33483815]
15. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 update: a report from the American Heart Association. Circulation 2020 Mar 03;141(9):e139-e596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]
16. National Diabetes Statistics Report 2020: Estimates of Diabetes and Its Burden in the United States. Centers for Disease Control and Prevention. Atlanta, GA, USA: Department of Health and Human Services, Centers for Disease Control and Prevention; 2020. URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2022-07-09]
17. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020 Aug;8(3):e000530 [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
18. Albritton J, Ortiz A, Wines R, Booth G, DiBello M, Brown S, et al. Video teleconferencing for disease prevention, diagnosis, and treatment : a rapid review. Ann Intern Med 2022 Feb;175(2):256-266. [doi: 10.7326/M21-3511] [Medline: 34871056]
19. Zachrison KS, Yan Z, Schwamm LH. Changes in virtual and in-person health care utilization in a large health system during the COVID-19 pandemic. JAMA Netw Open 2021 Oct 01;4(10):e2129973 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.29973] [Medline: 34705016]
20. Moran B, Frazier T, Brown LS, Case M, Polineni S, Roy L. A review of the effectiveness of audio-only telemedicine for chronic disease management. Telemed J E Health (forthcoming) 2022 Jan 12. [doi: 10.1089/tmj.2021.0285] [Medline: 35021882]
21. Breton M, Sullivan EE, Deville-Stoetzel N, McKinstry D, DePuccio M, Sriharan A, et al. Telehealth challenges during COVID-19 as reported by primary healthcare physicians in Quebec and Massachusetts. BMC Fam Pract 2021 Sep 26;22(1):192. [doi: 10.1186/s12875-021-01543-4] [Medline: 34563113]
22. Walsh C, Lewinski AA, Rushton S, Soliman D, Carlson SM, Luedke MW, et al. Virtual Care for the Longitudinal Management of Chronic Conditions: A Systematic Review. Health Services Research & Development. Washington, DC, USA: U.S. Department of Veterans Affairs; 2021. URL: http://vaww.hsrd.research.va.gov/publications/esp/virtual-care.cfm [accessed 2022-07-09]
23. Kilbourne AM, Elwy AR, Sales AE, Atkins D. Accelerating research impact in a learning health care system: VA's quality enhancement research initiative in the choice act era. Med Care 2017 Jul;55 Suppl 7 Suppl 1:S4-12 [FREE Full text] [doi: 10.1097/MLR.0000000000000683] [Medline: 27997456]
24. Gordon AM, Goldstein KM, Walsh C, Lewinski AA, Tabriz AS, Shaw RJ, et al. Virtual Care for the Longitudinal Management of Chronic Conditions. PROSPERO CRD42021239756. 2021. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021239756 [accessed 2022-07-05]
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. J Clin Epidemiol 2021 Jun;134:103-112. [doi: 10.1016/j.jclinepi.2021.02.003] [Medline: 33577987]
26. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Hoboken, NJ, USA: John Wiley and Sons; 2019.
27. 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]
28. EPOC resources for review authors. Cochrane Effective Practice and Organisation of Care. URL: https://epoc.cochrane.org/resources/epoc-resources-review-authors [accessed 2020-05-12]

29. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol 2017 Jul;87:4-13 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.05.006] [Medline: 28529184]
30. Klingeman H, Funnell M, Jhand A, Lathkar-Pradhan S, Hodish I. Type 2 diabetes specialty clinic model for the accountable care organization era. J Diabetes Complications 2017 Oct;31(10):1521-1526. [doi: 10.1016/j.jdiacomp.2017.05.011] [Medline: 28793967]
31. 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]
32. Jeong JY, Jeon JH, Bae KH, Choi YK, Park KG, Kim JG, et al. Smart care based on telemonitoring and telemedicine for type 2 diabetes care: multi-center randomized controlled trial. Telemed J E Health 2018 Aug;24(8):604-613. [doi: 10.1089/tmj.2017.0203] [Medline: 29341843]
33. Hansen C, Loges C, Seidl K, Eberhardt F, Tröster H, Petrov K, et al. INvestigation on Routine Follow-up in CONgestive HearT FAilure patients with remotely monitored implanted Cardioverter Defibrillators SysTems (InContact). BMC Cardiovasc Disord 2018 Jun 28;18(1):131 [FREE Full text] [doi: 10.1186/s12872-018-0864-7] [Medline: 29954340]
34. Rasmussen OW, Lauszus FF, Loekke M. Telemedicine compared with standard care in type 2 diabetes mellitus: a randomized trial in an outpatient clinic. J Telemed Telecare 2016 Sep;22(6):363-368. [doi: 10.1177/1357633X15608984] [Medline: 26468213]
35. Whitlock WL, Brown A, Moore K, Pavliscsak H, Dingbaum A, Lacefield D, et al. Telemedicine improved diabetic management. Mil Med 2000 Aug;165(8):579-584. [Medline: 10957848]
36. Rovner B. Reducing Emergency Diabetes Care for Older African Americans (PREVENT). Clinical Trials. 2018 Mar 15. URL: https://clinicaltrials.gov/show/NCT03466866 [accessed 2022-07-09]
37. Rodrigues D. Teleconsultation study in diabetic patients. World Health Organization International Clinical Trials Registry Platform. 2019 Dec 27. URL: https://trialsearch.who.int/Trial2.aspx?TrialID=RBR-8gpgyd [accessed 2022-07-09]
38. Komkov DS, Myasnikov RP, Fedotova NP, Boytsov SA. Effect of education and telephone monitoring on exercise tolerance and health-related quality of life in patients with chronic heart failure: a prospective randomized controlled parallel-group study. Eur J Prevent Cardiol 2015 Jun 14;22(1_suppl):S194. [doi: 10.1177/2047487315586749]
39. Mehrotra A, Chernew ME, Linetsky D, Hatch H, Cutler DA. The Impact of the COVID-19 Pandemic on Outpatient Visits: Practices Are Adapting to the New Normal. The Commonwealth Fund. 2020 Jun 25. URL: https://www.commonwealthfund.org/publications/2020/jun/impact-covid-19-pandemic-outpatient-visits-practices-adapting-new-normal [accessed 2021-07-12]
40. 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]
41. Yun JE, Park JE, Park HY, Lee HY, Park DA. Comparative effectiveness of telemonitoring versus usual care for heart failure: a systematic review and meta-analysis. J Card Fail 2018 Jan;24(1):19-28. [doi: 10.1016/j.cardfail.2017.09.006] [Medline: 28939459]
42. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote monitoring of patients with heart failure: an overview of systematic reviews. J Med Internet Res 2017 Jan 20;19(1):e18 [FREE Full text] [doi: 10.2196/jmir.6571] [Medline: 28108430]
43. Cajita MI, Gleason KT, Han HR. A systematic review of mHealth-based heart failure interventions. J Cardiovasc Nurs 2016;31(3):E10-E22 [FREE Full text] [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]
44. Faleh AlMutairi M, Tourkmani AM, Alrasheedy AA, ALHarbi TJ, Bin Rsheed AM, ALjehani M, et al. Cost-effectiveness of telemedicine care for patients with uncontrolled type 2 diabetes mellitus during the COVID-19 pandemic in Saudi Arabia. Ther Adv Chronic Dis 2021 Sep 8;12:20406223211042542 [FREE Full text] [doi: 10.1177/20406223211042542] [Medline: 34729144]
45. Tourkmani AM, ALHarbi TJ, Rsheed AM, Alrasheedy AA, ALMadani W, ALJuraisi F, et al. The impact of telemedicine on patients with uncontrolled type 2 diabetes mellitus during the COVID-19 pandemic in Saudi Arabia: findings and implications. J Telemed Telecare (forthcoming) 2021 Feb 01:1357633X20985763. [doi: 10.1177/1357633X20985763] [Medline: 33525952]
46. Hu Y, Wen X, Wang F, Yang D, Liu S, Li P, et al. Effect of telemedicine intervention on hypoglycaemia in diabetes patients: a systematic review and meta-analysis of randomised controlled trials. J Telemed Telecare 2019 Aug;25(7):402-413. [doi: 10.1177/1357633X18776823] [Medline: 29909748]
47. Lee SW, Chan CK, Chua SS, Chaiyakunapruk N. Comparative effectiveness of telemedicine strategies on type 2 diabetes management: a systematic review and network meta-analysis. Sci Rep 2017 Oct 04;7(1):12680 [FREE Full text] [doi: 10.1038/s41598-017-12987-z] [Medline: 28978949]
48. Chakranon P, Lai YK, Tang YW, Choudhary P, Khunti K, Lee SW. Distal technology interventions in people with diabetes: an umbrella review of multiple health outcomes. Diabet Med 2020 Dec;37(12):1966-1976. [doi: 10.1111/dme.14156] [Medline: 31631398]

49. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-930. [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]
50. Lewinski AA, Sullivan C, Allen KD, Crowley MJ, Gierisch JM, Goldstein KM, et al. Accelerating implementation of virtual care in an integrated health care system: future research and operations priorities. J Gen Intern Med 2021 Aug;36(8):2434-2442. [doi: 10.1007/s11606-020-06517-3] [Medline: 33496928]

Abbreviations

ALT: alanine aminotransferase
CHF: congestive heart failure
COPD: chronic obstructive pulmonary disease
EPOC: Effective Practice and Organization of Care
HbA$_{1c}$: glycosylated hemoglobin
OECD: Organisation for Economic Co-operation and Development
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
ROB: risk of bias
T2DM: type 2 diabetes mellitus
VHA: Veterans Health Administration

Edited by T Leung; submitted 12.02.22; peer-reviewed by V Stamenova, B Chaudhry; comments to author 26.05.22; revised version received 05.07.22; accepted 08.07.22; published 26.08.22.

Please cite as:
Lewinski AA, Walsh C, Rushton S, Soliman D, Carlson SM, Luedke MW, Halpern DJ, Crowley MJ, Shaw RJ, Sharpe JA, Alexopoulos AS, Tabriz AA, Dietch JR, Uthappa DM, Hwang S, Ball Ricks KA, Cantrell S, Kosinski AS, Ear B, Gordon AM, Gierisch JM, Williams Jr JW, Goldstein KM
Telehealth for the Longitudinal Management of Chronic Conditions: Systematic Review
J Med Internet Res 2022;24(8):e37100
URL: https://www.jmir.org/2022/8/e37100
doi:10.2196/37100
PMID:36018711

Review

Association Between Patient Factors and the Effectiveness of Wearable Trackers at Increasing the Number of Steps per Day Among Adults With Cardiometabolic Conditions: Meta-analysis of Individual Patient Data From Randomized Controlled Trials

Alexander Hodkinson[1,2], BSc, MSc, PhD; Evangelos Kontopantelis[1,2,3], PhD; Salwa S Zghebi[1,2], PhD; Christos Grigoroglou[1,2], PhD; Brian McMillan[1,2], MD; Harm van Marwijk[4], MD; Peter Bower[1,2], PhD; Dialechti Tsimpida[1,2], PhD; Charles F Emery[5], PhD; Mark R Burge[6], MD; Hunter Esmiol[6], PhD; Margaret E Cupples[7], MD; Mark A Tully[8], PhD; Kaberi Dasgupta[9,10], MD; Stella S Daskalopoulou[9,11], MD; Alexandra B Cooke[9], PhD; Ayorinde F Fayehun[12], PhD; Julie Houle[13], MD; Paul Poirier[14], PhD; Thomas Yates[15], MD; Joseph Henson[15], MD; Derek R Anderson[5], MD; Elisabeth B Grey[16], PhD; Maria Panagioti[1,2], PhD

[1]Division of Population Health, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom
[2]Health Services Research and Primary Care, National Institute for Health Research School for Primary Care Research, Manchester, United Kingdom
[3]Division of Informatics, Imaging & Data Sciences, Faculty of Biology, Medicine & Health, University of Manchester, Manchester, United Kingdom
[4]Department of Primary Care and Public Health, Brighton and Sussex Medical School, University of Brighton, Brighton, United Kingdom
[5]Department of Psychology, The Ohio State University College of Arts and Sciences, Columbus, OH, United States
[6]Department of Medicine, Endocrinology and Metabolism, University of New Mexico Health Sciences Center, Albuquerque, NM, United States
[7]Department of General Practice and Primary Care, Centre for Public Heath, Queen's University Belfast, Belfast, United Kingdom
[8]School of Medicine, Ulster University, Londonderry, United Kingdom
[9]Department of Medicine, McGill University, Montreal, QC, Canada
[10]Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Montreal, QC, Canada
[11]Centre for Translational Biology, Research Institute of the McGill University Health Centre, Montreal, QC, Canada
[12]Department of Family Medicine, University College Hospital, Ibadan, Nigeria
[13]Department of Nursing, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada
[14]Institut Universitaire de Cardiologie et de Pneumologie de Québec, Université Laval, Laval, QC, Canada
[15]Diabetes Research Centre, University of Leicester, Leicester, United Kingdom
[16]Centre for Motivation and Health Behaviour Change, Department for Health, University of Bath, Bath, United Kingdom

Corresponding Author:
Alexander Hodkinson, BSc, MSc, PhD
Division of Population Health
School of Health Sciences, Faculty of Biology, Medicine and Health
University of Manchester
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 (0)161 2753535
Email: alexander.hodkinson@manchester.ac.uk

Abstract

Background: Current evidence supports the use of wearable trackers by people with cardiometabolic conditions. However, as the health benefits are small and confounded by heterogeneity, there remains uncertainty as to which patient groups are most helped by wearable trackers.

Objective: This study examined the effects of wearable trackers in patients with cardiometabolic conditions to identify subgroups of patients who most benefited and to understand interventional differences.

Methods: We obtained individual participant data from randomized controlled trials of wearable trackers that were conducted before December 2020 and measured steps per day as the primary outcome in participants with cardiometabolic conditions including diabetes, overweight or obesity, and cardiovascular disease. We used statistical models to account for clustering of participants within trials and heterogeneity across trials to estimate mean differences with the 95% CI.

Results: Individual participant data were obtained from 9 of 25 eligible randomized controlled trials, which included 1481 of 3178 (47%) total participants. The wearable trackers revealed that over the median duration of 12 weeks, steps per day increased by 1656 (95% CI 918-2395), a significant change. Greater increases in steps per day from interventions using wearable trackers were observed in men (interaction coefficient –668, 95% CI –1157 to –180), patients in age categories over 50 years (50-59 years: interaction coefficient 1175, 95% CI 377-1973; 60-69 years: interaction coefficient 981, 95% CI 222-1740; 70-90 years: interaction coefficient 1060, 95% CI 200-1920), White patients (interaction coefficient 995, 95% CI 360-1631), and patients with fewer comorbidities (interaction coefficient –517, 95% CI –1188 to –11) compared to women, those aged below 50, non-White patients, and patients with multimorbidity. In terms of interventional differences, only face-to-face delivery of the tracker impacted the effectiveness of the interventions by increasing steps per day.

Conclusions: In patients with cardiometabolic conditions, interventions using wearable trackers to improve steps per day mostly benefited older White men without multimorbidity.

Trial Registration: PROSPERO CRD42019143012; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=143012

(J Med Internet Res 2022;24(8):e36337) doi:10.2196/36337

KEYWORDS

systematic review; individual patient data; meta-analysis; steps/day; wearable tracker; cardiometabolic conditions; diabetes; obesity; cardiovascular disease

Introduction

Background

Cardiometabolic conditions are the leading cause of death worldwide, accounting for more than 41 million deaths annually [1]. These conditions include obesity, diabetes mellitus, and cardiovascular disease (CVD); these 3 common, intersecting noncommunicable diseases affect almost two-thirds of the global population [2,3].

The World Health Organization recently recognized physical inactivity as a serious and growing public health problem and has set out to reduce it by 10% before 2025 [4]. The United Kingdom's National Institute for Health and Care Excellence also highlights the importance of physical activity for obesity management, successful aging, CVD prevention, and weight management during pregnancy [5]. The consequences of being physically inactive include unhealthy weight gain, dyslipidemia, and elevated blood pressure and blood glucose levels, all of which heighten the risk of developing a cardiometabolic condition [6].

Wearable physical activity trackers, such as accelerometers, pedometers, and the Fitbit (Fitbit Inc), are portable electrical or electromechanical devices that count each step a person takes by detecting the motion of the person along the body's long axis [7,8]. They have become very popular for motivating and monitoring (thereby increasing) physical activity in general and in people with cardiometabolic conditions in particular [9,10]. Systematic reviews have suggested that the validity of various wearable trackers, especially those measuring steps, is high, and these reviews have found that they are useful for tracking ambulatory physical activity in clinical populations [11-13].

Since many wearable trackers are affordable and user-friendly, they are viewed as a good practical method for monitoring basic physical activity [14], such as the number of steps per day, in high-risk people with chronic cardiometabolic conditions [15-17]. However, their long-term effectiveness in achieving the desired behavior changes (ie, increasing steps per day) in specific patient subgroups with cardiometabolic conditions is unclear, and they may only succeed in the short term as "quick fixes" [18].

Our recent meta-analysis of 38 randomized trials [19] suggested that interventions using wearable trackers are moderately effective at increasing physical activity, including steps per day, in people with cardiometabolic conditions. The most promising interventions were those that focused on the number of steps per day. However, without individual participant data (IPD), we could not conduct an assessment of patient factors, baseline effects, and interventional differences, nor could we analyze their importance [20]. This was a major constraint of the review findings.

In IPD meta-analysis, rather than extracting summary (ie, aggregate) data from study publications or from investigators, original research data are obtained directly from the researchers responsible for each study. These data are then re-analyzed centrally and combined in the meta-analysis. The IPD approach is becoming an increasingly popular tool compared to traditional aggregate-data meta-analysis, especially as the IPD approach avoids reliance on published results and provides an opportunity to investigate individual-level interactions, such as treatment-effect modifiers [20].

Objectives

In the present study, we undertook an IPD meta-analysis to identify whether belonging to certain subgroups of patients with cardiometabolic conditions, including groups with differing age, sex, ethnicity, and number or combination of cardiometabolic conditions, moderated the effectiveness of interventions using wearable trackers in improving physical activity, measured by the number of steps per day. We also examined the impact of interventional differences, such as behavior change, device placement, delivery method, and performance over time, on the effectiveness of interventions using wearable trackers in improving steps per day.

Methods

This IPD meta-analysis followed a registered (PROSPERO CRD42019143012) protocol. A statistical analysis plan was produced in advance of analysis and the findings are reported in accordance with the PRISMA-IPD statement [21].

Literature Search and Study Identification

We searched MEDLINE (Ovid), EMBASE (EBSCO), CENTRAL, CINAHL, and PsycINFO from inception until August 2018, without language restriction; this was updated in December 2020 (Multimedia Appendix 1, pages 3-11). Additional studies were obtained by citation tracking, extraction from previous systematic reviews, and searches of trial registers (ie, ClinicalTrials.gov and ICTRP). A list of all the search sources and the data collection and management process are detailed in the protocol for this paper.

Two researchers (AH and MP) independently identified the citations and then fully screened the relevant manuscripts according to the eligibility criteria. We included randomized controlled trials (RCTs) or cluster RCTs involving adults (aged 16 years or older) with a cardiometabolic condition, defined as a diagnosis (or high risk) of type 2 diabetes, CVD, or obesity or overweight.

We included studies with an intervention program designed around the daily usage of wearable trackers, such as pedometers, accelerometers, and fitness trackers, rather than studies that only measured performance at the beginning and end of the study. Studies were required to have a usual-care comparator. Participants' steps per day in both intervention and usual-care groups were measured in parallel using the same device. We further required that studies reported a physical activity assessment (ie, step count) at baseline and follow up using a separate wearable tracker that all participants received independently of their allocation (ie, both the intervention and control groups). We adopted this eligibility criterion to be able to reliably pool the results across the studies.

The primary outcome was the number of steps per day, measured with any wearable tracker. Wearable trackers, either mechanical (eg, spring levered) or electronic (eg, using GPS or actigraph functionality), mostly monitored daily steps as the main outcome of interest; intervention programs involving wearable trackers often set goals for increasing the number of daily steps incrementally over time and estimated and logged the total distance travelled [22]. We excluded wearable trackers that used other types of measurement output, because these vary considerably in terms of their performance, choice of activity measurement (ie, light, moderate, or vigorous activity, energy expenditure, sedentary time, or stationary pattern) and intensity and frequency of the physical activity (ie, bouts of exercise). This would have made their pooling more problematic.

Secondary outcomes included anthropometric measures, glycated hemoglobin level (mmols/mol), blood pressure, and cholesterol level.

Risk of bias (RoB) was independently assessed by 2 reviewers (AH and MP) using the Cochrane RoB tool [23], alongside the completeness and quality of the provided IPD. Results of IPD studies were also compared with studies that did not supply IPD.

Data Extraction and Assessment of IPD Integrity

We used IPD (obtained from November 2019 to October 2021) to determine demographic characteristics that we chose a priori in the protocol, such as age, sex, ethnicity, and comorbidity; intervention characteristics, such as objective, duration, use of a behavior change framework, delivery method and placement of wearable tracker; and primary and secondary outcomes.

Continuous variables were kept continuous, but some were also categorized when this was considered to be more clinically meaningful. For instance, patient groups were split by median age and by age range (20-49, 50-59, 60-69, and 70-90 years); the total number of comorbidities, predefined as cardiometabolic conditions including type 2 diabetes, hypertension, obesity or overweight, metabolic syndrome, and any cardiovascular condition; and BMI (normal 18.5-24.9 kg/m^2; overweight, 25-29.9 kg/m^2; obese, ≥30 kg/m^2). Descriptive characteristics (eg, age, comorbidity, education, and ethnicity) were analyzed for intervention and control groups using ANOVA for continuous variables and the chi-square test for categorical variables. Following this, ethnicity was classified into 2 groups, White European/North American and other ethnicities, to improve the analysis of covariance (ANCOVA), due to the limited number of patients in other ethnic groups.

Since all the trials provided above 70% of the IPD for the corresponding primary outcome, we imputed any missing values using the R package MICE: Multivariate Imputation by Chained Equations [24], following Rubin's principle for imputation [25]. The range of imputed values was bounded by the observed values of the primary outcome, and baseline covariates (study, intervention, age, sex, and baseline) were used to predict missing data. The algorithms' convergence was assessed, and sensitivity analyses were performed using only cases with available data (ie, complete cases).

Data Synthesis

Primary analyses used a 1-stage linear mixed effect model that incorporated random effects to allow for heterogeneity across trials [20,26,27], fitted using the Stata (version 16.1; StataCorp LLC) commands mixed and ipdforest to summarize the evidence by study and obtain forest plots [28]. Restricted maximum likelihood was used for model estimation, and centering of covariates by study-specific means was performed to avoid aggregation bias [29].

As the primary outcomes were all objectively measured using the standard unit of steps per day, the analysis was performed using the mean difference (MD) between intervention and control groups with the 95% CI. Differential effects were then investigated by adding patient covariate parameters (ie, age, sex, ethnicity, and number of cardiometabolic conditions) and interactions between covariates (ie, treatment-covariate interaction terms) to the linear mixed model for the primary outcome. Important and significant differential effects were displayed through subgroup analysis plots.

Results from a 2-stage random-effects (DerSimonian-Laird) meta-analysis were compared to results from 1-stage analyses using the ipdmetan command, for consistency. The Hartung-Knapp CI was used to account for uncertainty in the variance estimate [30].

Heterogeneity was examined by visually inspecting forest plots and using the I^2 statistic with the 95% CI [31]. Publication bias was assessed visually by using contour-enhanced funnel plots and the statistical Egger test for asymmetry [32]. The Egger test is performed using the following hypothesis, testing by *P* value: the null hypothesis is symmetry in the funnel plot, and the alternative hypothesis is asymmetry in the funnel plot. If $P \leq .05$, we reject the null hypothesis. If $P > .05$, we accept the null hypothesis and reject the alternative hypothesis [33]. We assessed for availability bias by comparing summary results of the non-IPD studies with those from IPD studies [34].

Sensitivity analyses of study-level factors were performed by (1) comparing studies that used a social cognitive theory framework as a guide for behavior change, (2) comparing studies that used a goal or goals as part of the intervention, (3) comparing the placement of the wearable tracker (ie, on the waist or wrist), (4) comparing the delivery of the intervention (ie, face to face or self-managed), (5) removing studies with high or unclear RoB based on allocation concealment, and (6) assessing the long-term performance (ie, beyond 6 months).

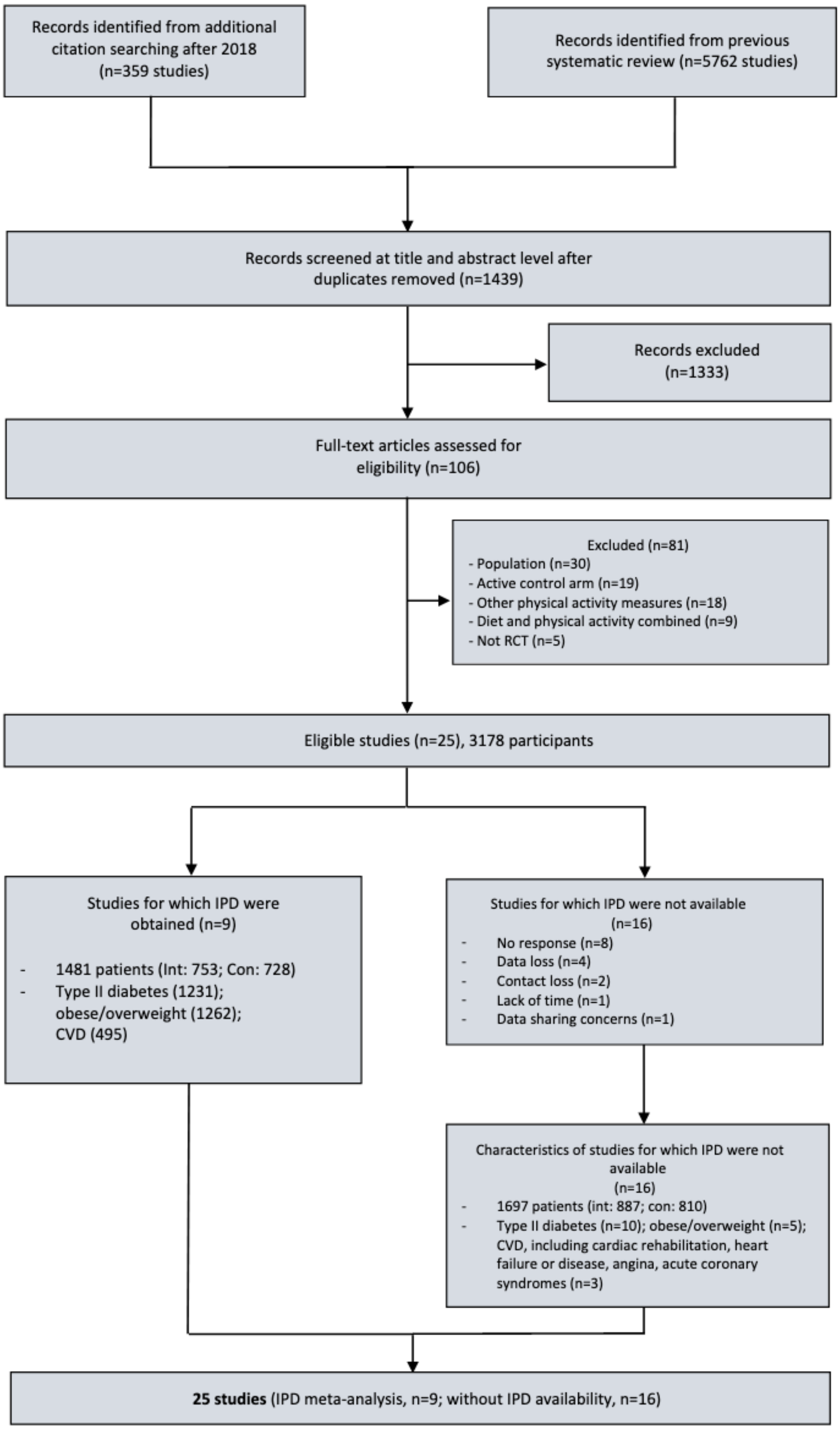
Ethical Considerations

Ethical approval and patient consent was not required as it had already been obtain from the primary authors during their trial period. Ethical waiver was provided and acknowledged by our University international review board.

Results

Of 25 eligible RCTs (including 3178 participants) that used wearable trackers and measured steps per day, we found that 9 studies (36%) had a median intervention duration of 12 (range 7-52) weeks, providing IPD for 47% of the total participants (1481/3178) (Figure 1). A list of the eligible studies is provided in Multimedia Appendix 1 (pages 12-14).

Figure 1. Identification and selection of studies providing individual participant data for meta-analysis of interventions involving wearable trackers for measuring steps per day in patients with cardiometabolic conditions. Con: control group; CVD: cardiovascular disease; int: intervention group; IPD: individual participant data; RCT: randomized controlled trial.



Characteristics of Studies

Of the 9 included IPD studies, 5 were from North America [35-40], 3 were from the United Kingdom [41-43], and 1 was from Nigeria [44]. Wearable trackers were used in all 9 studies [35,37-44] and the primary outcome was steps per day. Within the IPD sample, 822 of 1481 patients were men (56%) and 907 of 1481 were 60 years or older (61%; range 25-86 years). A total of 1231 of 1481 patients had (or were at risk of) type 2 diabetes (83%), 1262 of 1481 were obese or overweight (85%), and 495 of 1481 patients had CVD (33%). Characteristics of the studies that provided IPD and did not provide IPD are presented in Multimedia Appendix 1 (pages 15-22); baseline characteristics of the IPD are provided in Table 1.

Table 1. Baseline characteristics of the individual participant data and imbalance assessment between treatment arms. Percentages are proportions of observations to intervention or control arms, as applicable.

Characteristics	Intervention	Control	*P* value[a]
Steps per day (in 1481 patients in 9 studies) (n), mean (SD)	6071.25 (3060.72)	6072.11 (3064.40)	.99
Age (in 1481 patients in 9 studies) (years), mean (SD)	60.53 (9.70)	60.73 (10.06)	.68
Height (in 986 patients in 5 studies) (cm), mean (SD)	124.00 (74.12)	126.73 (72.35)	.56
BMI (in 1325 patients in 7 studies) (kg/cm^2), mean (SD)	32.03 (5.44)	32.11 (5.03)	.77
BMI by classification (in 1325 patients in 7 studies)			.40
Patients, N	680	645	
Normal (18.5-24.9 kg/m^2), n (%)	37 (5.4)	26 (4)	
Overweight (25-29.9 kg/m^2), n (%)	266 (39.1)	267 (41.4)	
Obese (≥30 kg/m^2), n (%)	377 (55.4)	352 (54.6)	
Ethnicity (in 1414 patients in 9 studies)			.27
Patients, N	721	693	
White European or North American, n (%)	534 (74.1)	510 (73.6)	
African American, n (%)	105 (14.6)	116 (16.7)	
Hispanic or Latino, n (%)	40 (5.5)	23 (3.3)	
Mixed ethnicity, n (%)	27 (3.7)	29 (4.2)	
Asian/Middle Eastern, n (%)	15 (2.1)	15 (2.2)	
Education status[a] (in 407 patients in 2 studies)			.60
Patients, N	210	183	
Low (not completed secondary education to A level), n (%)	42 (20)	33 (18)	
Medium (completed secondary education; ie, A level equivalent), n (%)	36 (17.1)	36 (19.7)	
High (any further or higher education), n (%)	94 (45.8)	71 (38.8)	
Preexisting CVD (in 495 patients in 3 studies), n/N (%)	44/230 (19.1)	84/265 (32)	.07
Preexisting type 2 diabetes (in 1231 patients in 4 studies), n/N (%)	370/619 (59.8)	403/612 (65.8)	.21
Preexisting hypertension (in 642 patients in 3 studies), n/N (%)	250/311 (80.4)	264/331 (79.8)	.83
Preexisting metabolic syndrome (in 471 patients in 2 studies), n/N (%)	194/240 (80.8)	201/231 (87)	.73
Depression score (in 347 patients in 2 studies), mean (SD)	2.24 (4.13)	2.29 (4.10)	.99
Smokers (in 578 patients in 3 studies), n/N (%)	87/282 (30.9)	84/296 (28.4)	.48

[a] Mean values were compared with a 2-tailed *t* test and categorical covariates were compared with the chi-squared test or ANOVA.

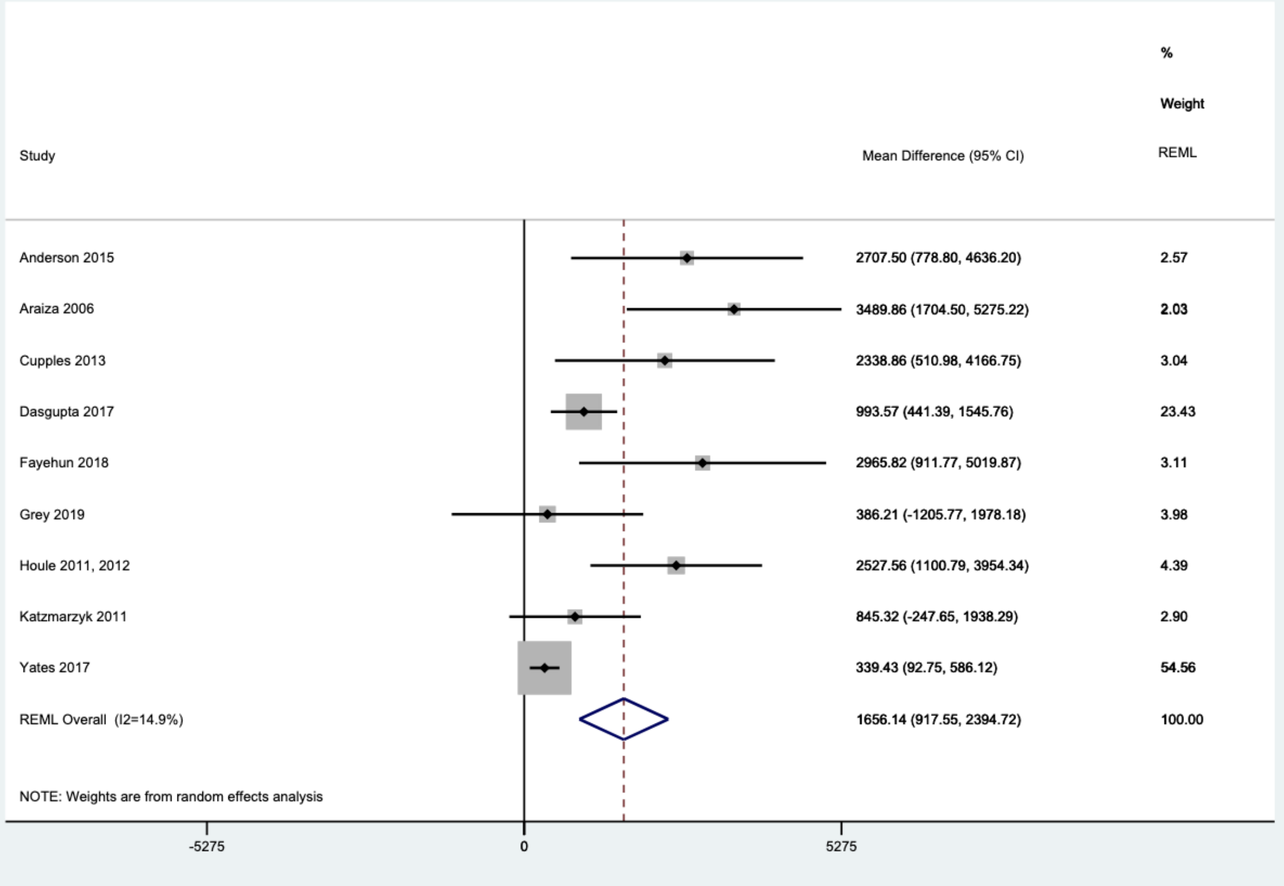
Risk of Bias Assessments

The RoB assessment of studies that contributed IPD compared with those that did not showed that the former had lower RoB across all domains (Multimedia Appendix 1, page 22). The assessments for each of the RoB domains of the IPD studies are available in Multimedia Appendix 1 (page 23), and the results of RoB assessments for non-IPD studies are available in our earlier systematic review [19].

Efficacy of Wearable Trackers on Increasing Steps per Day

In the 1-stage analysis involving all 9 studies and 1481 participants, wearable trackers were associated with small but significantly improved levels of physical activity over the median intervention duration of 12 (range 6-52) weeks. The number of steps was 1656 (95% CI 918-2395) greater in the intervention than the control group (Figure 2). These results are consistent with the 2-stage analysis (Multimedia Appendix 1, page 32). The mean number of steps per day at the end of treatment, without adjusting for baseline scores, was 6561 (SD 3336) in the wearable activity tracker intervention group and 6561 (SD 3340) in the control group; this was not a significant difference (*P*=.99). Visual and statistical evidence (*P*=.002; Egger test) of a small study effect was found in the funnel plots (Multimedia Appendix 1, pages 24-30). After removing the studies with high or unclear RoB from the analysis, the symmetry of the funnel plot improved somewhat (*P*=.06). Secondary outcomes showed insignificant differences (Multimedia Appendix 1, page 31).

Figure 2. Forest plot showing 1-stage meta-analysis of individual participant data from studies using wearable trackers to measure steps per day; the mean postintervention difference in steps per day is also shown. MD: mean difference; REML: restricted maximum likelihood.



Covariate interaction effects for the primary outcome showed that patients older than 50 years benefited more from using wearable trackers. The interaction coefficient for patients aged 50 to 59 years was 1175.16 (95% CI 377.46 to 1972.86) steps per day, with an individual group effect of 2006.83 (95% CI 1163.83 to 2849.82); in patients aged 60 to 69 years, the interaction coefficient was 981.37 (95% CI 222.39 to 1740.35), with an individual group effect of 1813.04 (95% CI 986.40 to 2639.68); and in patients aged 70 to 90 years, the interaction coefficient was 1059.98 (95% CI 200.29 to 1919.66), with an individual group effect of 1891.65 (95% CI 963.98 to 2819.31; I^2 15.5%). In contrast, for patients aged under 50 years the interaction coefficient was 831.67 (95% CI –97.00 to 1760.33) (Table 2).

The number of steps per day after using wearable trackers was greater among men than women (interaction coefficient –668.3, 95% CI –1156.8 to –179.93). For men, the mean number of steps per day was 2006 (95% CI 1204.4 to 2807.6), while for women, it was 1337.65 (95% CI 538.92 to 2136.37; I^2 16%) (Figure 3 and Figure 4). Patients with at least 2 comorbidities showed a significantly lower number of steps per day; the interaction coefficient for 2 comorbidities was –516.80 (95% CI –1188.34 to –10.74), and the mean number of steps was 1344.70 (95% CI 421.62 to 1843.87). For patients with 3 comorbidities, the interaction coefficient was –876.44 (95% CI –2071.88 to 509.41), and the mean number of steps was 570.17 (95% CI –304.66 to 870.08; I^2 15.7%). In contrast, for patients with only 1 comorbidity, the interaction coefficient was 1861.55 (95% CI 1061.6 to 2661.5). White patients also displayed a higher step count after using wearable trackers (2189, 95% CI 1276 to 3102) compared to other ethnic groups (1194, 95% CI 280 to 2107); the interaction coefficient was 995 (95% CI 360 to 1631) steps per day (I^2 21%).

Table 2. Differential effects of wearable trackers on physical activity measured by steps per day among specific subgroups of patients.

Characteristic	Mean difference in steps per day,[a] n (95% CI)	Treatment covariate interaction		
		Interaction coefficient (95% CI)	*P* value	I^2, % (95% CI)
Age				
≥60 years	1814.39 (996.51 to 2632.28)	1	N/A[b]	16.1 (5.0 to 41.1)
<60 years[c]	1566.83 (766.98 to 2366.68)	–247.56 (–762.0 to 266.92)	.35	
Age category				
20-49 years	831.67 (–97.00 to 1760.33)	1	N/A	15.5 (4.8 to 40.2)
50-59 years	2006.83 (1163.83 to 2849.82)	1175.16 (377.46 to 1972.86)	.004	
60-69 years	1813.04 (986.40 to 2639.68)	981.37 (222.39 to 1740.35)	.01	
70-90 years	1891.65 (963.98 to 2819.31)	1059.98 (200.29 to 1919.66)	.02	
Sex				
Men	2006 (1204.4 to 2807.6)	1	N/A	16.03 (5.0 to 40.91)
Women	1337.65 (538.92 to 2136.37)	–668.3 (–1156.8 to –179.93)	.01	
Ethnicity				
Other	1193.65 (280.31 to 2106.99)	1	N/A	20.5 (7.0 to 46.8)
White[d]	2189.0 (1276.3 to 3101.65)	995.30 (359.80 to 1630.80)	.002	
Comorbidities[e]				
1	1861.55 (1061.6 to 2661.5)	1	N/A	15.7 (4.6 to 41.7)
2	1344.70 (421.62 to 1843.87)	–516.80 (–1188.34 to –10.74)	.04	
3	570.17 (–304.66 to 870.08)	–876.44 (–2071.88 to –509.41)	.01	
4	1078.28 (468.72 to 2077.31)	N/A	N/A	
Cardiometabolic condition focus				
Other conditions[f]	1535.28 (–557.35 to 3627.91)	1	N/A	16.2 (0.5 to 87.3)
Type II diabetes	1942.47 (47.24 to 3837.70)	407.19 (–785.09 to 1599.47)	.50	

[a] Model accounted for baseline steps per day scores with analysis of covariance.

[b] N/A: not applicable.

[c] Per year of age.

[d] White versus all other ethnicities.

[e] Including type II diabetes, hypertension, angina, obese or overweight, and any other cardiovascular condition (excluding stroke).

[f] Including hypertension.

Figure 3. Gender effect for women. Forest plot showing 1-stage meta-analysis of individual participant data from women only, derived from studies using wearable trackers to measure steps per day; the mean postintervention difference in steps per day is also shown. MD: mean difference; REML: restricted maximum likelihood.

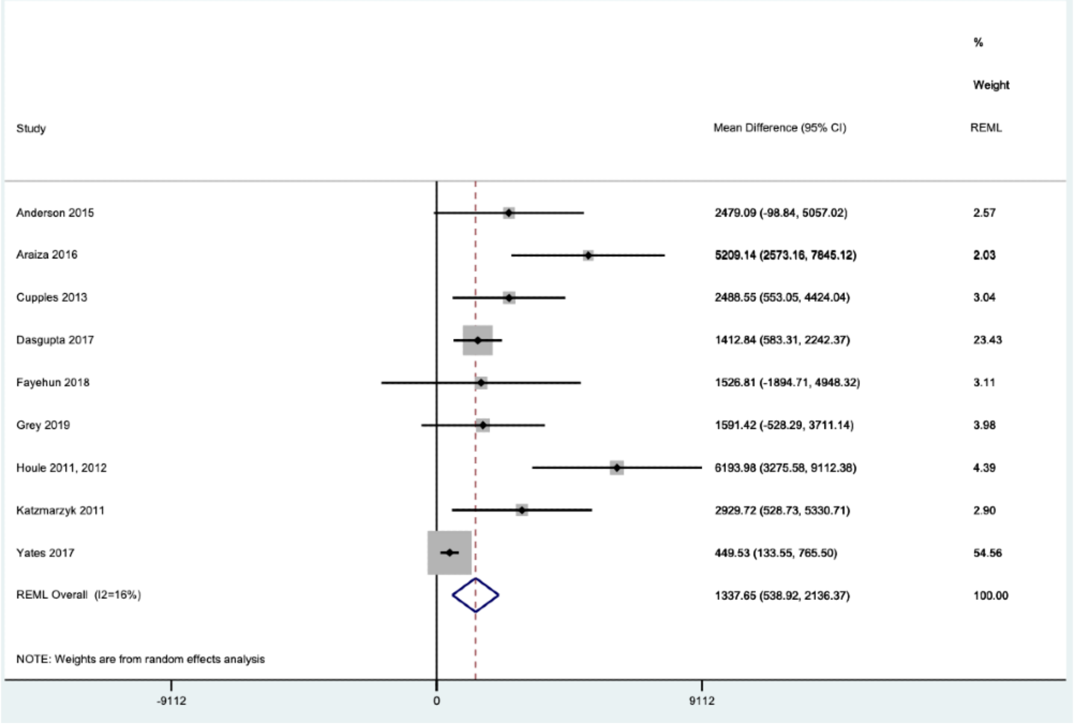
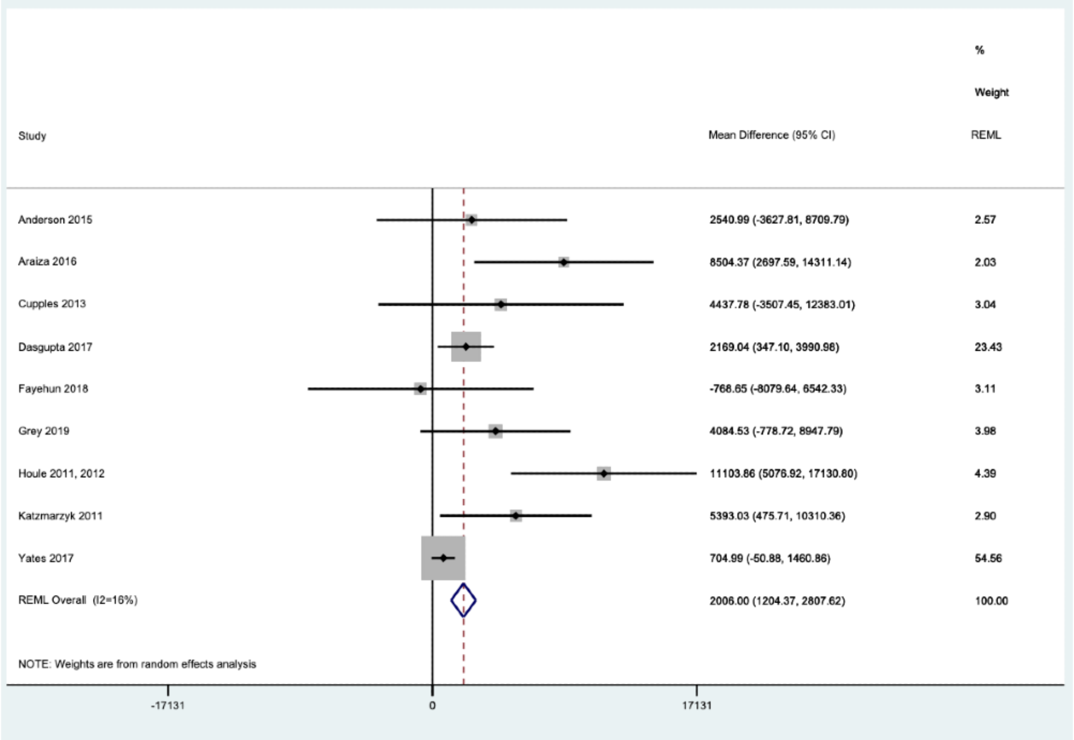


Figure 4. Gender effect for men. Forest plot showing 1-stage meta-analysis of individual participant data from men only, derived from studies using wearable trackers to measure steps per day; the mean postintervention difference in steps per day is also shown. MD: mean difference; REML: restricted maximum likelihood.



Further Sensitivity Analysis and Exploratory Analysis

Behavior change frameworks were used in 3 studies [35,37,38,45], but did not appear to improve the number of steps per day (the interaction coefficient was –1054, 95% CI –2785 to 677). Nevertheless, groups that did or did not use the program still showed a statistically significant change (part of program: 2476, 95% CI 935 to 4017; not part of program: 1422, 95% CI 634 to 2211). In addition, studies that set goals for reaching specific steps per day also did not appear to improve the number of steps per day, nor did the placement of the wearable tracker (ie, waist vs wrist) (Multimedia Appendix 1, page 33).

Small but nonsignificant improvements in steps per day were seen in studies with low RoB compared to studies with high or unclear RoB. Studies that assessed the number of steps per day beyond 26 weeks showed a lower (but nonsignificant) number of steps per day compared to studies that assessed performance in the short term (ie, less than 6 months). While both groups showed statistically significant changes (<26 weeks: 2000, 95% CI 1068 to 2932; ≥26 weeks: 1143, 95% CI 33 to 2254) when they used trackers over the longer term (ie, beyond 6 months) it was clear that performance waned by almost 1000 steps per day compared to short-term use (ie, up to 6 months). Studies that involved face-to-face delivery of the wearable tracker significantly improved the number of steps per day (2630, 95% CI 1835 to 3425) when compared to studies that involved only self-regulated use of the tracker (850, 95% CI 325 to 1375) (the interaction coefficient was 1780, 95% CI 826 to 2733).

Improvement in steps per day among the non-IPD studies was significantly higher with the wearable trackers; results showed almost double the number of steps per day (2854, 95% CI 1944 to 3763) compared to the IPD meta-analysis estimate (Multimedia Appendix 1, page 34). However, including the non-IPD studies in the meta-analysis increased bias from small studies and worsened the funnel plot asymmetry (Multimedia Appendix 1, pages 26-27).

Discussion

Principal Findings

This IPD meta-analysis confirms that interventions using wearable trackers were effective at increasing physical activity, as measured by total number of steps per day, in participants with cardiometabolic conditions, compared to control groups, over a median duration of 12 weeks. These improvements were slightly lower than estimated by our aggregate-data meta-analysis [19]. In this IPD meta-analysis, we identified differential effects in relation to age, sex, and ethnicity of the participants and the number of comorbidities present. Consistent with our aggregate-data meta-analysis [19], we observed that interventions that used wearable trackers with face-to-face delivery by a professional were more effective at increasing steps per day than patient self-managed interventions.

The benefits in terms of the mean difference in steps per day after using wearable trackers in the short term were not as large as seen in other meta-analyses [46,47], but at the end of treatment, the activity tracker mean score increased to 6561 steps per day, which does appear to meet the recommended number of daily steps as outlined in public health guidelines [48,49]. For example, the average daily steps recorded were just above the recommendation of 3000 to 6000 steps per day made by the United Kingdom's National Obesity Forum and the recommendation of 3000 steps per day made by Northern Ireland's Public Health Agency [50]. However, the large standard deviation (3336) from this mean score suggests that some patients were still underachieving and did not meet the goal for steps per day. Still, recent evidence [51] has shown that for each daily increase of 1000 steps in physically inactive individuals at baseline, the estimated risk reduction in all-cause mortality is 6% to 36%, while the reduction for CVD is 5% to 21%. An increase of 500 steps per day or the equivalent (eg, 5 minutes of brisk walking) is also considered the minimum clinically important difference in steps per day in inactive adults [52]. Further, a recent dose-response meta-analysis [53] of the association between steps per day and all-cause mortality risk indicated a strong inverse association; the risk decreased linearly from 2700 to 17,000 steps per day. More specifically, the hazard ratio for 10,000 steps per day was 0.44 (95% CI 0.31 to 0.63). Even the most extremely physically inactive patients, such as the ones in our study, are still likely to benefit from small gains in steps per day.

Wearable trackers were predominately more effective in White men, but were still somewhat effective in non-White men and women. Evidence gathered from various countries shows that women are less active than men (there is a global average of 31.7% inactive women vs 23.4% inactive men) and that barriers to women's involvement in sports are numerous and complex [54-56]. Men tend to have more intrinsic motivators linked to and leading to physical activity, such as the need or desire to improve health, prevent disease onset, and improve body shape, and are also more competitive [57]. In contrast, different stimuli appear to motivate women of various ages to undertake physical exercise, such as emotional involvement, socialization, mental and physical well-being, and the achievement and maintenance of a positive self-image [58]. Women may also have less time due to daily household chores. Therefore, policies that address the sex gap in physical activity could start with better access and investment and by altering sociocultural norms. In relation to ethnicity, there is evidence that non-White participants have lower levels of physical activity and that their participation in and benefit from physical activity programs are suboptimal, due to lower access and socioeconomic and sociocultural bias [59]. Our findings suggest that non-White women in particular are less likely to achieve significant benefits from interventions involving wearable trackers. This may be because of the increased likelihood that they encounter social, economic, and cultural barriers to physical activity that are unique to them.

Wearable trackers were also most effective for improving physical activity in participants aged 49 years or more, and were surprisingly ineffective in the 365 participants aged 50 years or less. Although this could be a sample size issue, there are some possible explanations for this finding. People under 50 years old may be less likely to engage with wearable trackers due to caring and work responsibilities, and they may have less time to participate in daily physical activity [60]. Conversely, patients aged between 50 and 60 years generally maintained better daily physical activity levels than those over the age of 60. This may be due to multiple comorbidities, which are likely to be the leading cause of reduced physical activity in the older age group [61].

Participants with a cardiometabolic condition have a relatively higher likelihood of developing additional comorbidities, meaning that many of these high-risk patients are often diagnosed with 2 or more diseases [62]. As we have clearly shown in this study, most patients do not meet the international clinical guidelines for recommended steps per day, and when cardiometabolic risk is combined with a comorbidity, the steps-per-day performance was reduced even further. This result reinforces several key studies and policy guidelines, which show that a worsening steps-per-day performance is highly associated with multimorbidity [1,2,5,6,63]. Moreover, it is well known that low physical activity increases weight, BMI, and waist circumference, which are all key predictors for further exacerbation of comorbid chronic diseases, including diabetes, hypertension, and dyslipidemia [64]. For health care professionals, effective and practical management of patients with multimorbidity is important. A more sensitive understanding of their lifestyles and their tendency toward extremely low levels of physical activity will facilitate the support of those most in need of it.

Secondary cardiometabolic biomarkers, such as blood sugar glucose, measured as hemoglobin A_{1c} (HbA_{1c}), blood pressure, cholesterol, and BMI or weight, were all found to not be statistically significant, which may be unsurprising given the short duration of the interventions (only 12 weeks). However, one study [65] did find small but significant reductions in BMI and systolic blood pressure when using pedometers, although the patient population in this study involved a variety of outpatients, who may not have been suffering from the same severe underlying cardiometabolic health conditions. Another study [47] investigated the effectiveness of setting physical activity goals in patients with type 2 diabetes who used step counters and did not report any significant reduction in HbA_{1c} level. Similarly, a meta-analysis [46] that compared accelerometers and pedometers for improving physical activity levels and HbA_{1c} levels in people with type 2 diabetes showed no significant differences between either type of tracker. While our original review [19] did show significant reductions in HbA_{1c} levels, this can only be considered a small reduction in effect size, and is not dissimilar to our findings based on IPD (–0.19 vs –0.13).

Strengths and Limitations

This is the first IPD meta-analysis to assess the differential effects of wearable trackers for important physical activity and clinically relevant outcomes in participants with cardiometabolic conditions. Strengths include a clear standardization of definitions and outcomes, imputation of missing data, a robust analysis that included ANCOVA [66], exploration of the potential for differential effects [29], extensive data checking and work with study investigators to ensure the quality of the data set, and the inclusion of studies that mostly had low RoB. Nevertheless, we were unable to obtain IPD from 16 studies, which meant that 1697 of 3178 (53%) of potential patients were missing. However, these studies were mainly small and generally had a higher RoB than the included studies. Over two-thirds of the IPD were from White participants, meaning that all other ethnic groups (Black African, Asian, Hispanic, and others) had to be combined into one category to allow for adequate power in the subgroup analysis. Nevertheless, we found that White patients had significantly higher step counts than other ethnic groups. This result is similar to findings from a recent prospective cohort study that assessed the association of steps per day with premature all-cause mortality among Black and White men and women with coronary artery risk and showed that Black participants took fewer steps than White participants (median 8670 steps/day, IQR 6810 to 10,811, vs median 9441 steps/day, IQR 7704 to 11,329, respectively; $P<.001$) [67]. As 83% of the patients had or were at risk of type 2 diabetes and 85% of the patients were at least overweight or obese, this meant that these 2 conditions largely overlapped in the patient population, meaning that it was not possible to properly adjust for this in the analysis. In addition, as only 33% of the patients had CVD, and no other condition data were provided in the IPD, these 2 conditions could only be compared with CVD. Thus, we urge that these results are interpreted with some caution and encourage better coding of condition data, which would allow for more detailed analyses. Only 3 studies [37,38,42,43] used a behavior change framework as part of their intervention design, and only 3 studies [36-38,43] collected data over the longer term (ie, at least 1 year). Both are clear weaknesses that limit our understanding of sustained effects over time, which is an important gap in knowledge [68]. Differences in tracker functionality may also have significant effects on their performance. For instance, trackers are often criticized for not measuring daily steps precisely enough, particularly when the tracker becomes tilted below the waistline in overweight or obese individuals [69-71]. However, some more expensive wearable trackers that can sense movement in a tilted position have shown promise [72]. Diversity in wearing time and self-monitoring across the studies made it impossible to effectively categorize this information for meaningful subgroup analyses.

Moreover, following an update to our aggregate review, we found that 8 studies, which included 414 participants, used wearable trackers to measure physical activity with variable measurement outputs rather than steps per day. We excluded these 8 studies from our IPD meta-analysis because the assessment of the differential effects would have been underpowered, with only 414 participants available. Once the evidence base is more developed, we strongly encourage future efforts to compare our findings with those of studies using additional measurement outputs or other wearable technologies (eg, smartphone apps, smartwatches, and wristbands). Finally, our last search update (in December 2020) could be considered marginally outdated for an aggregate-data meta-analysis. However, data acquisition, preparation, and analysis for IPD meta-analyses requires considerably more time and resources than aggregate-data meta-analyses. We strongly recommend universal open access to trial data to speed access to IPD in the future.

Conclusion

Interventions using wearable trackers were effective at providing a small mean improvement in steps per day over short periods of use in participants with a cardiometabolic condition. Interventions with wearable trackers that were delivered and guided by a professional were most effective in White men and in those aged 50 years and older with only one comorbidity. Future research should look at ways to extend the beneficial effects of interventions with wearable trackers to other patients (particularly women) with cardiometabolic conditions and for longer periods.

Acknowledgments

We thank the authors who agreed to participate and supply their individual participant data: CFE, MRB, HE, MEC, MAT, KD, SSD, ABC, AFF, EBG, JH, PP, TY, and JH. This study was funded by the Evidence Synthesis Working Group (project 390), which is supported by the National Institute for Health Research School for Primary Care Research. The National Institute for Health Research Greater Manchester Patient Safety Translational Research Centre has partly funded the time of MP.

Data Availability

A data sharing agreement form was completed by each participant that provided individual participant data; there has not been any agreement to share the data. Therefore, authors seeking individual participant data should contact the principal investigators from the individual trials if they desire access. However, we aim to discuss with the authors the possibility of making the individual participant data database publicly available.

Disclaimer

This paper presents independent research funded by the National Institute for Health Research School for Primary Care Research Project 390. The views expressed are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care. The research team members were independent from the funding agencies. The funders had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, and approval of the manuscript.

Authors' Contributions

MP, EK, and PB developed the initial research idea and obtained funding for this study. AH, MP, EK, HVM, and PB formulated the research questions and designed the study. AH searched for published work, selected articles, extracted and analyzed data, and drafted the protocol, statistical analysis plan, and manuscript. MP, SSZ, and CG helped the article search, data extraction, and quality checking. MP substantially contributed to designing the searches and the statistical analysis plan, writing the manuscript, and interpreting the findings. EK provided statistical support for using the ipdforest command in Stata. MP, EK, PB, HVM, BM, SSZ, CG, DT, and all other authors substantially contributed to the manuscript by providing review comments and edits. All authors have read and approved the final manuscript. The manuscript’s guarantor (AH) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported, that no important aspects of the study have been omitted, and that any discrepancies from the study as planned have been explained. The corresponding author attests that all listed authors meet the authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary appendix booklet.
[DOCX File , 362 KB - jmir_v24i8e36337_app1.docx]

References

1. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021 Apr;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
2. Miranda JJ, NCD Risk Factor Collaboration (NCD-RisC)—Americas Working Group. Trends in cardiometabolic risk factors in the Americas between 1980 and 2014: a pooled analysis of population-based surveys. Lancet Glob Health 2020 Jan;8(1):e123-e133 [FREE Full text] [doi: 10.1016/S2214-109X(19)30484-X] [Medline: 31839128]
3. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1659-1724 [FREE Full text] [doi: 10.1016/S0140-6736(16)31679-8] [Medline: 27733284]
4. Physical activity. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2022-07-03]
5. Physical activity. National Institute for Health and Care Excellence. 2020. URL: https://www.nice.org.uk/guidance/lifestyle-and-wellbeing/physical-activity [accessed 2022-07-03]

6. Global Recommendations on Physical Activity for Health. World Health Organization. URL: https://www.who.int/publications/i/item/9789241599979 [accessed 2022-07-03]
7. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. J Biomed Inform 2019 May;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
8. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021 Apr 21;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
9. Afshin A, Babalola D, Mclean M, Yu Z, Ma W, Chen C, et al. Information Technology and Lifestyle: A Systematic Evaluation of Internet and Mobile Interventions for Improving Diet, Physical Activity, Obesity, Tobacco, and Alcohol Use. J Am Heart Assoc 2016 Aug 31;5(9):e003058 [FREE Full text] [doi: 10.1161/JAHA.115.003058] [Medline: 27581172]
10. Pedometers can help people get more active as part of an exercise programme. National Institute for Health and Care Research. URL: https://evidence.nihr.ac.uk/alert/pedometers-can-help-people-get-more-active-as-part-of-an-exercise-programme-/ [accessed 2022-07-03]
11. Tudor-Locke C, Williams JE, Reis JP, Pluto D. Utility of pedometers for assessing physical activity: convergent validity. Sports Med 2002;32(12):795-808. [doi: 10.2165/00007256-200232120-00004] [Medline: 12238942]
12. Bassett DR, John D. Use of pedometers and accelerometers in clinical populations: validity and reliability issues. Phys Ther Rev 2013 Jul 19;15(3):135-142. [doi: 10.1179/1743288x10y.0000000004]
13. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
14. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159-151 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
15. Franssen WMA, Franssen GHLM, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020 May 11;17(1):57-51 [FREE Full text] [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
16. Wise JM, Hongu N. Pedometer, Accelerometer, And mobile technology for Promoting Physical Activity. The University of Arizona Cooperative Extension. URL: https://extension.arizona.edu/sites/extension.arizona.edu/files/pubs/az1491-2014.pdf [accessed 2022-07-03]
17. Ayabe M, Ishii K, Takayama K, Aoki J, Tanaka H. Comparison of interdevice measurement difference of pedometers in younger and older adults. Br J Sports Med 2010 Feb 28;44(2):95-99. [doi: 10.1136/bjsm.2007.045179] [Medline: 18308892]
18. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. Am J Health Promot 2019 Jun 26;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]
19. Hodkinson A, Kontopantelis E, Adeniji C, van Marwijk H, McMillian B, Bower P, et al. Interventions Using Wearable Physical Activity Trackers Among Adults With Cardiometabolic Conditions: A Systematic Review and Meta-analysis. JAMA Netw Open 2021 Jul 01;4(7):e2116382 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.16382] [Medline: 34283229]
20. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 2010 Feb 05;340(feb05 1):c221-c221. [doi: 10.1136/bmj.c221] [Medline: 20139215]
21. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, PRISMA-IPD Development Group. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA 2015 Apr 28;313(16):1657-1665. [doi: 10.1001/jama.2015.3656] [Medline: 25919529]
22. Lee I, 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]
23. Altman DG, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, NY: John Wiley & Sons; 2008.
24. Van Buuren S. MICE: Multivariate Imputation by Chained Equations. Carnegie Mellon University. URL: http://lib.stat.cmu.edu/R/CRAN/web/packages/mice/index.html [accessed 2022-07-03]
25. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons; 1987.
26. Kontopantelis E. A comparison of one-stage vs two-stage individual patient data meta-analysis methods: A simulation study. Res Synth Methods 2018 Sep 21;9(3):417-430 [FREE Full text] [doi: 10.1002/jrsm.1303] [Medline: 29786975]
27. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. Stat Med 2017 Feb 28;36(5):855-875 [FREE Full text] [doi: 10.1002/sim.7141] [Medline: 27747915]
28. Kontopantelis E, Reeves D. A Short Guide and a Forest Plot Command (Ipdforest) for One-Stage Meta-Analysis. Stata J 2013 Sep 01;13(3):574-587 [FREE Full text] [doi: 10.1177/1536867x1301300308]

29. Hua H, Burke DL, Crowther MJ, Ensor J, Tudur Smith C, Riley RD. One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information. Stat Med 2017 Feb 28;36(5):772-789 [FREE Full text] [doi: 10.1002/sim.7171] [Medline: 27910122]
30. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 2014 Feb 18;14(1):25 [FREE Full text] [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
31. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
32. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011 Jul 22;343(jul22 1):d4002-d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]
33. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
34. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. BMJ 2012 Jan 03;344(1):d7762-d7762. [doi: 10.1136/bmj.d7762] [Medline: 22214758]
35. Anderson DR. Health Beliefs, Will to Live, Hope, and Social Support in a Pedometer-Based Exercise Intervention among Cardiac Rehabilitation Patients. The Ohio State University. 2015. URL: https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1434901973&disposition=inline&psig=AOvVaw1mbfdGQZAw0Jr6ZoSCUHuD&ust=1657473325229075 [accessed 2022-07-12]
36. Dasgupta K, Rosenberg E, Joseph L, Cooke AB, Trudeau L, Bacon SL, SMARTER Trial Group. Physician step prescription and monitoring to improve ARTERial health (SMARTER): A randomized controlled trial in patients with type 2 diabetes and hypertension. Diabetes Obes Metab 2017 May 22;19(5):695-704 [FREE Full text] [doi: 10.1111/dom.12874] [Medline: 28074635]
37. Houle J, Doyon O, Vadeboncoeur N, Turbide G, Diaz A, Poirier P. Innovative program to increase physical activity following an acute coronary syndrome: randomized controlled trial. Patient Educ Couns 2011 Dec;85(3):e237-e244. [doi: 10.1016/j.pec.2011.03.018] [Medline: 21546203]
38. Houle J, Doyon O, Vadeboncoeur N, Turbide G, Diaz A, Poirier P. Effectiveness of a pedometer-based program using a socio-cognitive intervention on physical activity and quality of life in a setting of cardiac rehabilitation. Can J Cardiol 2012 Jan;28(1):27-32. [doi: 10.1016/j.cjca.2011.09.020] [Medline: 22177854]
39. Katzmarzyk PT, Champagne CM, Tudor-Locke C, Broyles ST, Harsha D, Kennedy BM, et al. A short-term physical activity randomized trial in the Lower Mississippi Delta. PLoS One 2011 Oct 25;6(10):e26667 [FREE Full text] [doi: 10.1371/journal.pone.0026667] [Medline: 22046325]
40. Araiza, Hewes H, Gashetewa C, Vella CA, Burge MR. Efficacy of a pedometer-based physical activity program on parameters of diabetes control in type 2 diabetes mellitus. Metabolism 2006 Oct;55(10):1382-1387. [doi: 10.1016/j.metabol.2006.06.009] [Medline: 16979410]
41. Cupples M, Dean D, Tully MA, Taggart M, McCorkell G, O'Neill S, et al. Using Pedometer Step-Count Goals to Promote Physical Activity in Cardiac Rehabilitation: A Feasibility Study of a Controlled Trial. Int J Phys Med Rehabil 2013;04(07):1 [FREE Full text] [doi: 10.4172/2329-9096.1000157]
42. Grey EB, Thompson D, Gillison FB. Effects of a Web-Based, Evolutionary Mismatch-Framed Intervention Targeting Physical Activity and Diet: a Randomised Controlled Trial. Int J Behav Med 2019 Dec 25;26(6):645-657 [FREE Full text] [doi: 10.1007/s12529-019-09821-3] [Medline: 31654276]
43. Yates T, Edwardson CL, Henson J, Gray LJ, Ashra NB, Troughton J, et al. Walking Away from Type 2 diabetes: a cluster randomized controlled trial. Diabet Med 2017 May 19;34(5):698-707. [doi: 10.1111/dme.13254] [Medline: 27589017]
44. Fayehun AF, Olowookere OO, Ogunbode AM, Adetunji AA, Esan A. Walking prescription of 10 000 steps per day in patients with type 2 diabetes mellitus: a randomised trial in Nigerian general practice. Br J Gen Pract 2018 Feb;68(667):e139-e145 [FREE Full text] [doi: 10.3399/bjgp18X694613] [Medline: 29335328]
45. Greaney ML, Askew S, Wallington SF, Foley PB, Quintiliani LM, Bennett GG. The effect of a weight gain prevention intervention on moderate-vigorous physical activity among black women: the Shape Program. Int J Behav Nutr Phys Act 2017 Oct 16;14(1):139 [FREE Full text] [doi: 10.1186/s12966-017-0596-6] [Medline: 29037247]
46. Baskerville R, Ricci-Cabello I, Roberts N, Farmer A. Impact of accelerometer and pedometer use on physical activity and glycaemic control in people with Type 2 diabetes: a systematic review and meta-analysis. Diabet Med 2017 May 19;34(5):612-620. [doi: 10.1111/dme.13331] [Medline: 28173623]
47. Qiu S, Cai X, Chen X, Yang B, Sun Z. Step counter use in type 2 diabetes: a meta-analysis of randomized controlled trials. BMC Med 2014 Feb 27;12(1):36 [FREE Full text] [doi: 10.1186/1741-7015-12-36] [Medline: 24571580]
48. Physical activity; walking and cycling: Public health guideline [PH41]. National Institute for Health and Clinical Excellence. URL: https://www.nice.org.uk/guidance/ph41 [accessed 2018-09-10]
49. School Health Guidelines to Promote Healthy Eating and Physical Activity. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/mmwr/pdf/rr/rr6005.pdf [accessed 2022-07-03]

50. Tudor-Locke C, Burkett L, Reis JP, Ainsworth BE, Macera CA, Wilson DK. How many days of pedometer monitoring predict weekly physical activity in adults? Prev Med 2005 Mar;40(3):293-298. [doi: 10.1016/j.ypmed.2004.06.003] [Medline: 15533542]
51. Hall KS, Hyde ET, Bassett DR, Carlson SA, Carnethon MR, Ekelund U, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. Int J Behav Nutr Phys Act 2020 Jun 20;17(1):78 [FREE Full text] [doi: 10.1186/s12966-020-00978-9] [Medline: 32563261]
52. Rowlands A, Davies M, Dempsey P, Edwardson C, Razieh C, Yates T. Wrist-worn accelerometers: recommending ~1.0 m as the minimum clinically important difference (MCID) in daily average acceleration for inactive adults. Br J Sports Med 2021 Jul 14;55(14):814-815. [doi: 10.1136/bjsports-2020-102293] [Medline: 32928741]
53. Jayedi A, Gohari A, Shab-Bidar S. Daily Step Count and All-Cause Mortality: A Dose-Response Meta-analysis of Prospective Cohort Studies. Sports Med 2022 Jan 21;52(1):89-99. [doi: 10.1007/s40279-021-01536-4] [Medline: 34417979]
54. The Lancet Public Health. Time to tackle the physical activity gender gap. Lancet Public Health 2019 Aug;4(8):e360 [FREE Full text] [doi: 10.1016/S2468-2667(19)30135-5] [Medline: 31345750]
55. Barker J, Smith Byrne K, Doherty A, Foster C, Rahimi K, Ramakrishnan R, et al. Physical activity of UK adults with chronic disease: cross sectional analysis of accelerometer measured physical activity in 96 706 UK Biobank participants. Int J Epidemiol 2019 Aug 01;48(4):1386 [FREE Full text] [doi: 10.1093/ije/dyz148] [Medline: 31280303]
56. Manjoo P, Joseph L, Pilote L, Dasgupta K. Sex differences in step count-blood pressure association: a preliminary study in type 2 diabetes. PLoS One 2010 Nov 22;5(11):e14086 [FREE Full text] [doi: 10.1371/journal.pone.0014086] [Medline: 21124929]
57. 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]
58. Vaccarezza M, Papa V, Milani D, Gonelli A, Secchiero P, Zauli G, et al. Sex/Gender-Specific Imbalance in CVD: Could Physical Activity Help to Improve Clinical Outcome Targeting CVD Molecular Mechanisms in Women? Int J Mol Sci 2020 Feb 21;21(4):1477 [FREE Full text] [doi: 10.3390/ijms21041477] [Medline: 32098263]
59. Such E, Salway S, Copeland R, Haake S, Domone S, Mann S. A formative review of physical activity interventions for minority ethnic populations in England. J Public Health (Oxf) 2017 Dec 01;39(4):e265-e274 [FREE Full text] [doi: 10.1093/pubmed/fdw126] [Medline: 27899479]
60. Khanom A, Evans BA, Lynch R, Marchant E, Hill RA, Morgan K, et al. Parent recommendations to support physical activity for families with young children: Results of interviews in deprived and affluent communities in South Wales (United Kingdom). Health Expect 2020 Apr;23(2):284-295 [FREE Full text] [doi: 10.1111/hex.13020] [Medline: 31898386]
61. Watson KB, Carlson SA, Gunn JP, Galuska DA, O'Connor A, Greenlund KJ, et al. Physical Inactivity Among Adults Aged 50 Years and Older - United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Sep 16;65(36):954-958 [FREE Full text] [doi: 10.15585/mmwr.mm6536a3] [Medline: 27632143]
62. Kivimäki M, Kuosma E, Ferrie JE, Luukkonen R, Nyberg ST, Alfredsson L, et al. Overweight, obesity, and risk of cardiometabolic multimorbidity: pooled analysis of individual-level data for 120 813 adults from 16 cohort studies from the USA and Europe. Lancet Public Health 2017 Jun;2(6):e277-e285 [FREE Full text] [doi: 10.1016/S2468-2667(17)30074-9] [Medline: 28626830]
63. Knaeps S, Bourgois JG, Charlier R, Mertens E, Lefevre J, Wijndaele K. Ten-year change in sedentary behaviour, moderate-to-vigorous physical activity, cardiorespiratory fitness and cardiometabolic risk: independent associations and mediation analysis. Br J Sports Med 2018 Aug 04;52(16):1063-1068 [FREE Full text] [doi: 10.1136/bjsports-2016-096083] [Medline: 27491779]
64. Carson V, Rinaldi RL, Torrance B, Maximova K, Ball GDC, Majumdar SR, et al. Vigorous physical activity and longitudinal associations with cardiometabolic risk factors in youth. Int J Obes (Lond) 2014 Jan 26;38(1):16-21. [doi: 10.1038/ijo.2013.135] [Medline: 23887061]
65. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
66. Vickers AJ, Altman DG. Statistics notes: Analysing controlled trials with baseline and follow up measurements. BMJ 2001 Nov 10;323(7321):1123-1124 [FREE Full text] [doi: 10.1136/bmj.323.7321.1123] [Medline: 11701584]
67. Paluch AE, Gabriel KP, Fulton JE, Lewis CE, Schreiner PJ, Sternfeld B, et al. Steps per Day and All-Cause Mortality in Middle-aged Adults in the Coronary Artery Risk Development in Young Adults Study. JAMA Netw Open 2021 Sep 01;4(9):e2124516 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.24516] [Medline: 34477847]
68. Khunti K, Griffin S, Brennan A, Dallosso H, Davies MJ, Eborall HC, et al. Promoting physical activity in a multi-ethnic population at high risk of diabetes: the 48-month PROPELS randomised controlled trial. BMC Med 2021 Jun 03;19(1):130-131 [FREE Full text] [doi: 10.1186/s12916-021-01997-4] [Medline: 34078362]
69. Crouter S, Schneider PL, Bassett DR. Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. Med Sci Sports Exerc 2005 Oct;37(10):1673-1679. [doi: 10.1249/01.mss.0000181677.36658.a8] [Medline: 16260966]

70. Feito Y, Bassett DR, Tyo B, Thompson DL. Effects of body mass index and tilt angle on output of two wearable activity monitors. Med Sci Sports Exerc 2011 May;43(5):861-866. [doi: 10.1249/MSS.0b013e3181fefd40] [Medline: 20962689]
71. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step Counting: A Review of Measurement Considerations and Health-Related Applications. Sports Med 2017 Jul 22;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
72. De Vries SI, Van Hirtum HWJEM, Bakker I, Hopman-Rock M, Hirasing RA, Van Mechelen W. Validity and reproducibility of motion sensors in youth: a systematic update. Med Sci Sports Exerc 2009 Apr;41(4):818-827. [doi: 10.1249/MSS.0b013e31818e5819] [Medline: 19276851]

Abbreviations

ANCOVA: analysis of covariance
CVD: cardiovascular disease
HbA$_{1c}$: hemoglobin A$_{1c}$
IPD: individual participant data
MD: mean difference
RCT: randomized controlled trial
RoB: risk of bias

Edited by R Kukafka; submitted 11.01.22; peer-reviewed by C Brakenridge, A Martinko; comments to author 23.01.22; revised version received 14.02.22; accepted 04.04.22; published 30.08.22.

Please cite as:
Hodkinson A, Kontopantelis E, Zghebi SS, Grigoroglou C, McMillan B, Marwijk HV, Bower P, Tsimpida D, Emery CF, Burge MR, Esmiol H, Cupples ME, Tully MA, Dasgupta K, Daskalopoulou SS, Cooke AB, Fayehun AF, Houle J, Poirier P, Yates T, Henson J, Anderson DR, Grey EB, Panagioti M
Association Between Patient Factors and the Effectiveness of Wearable Trackers at Increasing the Number of Steps per Day Among Adults With Cardiometabolic Conditions: Meta-analysis of Individual Patient Data From Randomized Controlled Trials
J Med Internet Res 2022;24(8):e36337
URL: https://www.jmir.org/2022/8/e36337
doi:10.2196/36337
PMID:36040779

Review

Digital Interventions to Enhance Readiness for Psychological Therapy: Scoping Review

Jacinta Jardine[1], BA, HDip; Robert Bowman[1], MA, MSc; Gavin Doherty[1], BA, DPhil

School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

Corresponding Author:
Jacinta Jardine, BA, HDip
School of Computer Science and Statistics
Trinity College Dublin
Hamilton Building
College Green
Dublin, D02 W272
Ireland
Phone: 353 861505715
Email: jjardine@tcd.ie

Abstract

Background: Psychological therapy is an effective treatment method for mental illness; however, many people with mental illness do not seek treatment or drop out of treatment early. Increasing client uptake and engagement in therapy is key to addressing the escalating global problem of mental illness. Attitudinal barriers, such as a lack of motivation, are a leading cause of low engagement in therapy. Digital interventions to increase motivation and readiness for change hold promise as accessible and scalable solutions; however, little is known about the range of interventions being used and their feasibility as a means to increase engagement with therapy.

Objective: This review aimed to define the emerging field of digital interventions to enhance readiness for psychological therapy and detect gaps in the literature.

Methods: A literature search was conducted in PubMed, PsycINFO, PsycARTICLES, Scopus, Embase, ACM Guide to Computing Literature, and IEEE Xplore Digital Library from January 1, 2006, to November 30, 2021. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) methodology was applied. Publications were included when they concerned a digitally delivered intervention, a specific target of which was enhancing engagement with further psychological treatment, and when this intervention occurred before the target psychological treatment.

Results: A total of 45 publications met the inclusion criteria. The conditions included depression, unspecified general mental health, comorbid anxiety and depression, smoking, eating disorders, suicide, social anxiety, substance use, gambling, and psychosis. Almost half of the interventions (22/48, 46%) were web-based programs; the other formats included screening tools, videos, apps, and websites. The components of the interventions included psychoeducation, symptom assessment and feedback, information on treatment options and referrals, client testimonials, expectation management, and pro-con lists. Regarding feasibility, of the 16 controlled studies, 7 (44%) measuring actual behavior or action showed evidence of intervention effectiveness compared with controls, 7 (44%) found no differences, and 2 (12%) indicated worse behavioral outcomes. In general, the outcomes were mixed and inconclusive owing to variations in trial designs, control types, and outcome measures.

Conclusions: Digital interventions to enhance readiness for psychological therapy are broad and varied. Although these easily accessible digital approaches show potential as a means of preparing people for therapy, they are not without risks. The complex nature of stigma, motivation, and individual emotional responses toward engaging in treatment for mental health difficulties suggests that a careful approach is needed when developing and evaluating digital readiness interventions. Further qualitative, naturalistic, and longitudinal research is needed to deepen our knowledge in this area.

(J Med Internet Res 2022;24(8):e37851) doi:10.2196/37851

KEYWORDS

readiness for change; stages of change; digital; motivation; engagement; uptake; mental health; mental illness; mobile phone

Introduction

Background

Mental illness is a pervasive global problem, estimated to be the second most prominent cause of the global burden of disease, surpassed only by cardiovascular disease [1]. Psychological therapy is both an effective and acceptable treatment for common mental illnesses such as anxiety and depression [2,3], with comparable outcomes across all approaches (ie, cognitive behavioral therapy [CBT], psychodynamic, and person centered) [4] and delivery formats (ie, face-to-face [FTF] and digital format) [5]. Despite the demonstrated effectiveness of psychological therapy, there remains an alarming difference between the number of people with a mental illness and the number of people being treated, often referred to as the *mental health treatment gap* [6]. This gap is substantial and ever expanding, as prevalence rises without a corresponding rise in treatment outreach or provision [6].

One significant problem that perpetuates this treatment gap is *client engagement* in therapy [7,8]. Engagement is a term with many associated meanings [9,10]; for the purposes of this review, we use it as an overarching term to represent client uptake (ie, whether the client begins treatment), as well as the client's ongoing, active participation in treatment. Client engagement is essential for clients to obtain said favorable outcomes [11-13]; however, because it is an internal cognitive state, it is difficult to measure [9]. Consequently, engagement is often inferred from observing more easily quantifiable metrics, such as adherence, dropout, and use in the digital realm. The uptake rates of digital mental health treatments are estimated to range from 3% to 25% [14]. Low use and high dropout rates are persistent problems when it comes to digital solutions [15], although similar problems also affect FTF modalities; between 17% and 25% of clients are estimated to drop out of FTF psychotherapy [8,16,17]. Considering that only 20% of the people with mental health problems seek treatment in the first place [18], the problem becomes even more apparent. Therefore, increasing client engagement is a key focus area in the wider mental health sphere [10].

Barriers to Engagement

Many of the practical barriers that have historically impeded access to and engagement with FTF psychological therapy (eg, cost, accessibility, and time constraints) [19] have been reduced with the emergence of digitally delivered treatments. However, this new treatment modality introduces its own set of novel barriers, such as internet anxiety, privacy concerns, lack of confidence in using technology, and disbelief in the effectiveness of the treatments themselves [20-22].

Arguably, the most significant barriers to engagement across all types of therapy delivery stem from the client's attitude toward seeking help and engaging in therapy [23]. Among these attitudinal barriers, low perceived need, a preference to deal with the problem on one's own, and internalized self-stigma are the most common [21,23,24].

Motivation to Change

Motivation is a term used to describe the analytical and habitual processes that energize and direct behavior [25], thus encompassing the attitudinal barriers discussed earlier, among other factors. It is easy to assume that individuals presenting for treatment are motivated to engage in the process and make changes in their lives; however, research indicates that up to 80% of the people who seek treatment are not ready to change and that a leading cause of treatment dropout and low adherence is a lack of motivation [26,27].

The most prominent theory explaining motivation for therapy and readiness for change is the *Transtheoretical Model* (TTM), which posits that clients move through a series of stages on their journey toward and through the process of change [28]. This theory describes behavior change not only in terms of action but also as a wider contemplative process that begins before a person is even considering change [17]. The *stages of change* presented in the TTM are precontemplation, contemplation, preparation, action, and maintenance [28]; the stage a client is in before treatment positively correlates with their outcomes after treatment (ie, the further along they are in terms of the stages, the better their outcomes) [29,30].

The mechanism by which a client's stage of change affects their overall therapy outcomes could manifest in their initial experiences of treatment [31]. For example, if the client is in the action stage at the onset of treatment, they can fully engage with the process immediately rather than spending initial sessions or interactions in ambivalence, thus delaying improvements [32-34]. As symptom changes that occur early in treatment are linked to greater overall treatment success [11,33], targeting those clients who are not yet in the action stage of change *before* they commence treatment, with interventions designed to move them toward action, could mean that more clients begin therapy, stay engaged, and reach positive outcomes.

Motivational Interviewing

There are several FTF pretherapy interventions aimed at moving clients through the stages of change and preparing them for subsequent therapy. Examples of such pretherapy interventions include motivational interviewing (MI), role induction, and vicarious therapy pretraining [27]. MI is arguably the most established of these interventions due to its significant effects on client adherence to subsequent therapy as well as treatment outcomes [26,32,35]. MI is a collaborative, discursive therapeutic approach that aims to guide rather than direct clients, fostering autonomy through open questions and evoking the client's personal reasons for change [36,37]. The specific techniques or tools used by MI practitioners (we will refer to these as "components") include exploring reasons for change, weighing up the pros and cons of change, developing discrepancy between the client's ideal and current states, and building confidence and self-efficacy [38]. The key causal model of MI is that client speech affects client outcome [39], meaning that the more favorably a client talks about behavior change, the more likely they are to make the change. Helping clients achieve this "change talk" is a highly nuanced, conversational art undertaken by skilled practitioners over multiple sessions

[36]. Originally developed as a brief stand-alone intervention for alcohol misuse, MI is now used as a pretherapy intervention for a range of mental illnesses, including anxiety and depression [35]. However, owing to its current FTF delivery format, traditional MI is not a widely available or accessible option for millions of people experiencing mental health difficulties around the world. Finding a feasible way to deliver interventions, such as MI, in a more accessible format could help more clients become motivated and begin treatment ready to take action.

Digital Readiness Interventions

Digital methods of intervention delivery hold promise as a way of creating accessible and timely solutions that can be easily scaled to cover entire populations, including those who have not yet reached out for help [10]. A recent systematic review of technology-assisted MI indicated its potential in this area [38]. However, in most of the included studies, the MI components were integrated with other approaches (eg, CBT) and used as stand-alone digital treatments targeted at changing problem behaviors, such as alcohol use and smoking [38], rather than as motivational *pretherapy* interventions. The extent to which digitally delivered MI has been used as a readiness intervention to prepare clients for therapy is unclear. Furthermore, little is known about the feasibility of delivering such a conversational and highly tailored process via digital means [38].

Outside MI, other digital approaches have begun to emerge, such as engagement-facilitation interventions, which aim to increase both the uptake of and adherence to web-based mental health programs [40]. The components of these interventions differ from those used in MI; for example, engagement-facilitation interventions include components such as expectation setting, psychoeducation about symptoms and treatment, treatment belief enhancement, symptom assessment, and assessment feedback [20]. At present, little is known about the full range of different types of digital interventions that are being used to prepare clients for further therapy, the components of these interventions, or the design processes used in their development. Furthermore, research in this area is spread across the digital health, behavior change, and human-computer interaction fields. Thus, a review is needed to scope this topic and clarify the current dispersed and diverse body of research.

Collaboration with clinical professionals and human-centered design processes are key to developing effective mental health interventions, given their sensitive and complex nature [41,42]. As this is an emerging field, formative research exploring intervention design, development, and evaluation can provide insights into opportunities, barriers, and design strategies that can be used to create effective and acceptable solutions.

This Study

The aim of this study was to define the emerging field of digital interventions to enhance readiness for psychological therapy. By exploring the current state of research in this area, we hope to identify the conceptual boundaries of the topic and identify gaps in the literature. Our research questions were as follows:

1. What types of digital interventions have been used to prepare clients for psychological therapy?
2. What components have been used in these interventions and which of these show evidence of effectiveness?
3. What design processes have been used to develop these interventions?
4. Is the digital delivery of preparatory interventions to enhance readiness for psychological treatment feasible?

Methods

Protocol and Structure

The protocol for this review was registered with the Open Science Framework on March 26, 2021 [43]. We used the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews) guidelines to structure our review [44].

Study Design

We chose a scoping review approach because the research studies in question are heterogeneous in nature and spread across multiple disciplines; they use different study designs to measure different outcomes, with different populations, in different contexts. As this is an emerging field, there are few boundaries on the extent, range, and nature of evidence [44], and the terminology used in the published literature is inconsistent and varied [45]. Therefore, this exploratory review type is well suited.

Eligibility Criteria

Publications were included for assessment if they met the following criteria: (1) the article concerns an intervention, a specific target of which is enhancing engagement with *further psychological* treatment or therapy; (2) the intervention is delivered *digitally* (ie, the primary active content of the intervention is digital), but studies that use technology solely as a means of synchronous communication (eg, web chat or video calls) were excluded; (3) the intervention took place *before* the target psychological treatment (ie, not combined or performed in tandem with the target treatment); (4) the article was written in English; (5) the article was published in a peer-reviewed publication between 2006 and 2021; and (6) the intervention was designed for adult or adolescent populations (ie, age ≥12 years).

The rationale for examining only recent evidence (past 15 years) is that digital technology is advancing rapidly; older studies may be out of date in terms of client attitudes and acceptance of technology [46]. Comparable time frames have been used in many recent reviews on digital mental health technologies [38,46,47]. We included adolescent populations in our review because research indicates that the main barriers to engagement with mental health treatments are comparable across adult and adolescent populations [48].

Search Strategy

The following electronic databases were searched: PubMed, PsycINFO, PsycARTICLES, Scopus, Embase, ACM Guide to Computing Literature, and IEEE Xplore Digital Library. Search terms reflected the 3 main eligibility criteria (Table 1).

Table 1. Search terms.

Criteria	MeSH[a] terms	Free-text terms
Target treatment (further psychological treatment or therapy)	"Mental Health" OR "Psychotherapy" OR "Stress, Psychological" OR "Anxiety Disorders" OR "Mood Disorders"	"CBT" OR "psychological" OR "mental ill-health" OR "anxiety" OR "depressi*" OR "stress" OR "wellbeing" OR "well-being" OR "resilience" OR "mood" OR "disorder*" OR "phobia*"
Digital delivery	"Therapy, Computer-Assisted" OR "Internet" OR "Digital Technology"	"digital" OR "technolog*" OR "comput*" OR "e-health" OR "ehealth" OR "m-health" OR "mhealth" OR "mobile" OR "online" OR "web" OR "web-based" OR "smartphone*"
Intervention type (readiness intervention; takes place before the target treatment)	"Transtheoretical Model" OR "Motivational Interviewing"	"readiness" OR "pre-therapy" OR "pre-treatment" OR prepar* OR "prelude" OR "prequel" OR "prior" OR "stage of change" OR "stages of change" OR "motivation to change" OR "motivational enhancement" OR "motivation interview" OR "motivational intervention"

[a] MeSH: Medical Subject Headings.

Data Collection

An initial exploratory search of PubMed and ACM databases was conducted, and words contained in the titles and abstracts of retrieved papers were analyzed. The search terms were adjusted based on the identified papers, and the final search strategy was decided. Once the protocol was registered with the Open Science Framework, a full search was undertaken across all included databases in March 2021; the search was updated in November 2021. Additional records were retrieved by checking the reference lists of included articles.

The first and second authors (JJ and RB) began by independently reviewing a subset (1300/9412, 15%) of the titles and abstracts against the eligibility criteria and comparing their findings. Discrepancies were found; hence, the eligibility criteria were clarified through discussion between the 2 authors by using relevant examples from the first sample reviewed. A further subset (1300/9412, 15%) was reviewed, the findings were compared, and consensus in decision-making about inclusion and exclusion was reached. The remaining articles were then split between the 2 reviewers (JJ and RB), who independently assessed the titles and abstracts. The final list of selected articles was reviewed by both reviewers. The first author (JJ) then retrieved the full text of the selected articles, and both reviewers independently evaluated them against the eligibility criteria. Reasons for exclusion were recorded, and where there were discrepancies, a discussion was held between the 2 reviewers, and a consensus was reached on the final selection of articles.

Data Analysis

Data charting was performed in Excel (Microsoft Corporation) by the first author (JJ), with checks and calibration by the second author (RB). The data charting form contained general study details as well as variables related to research questions, including target treatment, intervention type (eg, technology used, duration, and interaction level), intervention pathway (eg, how and when the intervention was delivered to clients and the relationship between the intervention and target treatment), intervention components, the model or framework used, measures and outcomes, user experience or acceptability, design process, critical appraisal (eg, limitations in the study, biases, strength of methodology, and generalizability of results), and key learnings.

Synthesis of Results

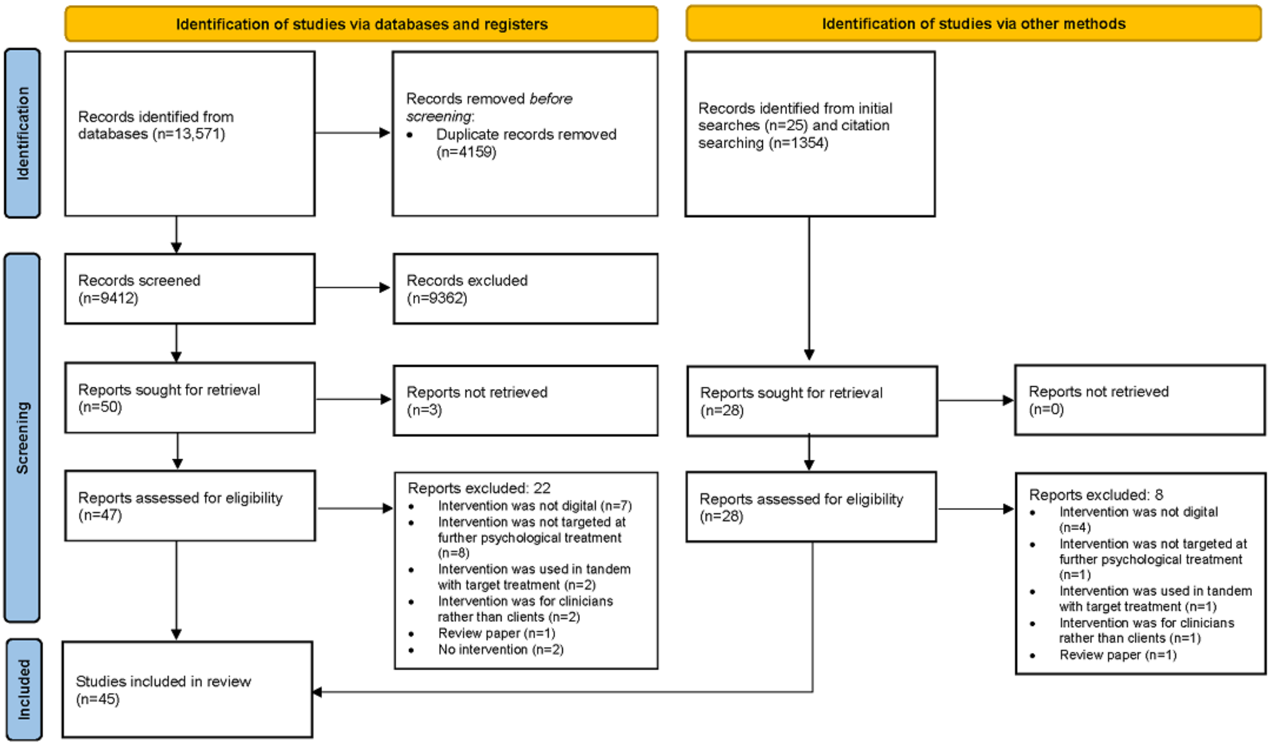
The charted data were further summarized based on the key characteristics of the data. For example, within a charted column such as *the target treatment* or *duration*, findings were assessed in relation to each other, and overarching categories were created based on the most common results. The frequency of occurrences was then examined, and result tables were created. In terms of more complex findings, such as components and outcomes, separate Excel worksheets were created, where individual studies or interventions could be explored in more detail. Frequent checks of the full paper were conducted to validate the initial charting.

Results

Study Selection

The search resulted in 13,571 hits. A further 1379 studies were identified via other sources. After removing duplicates and screening titles, abstracts, and full texts, 45 (0.30%) papers met the eligibility criteria (Figure 1).

Figure 1. Flow diagram of the study selection process.



Study Characteristics

The studies included in this review (Table 2) were mainly conducted in the United States (17/45, 38%), Australia (9/45, 20%), and Germany (7/45, 15%). Only 2% (1/45) of the studies included multiple countries [49]. In terms of study design, among the 45 studies, 19 (43%) were randomized controlled trials, 7 (16%) were observational studies, 6 (13%) were protocols, 4 (9%) were studies exploring the development process of interventions, 3 (7%) were pre-post designs, 2 (4%) were nonrandomized controlled trials, 2 (4%) were historically controlled studies, 1 (2%) was a qualitative evaluation, and 1 (2%) study presented the results of multiple studies (2 randomized controlled trials and 1 pre-post study). Only 26% (12/45) of the studies included qualitative data collection, 18% (8/45) used mixed methods [50-57], and 9% (4/45) were purely qualitative [58-61]. Depression (9/45, 20%) and unspecified general mental health (9/45, 20%) were the most common target conditions covered, with comorbid anxiety and depression being the next most frequent (6/45, 14%). The other conditions and problems targeted included smoking (5/45, 11%), eating disorders (5/45, 11%), suicide (3/45, 7%), social anxiety (3/45, 7%), substance use (2/45, 4%), gambling (2/45, 4%), and psychosis (1/45, 2%).

Table 2. Characteristics of the included papers.

Author, study, and year	Study design	Country	Condition
Christensen et al [62], 2006	RCT[a]	Australia	Depression
Reis and Brown [63], 2006	RCT	The United States	General mental health
Haas et al [57], 2008	Observational	The United States	Suicide
Costin et al [64], 2009	RCT	Australia	Depression
Olson et al [56], 2009	Historically controlled	The United States	General mental health
Titov et al [65], 2010	RCT	Australia	Social anxiety
Brunette et al [66], 2011	NRCT[b]	The United States	Smoking
Johansen et al [67], 2011	RCT	The United States	General mental health
Strassle et al [68], 2011	RCT	The United States	General mental health
Ferron et al [69], 2012	Observational	The United States	Smoking
Reins et al [70], 2013	Protocol	Germany	Depression
Hötzel et al [71], 2014	RCT	Germany	Eating disorder
Taylor-Rodgers and Batterham [72], 2014	RCT	Australia	Anxiety and depression
Ahmedani et al [73], 2015	Pre-post	The United States	Depression
Ebert et al [14], 2015	RCT	Germany	Depression
King et al [74], 2015	RCT	The United States	Suicide
Batterham et al [75], 2016	RCT	Australia	Anxiety and depression
BinDhim et al [49], 2016	Observational	Australia, the United Kingdom, Canada, New Zealand, and the United States	Depression
Moessner et al [76], 2016	Observational	Germany	Eating disorder
Birnbaum et al [77], 2017	Observational	The United States	Psychosis
Bommelé et al [78], 2017	NRCT	The Netherlands	Smoking
Brown et al [55], 2017	Development process	The United States	Smoking
Griffiths et al [79], 2017	RCT	Australia	Social anxiety
Krampe et al [80], 2017	RCT	Germany	General mental health
Metz et al [81], 2017	Protocol	The Netherlands	General mental health
Muir et al [58], 2017	Development process	The United Kingdom	Eating disorder
Liu et al [60], 2018	Development process	New Zealand	General mental health
Suka et al [82], 2018	Observational	Japan	Depression
Batterham et al [40], 2019	Protocol	Australia	Anxiety and depression
Dannenberg et al [59], 2019	Development process	The United States	Depression
Denison-Day et al [54], 2019	RCT	The United Kingdom	Eating disorder
Dreier et al [50], 2019	Protocol	Germany	Suicide
Ebert et al [83], 2019	RCT	Germany	General mental health
Johansen et al [61], 2019	Qualitative	Norway	Gambling
McLean et al [84], 2019	Observational	Australia	Eating disorder
Shand et al [85], 2019	Protocol	Australia	Depression
Beck et al [51], 2020	Pre-post	Canada	Anxiety and depression
Brunette et al [86], 2020	RCT	The United States	Smoking
Duffy et al [53], 2020	Pre-post	The United Kingdom	Anxiety and depression
Peter et al [87], 2020	RCT	The United States	Gambling
Keller et al [52], 2021	RCT and pre-post	The United States	General mental health

Author, study, and year	Study design	Country	Condition
Olthof et al [88], 2021	Protocol	The Netherlands	Substance use
Soucy et al [89], 2021	RCT	Canada	Anxiety and depression
Tobias et al [90], 2021	RCT	The United States	Social anxiety
Yoon et al [91], 2021	Historically controlled	The United States	Substance use

[a] RCT: randomized controlled trial.

[b] NRCT: nonrandomized controlled trial.

Types of Intervention

To assess the interventions analyzed in the included papers, we first distinguished the interventions themselves from the papers. A total of 6 studies in the sample [62,64,65,67,86,90] assessed 2 distinctly different interventions in their analysis (ie, the components of the interventions were distinct); therefore, we separated them into individual records. In all, 2 studies assessed slight variations in the same intervention [52,87]; however, we did not segregate these studies because they only reflected minor variations in what was essentially the same intervention. Furthermore, 3 sets of 2 studies in the sample analyzed the same interventions: [66,69], [54,58], and [51,89]; therefore, we grouped them together. The final list of 48 interventions is analyzed in this section and the subsequent one.

We explored the interventions under several categories: intervention format, target treatment or therapy for which the intervention was designed to prepare clients for, the level of support provided, whether the intervention was designed for repeated or once-off use, the duration of the intervention, the theoretical model used, and how the intervention was delivered to the client (Table 3). We found many variations in the types of interventions used to prepare people for psychological therapy. Almost half of the interventions (22/48, 46%) were web-based programs; the other formats included screening tools, videos, apps, and websites. Many of the included interventions were not designed to prepare clients for a specific treatment but instead to encourage general professional help seeking (27/48, 56%). Of those targeted at specific treatments, FTF therapy was the most common (14/48, 29%), followed by web-based therapy (6/48, 13%), and phone therapy (1/48, 2%). In terms of the duration of the interventions, those that specified a duration ranged from 15 seconds to 6 months, with most interventions taking <90 minutes to complete (23/48, 48%). We also investigated how and when the interventions were delivered to clients. Most of the interventions (32/48, 67%) were delivered to clients who had not already sought help via outreach methods such as social media, marketing, or email. Excluding a study that was unclear, the remaining 15 (31%) interventions were delivered to clients who had already sought help or were investigating help.

Table 3. Types of interventions in the selected studies (N=48).

Category	Studies, n (%)
Intervention format	
Web-based program	22 (46)
Screening tool	7 (15)
Video	6 (13)
App	4 (8)
Website	3 (6)
Automated emails and website	2 (4)
Screening tool and messaging	2 (4)
Advertisement	1 (2)
Advertisement and website	1 (2)
Target treatments	
General professional help	27 (56)
Specific treatments	21 (44)
Specific face-to-face therapy	14 (29)
Specific web-based therapy	6 (13)
Specific phone therapy	1 (2)
Support	
No support	35 (73)
Supported	13 (27)
Asynchronous (clinician)	5 (10)
Synchronous (digital)	4 (8)
Synchronous (clinician)	2 (4)
Asynchronous and synchronous (peer)	1 (2)
Asynchronous and synchronous (clinician)	1 (2)
Use	
Once-off	28 (58)
Repeated	20 (42)
Intervention duration (estimated or average)	
Duration (minutes or hours)	25 (52)
≤30 minutes	14 (29)
31-90 minutes	9 (19)
91 minutes-4.5 hours	2 (4)
Duration (weeks)	16 (33)
1-4	9 (19)
≥4	7 (15)
Duration not specified	12 (25)
Theoretical models[a]	
No model mentioned	16 (33)
Motivational interviewing	16 (33)
Cognitive behavioral therapy	6 (13)
Transtheoretical model	4 (8)
Theory of planned behavior	4 (8)

Category	Studies, n (%)
Intervention delivery	
Outreach (clients had not sought help)	32 (67)
Social media	9 (19)
Clinician or health service referral	8 (17)
Print marketing (flyers or brochures)	6 (13)
Trial panel (eg, Amazon Mechanical Turk)	6 (13)
Email (student email, newsletters, or from the electronic medical record portal)	5 (10)
Digital marketing (web-based advertisements or media)	5 (10)
Postal screening questionnaire	4 (8)
General practitioner waiting room	2 (4)
Events (community events or school workshops)	2 (4)
Before target treatment (clients had already sought help)	12 (25)
Before first use or session	6 (13)
On waiting list for treatment or assessment	3 (6)
During intake	2 (4)
Before intake	1 (2)
Self-selected (clients were interested in help)	3 (6)
Downloaded screening app	1 (2)
Via e-mental health portal	1 (2)
Via referral website for clinic	1 (2)
Unclear	1 (2)

[a] Other models used in only 1 or 2 studies: health belief model, acceptance and commitment therapy, self-determination theory, unified theory of acceptance and use of technology, screening brief intervention and referral to treatment, motivational enhancement therapy, theory of reasoned action, and extended parallel process model.

Intervention Components

The 48 interventions examined in the included studies all comprised several different topics and tools, which we refer to as components. The most prevalent component was general psychoeducation (40/48, 83%), followed by symptom assessment (23/48, 48%) and information on various treatment options (21/48, 44%). Refer Table 4 for a list of the 14 most common components. Other components included in <4 studies were self-monitoring, data security information, personal strengths, therapeutic alliance and roles in therapy, acceptance, imaginative exercises (eg, imagining ideal life or future with or without treatment), MI techniques (eg, importance and readiness rulers), and information about the effectiveness or advantages of a specific target treatment.

Identifying components that showed evidence of effectiveness was difficult owing to the variety of interventions and components covered in this review, as well as the diversity in the outcomes of the experimental studies (see the *Feasibility* section for a closer look at these outcomes). Some studies that compared 2 interventions with different components found no differences among the outcomes of these interventions [64,86]; however, other studies found the opposite (ie, different components in similarly delivered interventions resulted in significantly different outcomes [65,67,90]). In 2 studies aimed at social anxiety [65,90], the addition of components such as the pro-con list, goal setting, values, and planning led to significantly greater engagement with further treatment or help-seeking behaviors than did interventions without these components. Interestingly, another study found that the effectiveness of the components depended on the condition in question; in this case, providing tailored feedback on screening was detrimental when it came to social anxiety but not depression [75].

Table 4. Components used in the included interventions.

Component	Description	Frequency, n (%)
Psychoeducation	Information about condition, symptoms, risks, prevalence, treatment benefits, recovery chances, and myth busting	40 (83)
Assessments	Self-administered assessments of symptoms or behavior	23 (48)
Treatment options	Information about potential treatment options	21 (44)
Assessment feedback	Tailored or generic feedback on assessments; for example, severity relevant to the general population	18 (38)
Referral information	Direct contact information or guidance for further treatment	17 (35)
Testimonials	Videos or written stories from people with similar issues or from those who have been through treatment	16 (33)
Expectation management	Guiding expectations on treatment or help seeking and expectation setting	16 (33)
Pro-con list	Cost-benefit analysis of change, treatment, or help seeking	15 (31)
Coping skills	Cognitive behavioral therapy skills (eg, cognitive restructuring or behavioral activation), relaxation, mindfulness, and emotion regulation	10 (21)
Planning	Planning for change or treatment or planning for overcoming obstacles to change or treatment (implementation intentions)	8 (17)
Goal setting	Personal goals, life goals, and treatment goals	8 (17)
Values	Using values to develop discrepancy between ideal and actual self	5 (10)
Self-efficacy	Building belief in ability to change, self-esteem, and positive self-affirmations	4 (8)
Problem-solving	Identifying problems, brainstorming solutions, and solution planning	4 (8)

Design Processes

Only 18 of the 45 (40%) included papers discussed how the intervention was designed or developed. Of these 18 studies, only 4 (22%) mentioned the design approach: a study used a user-centered design [59], one used a person-based approach [58], another used a participatory design [40], and the final study used a participatory design, ethnography, and co-design [60]. In terms of the design methods used in the development of the interventions, 9 (20%) studies included consultation with experts or input from expert groups [56,58-60,63,77,84,89,91], 5 (11%) studies used expert or user surveys [50,55,58,60,84], 4 (22%) conducted focus groups with users [40,59,60,78], and 3 (7%) conducted interviews with either users or experts [55,59,60]. A total of 2 (4%) studies reported using working groups comprising users with lived experience and experts to cocreate the intervention [60,77], and 2 (4%) used expert-only working groups [56,63]. A total of 12 (27%) studies reported conducting user testing of their interventions, usually with an iterative process of implementing feedback. A study conducted feasibility testing with clinicians [56].

Feasibility

To better understand the effectiveness of the included interventions, we took the controlled studies (24/45, 53%) and charted their outcomes (Table 5). The outcome measures across the studies were diverse and ranged from behavior to intentions and attitudes toward further treatment. Other associated factors such as symptom improvement, mental health literacy, stigma, and acceptance were also used as proxy measures to infer subsequent behavior or action. Controls included treatment as usual, wait-list, no intervention, intervention control, and attention controls. For the attention controls, we distinguished between nonspecific treatment component controls and specific treatment component controls [92].

Of the 16 studies that measured actual behavior or action (eg, engagement with target treatment or help-seeking behavior), 7 (44%) showed evidence of intervention effectiveness compared with controls [63,65,66,74,87,89,90]. However, these results should be considered in the context of other findings in the studies. For example, a study of an MI-based program aimed at preparing clients for web-based CBT found that clients in the intervention group (IG) spent longer time using the target treatment than those in the control group (CG), but their symptoms were actually worse after the treatment [89]. The participants in this study were highly motivated to engage in treatment at screening, which should also be noted along with the results.

A further 7 studies found no differences between controls and interventions in terms of behavior [54,62,64,67,68,78,86]; however, the other results in these studies provide vital qualifying information. For example, Denison-Day et al [54] offered the intervention to clients in the IG but allowed for natural uptake, meaning that only 34% of the IG actually used the intervention. Hence, no differences were found among groups when 98% of those who actually used the intervention engaged in further treatment. The type of control also had a considerable impact on whether the interventions were found to be “effective” (eg, Brunette et al [86] found no differences among groups, but both groups were given interventions, and both had high subsequent use of target treatment). In some studies with no intervention controls, both groups were found to exhibit high adherence to the target treatment [68].

A total of 2 studies indicated worse behavioral outcomes for the IG compared with the CG [75,80]. Again, the control and other results need to be considered; in the study by Krampe et al [80], the “treatment-as-usual” CG received both the digital intervention and MI-based FTF psychotherapy sessions, whereas the IG received the digital intervention alone, and their results showed that the digital intervention was comparable with the FTF control for those with high readiness to change scores [80]. In the study by Batterham et al [75], the IG received tailored feedback after screening based on symptom severity, whereas the CG received generic, untailored feedback. For clients with social anxiety, tailored feedback led to lower treatment use and intentions to seek help rather than generic advice. Study attrition was lower in the IG than in the CG; however, this is another factor to consider along with these results [75].

Considering the other variables measured in these studies, the findings are mixed. Some indicated that the interventions increased help-seeking intentions [72,83], whereas others showed no effect on intentions [64,71,79], despite their effectiveness in improving attitudes toward treatment or motivation to change. Some indicated improved symptoms [62,71,78], whereas others reported reduced stigma or improved literacy [14,52,72,79]. In addition, all the pre-post studies in the review found that their interventions either reduced client symptoms [53,73] or increased client interest in further treatment [51,73], and all the observational studies in the sample indicated positively skewed effects of their interventions on help-seeking actions, behaviors, or intentions [49,57,69,76,77,82,84].

No obvious patterns were observed among intervention format, support level, duration, components (see the *Intervention Components* section), condition, target treatment or intervention delivery, and whether interventions were effective. Several studies that compared interactive and noninteractive interventions suggested that interactivity is important for effectiveness [62,87,90]; however, the opposite result was also found [86].

Table 5. Outcomes of the controlled studies in the sample (standardized measures are abbreviated).

Study	Study design	Control	Sample size, N	Intervention	Measures	Significant outcomes
Olson et al [56]	Historically controlled	TAU[a]	163	Screening tool	Acceptance and quality of physician appointment survey; qualitative physician feedback	IG[b] more likely to discuss alcohol and tobacco use with physician but not mood disorders. IG increased acceptance of subsequent physician appointment
Yoon et al [91]	Historically controlled	TAU	301	Screening tool	Screen for unhealthy drinking behaviors and alcohol use disorders; motivation to change and referral interest survey; acceptance survey	CG[c] used to compare response rate only (responses were comparable). Only 16% of the IG had unhealthy drinking habits. Of these, 14% were interested in further help, and 40% would cut back on their own
Bommelé et al [78]	NRCT[d]	NTCC[e]	757	WP[f]	PO[g]: receptivity to information, motivation to change, self-efficacy and referral interest survey; SO[h]: cigarettes per day and quit attempts	IG more receptive to information than CG after the intervention but not at the 2-week or 2-month follow-up. IG had reduced smoking at all time points. No differences in quit attempts or referral
Brunette et al [66]	NRCT	Wait-list	41	WP	PO: treatment seeking and motivation to change survey (verified by medical records); SO: FTND[i]; 1 item from SCS[j]; ATS[k]	IG more likely to have taken action toward change than CG (eg, attempting to quit, meeting with a clinician to discuss, or start treatment)
Strassle et al [68]	RCT	No intervention	68	Video	PO: return for second session of TT[l]; SO: SCL-90[m]; IIP-32[n]; CASF-P[o]; therapist measures: GAF[p]; CASF-T[q]	No differences between IG and CG in adherence to TT, therapeutic alliance, or TT outcomes (all clients had high adherence to TT)
Ebert et al [14]	RCT	No intervention	128	Video	PO: acceptance survey; SO: expectations, social opinions, internet concerns, help-seeking attitudes, and web-based therapy literacy survey	IG had higher acceptance, expectations, and literacy and lesser internet concerns than CG. No differences in social opinions or help-seeking attitudes
Ebert et al [83]	RCT	No intervention	1374	Screening tool	PO: intention to seek help survey; moderators: CIDIS[r]; AUDIT[s]; CSSR[t]; SITBI[u]; subjective health, lifetime and current treatment use, intention to use mental health services, barriers to treatment use, and readiness to change survey	IG had higher intentions to seek help than CG. Intervention was more effective for those with panic disorder and worse physical health and those who were nonheterosexual. No effect of intervention for those in the action stage of change
Soucy et al [89]	RCT	No intervention	231	WP	PO: CQ[v]; TT lessons accessed; GAD-7[w]; PHQ-9[x]; SO: motivation to engage in TT survey; acceptance survey; K10[y]; SDS[z]	IG spent longer in TT than did CG. IG had higher anxiety and perceived disability at post-TT period than did CG. No differences in motivation or acceptance
Christensen et al [62]	RCT	NTCC	414	2 IGs: W[aa] and WP	CES-D[ab]; help- and treatment-seeking survey	Both W and WP reduced depression symptoms compared with CG. W less likely to seek informal help than CG. WP more likely to use certain evidence-based treatments
Reis and Brown [63]	RCT	NTCC	125	Video	Therapist measure: TSQ[ac]	IG had lower dropout from TT than did CG
Costin et al [64]	RCT	NTCC	348	2 IGs: both automated emails and W	PO: AHSQ[ad]; informal help-seeking survey; SO: GHSQ[ae]; beliefs about help-seeking survey; depression and help-seeking literacy survey; CES-D; acceptance survey	No differences among IGs or between IGs and CG in help-seeking behavior, intentions, literacy, or depression symptoms. IGs had more positive beliefs about formal help than did CG
Johansen et al [67]	RCT	NTCC	105	2 IGs: WA[af] video and EA[ag] video	Acceptance survey; PANAS[ah]; WAI-S[ai] (client and therapist); return for second session of TT	WA had higher negative affect and lower therapist-rated alliance than CG. No difference in client-rated alliance among IGs. No differences in adherence to TT between IGs and CG

Study	Study design	Control	Sample size, N	Intervention	Measures	Significant outcomes
Taylor-Rodgers and Batterham [72]	RCT	NTCC	67	WP	PO: A-Lit[aj]; D-Lit[ak]; LSS[al]; DSS[am]; GASS[an]; SOSS[ao]; ATSPPH-SF[ap]; GHSQ; SO: PHQ-9; GAD-7; acceptance and adherence survey	IG had increased anxiety literacy, help-seeking attitudes and intentions, and reduced depression stigma compared with CG. No differences in symptoms, acceptance, or adherence
Griffiths et al [79]	RCT	NTCC	83	WP	PO: GHSQ; SO: ATSPPH-SF; SA-Lit[aq]; SASS-I[ar]; perceived need for treatment and interest in TT; acceptance survey	IG had higher literacy, perceived need, and positive attitudes toward treatment than did CG. No differences in help-seeking intentions or stigma
King et al [74]	RCT	STCC	76	Screening tool and messaging	Perceived need for help and treatment use survey; 2 items from DDS[as]; readiness to access help survey	IG had higher readiness to access help and use treatment and lower stigma than did CG at the 2-month follow-up
Batterham et al [75]	RCT	STCC	2773	Screening tool	PO: AHSQ; SO: PHQ-9; SOPHS[at] 2 items from GHSQ; AQoL-4D[au]; self-reported days out of role	IG had higher study attrition than did CG. For social anxiety, IG had lower treatment use and intentions to seek help than did CG, no differences found for depression
Peter et al [87]	RCT	STCC	805	2 IGs: screening tools—IM[av] and NM[aw]	PO: choice between BBGS[ax] and 3 items from GBQ[ay]; moderators: gambling history, psychological distress, and treatment interest survey	IM more likely to complete gambling screener than NM or CG
Titov et al [65]	RCT	Intervention control	108	2 IGs: WPs—Education and Education+Motivation	PO: SIAS[az]; SPS[ba]; SO: PHQ-9; K-10, SDS, and CEQ[bb]; literacy and motivation to change survey; time spent, log-ins, and homework downloads of TT	Education+Motivation had higher use of TT than Education. No differences in TT outcomes or acceptability. No differences in motivation to change
Tobias et al [90]	RCT	Intervention control	267	2 IGs: WPs—Education and Education+Motivation	Motivation for individual treatment steps, attitudes toward and intentions to seek treatment, perceived ability to engage in treatment seeking, and treatment use survey; CSQ-8[bc]	Education+Motivation had improved treatment-seeking attitudes and behaviors, compared with Education. Both groups improved on all outcomes
Brunette et al [86]	RCT	Intervention control	162	2 IGs: WPs—IWP[bd] and DEP[be]	PO: treatment use (verified by medical records); SO: expired carbon monoxide; TFB[bf] (quit attempts); PUEUS[bg]	No differences between IWP and DEP in TT use, quit attempts, or abstinence (both groups had high use of TT)
Denison-Day et al [54]	RCT	TAU	313	WP	PO: attendance at initial assessment appointment; SO: use of TT, acceptance, and motivation (interview)	No differences between IG and CG in attendance at initial appointment. Only 34% of the IG used the intervention, and of these, 98% attended the appointment
Krampe et al [80]	RCT	TAU	220	Screening tool	PO: treatment use; SO: URICA[bh]; BSI-GSI[bi]	IG had lower treatment use and worse symptoms than CG. IG and CG were comparable for those with high readiness to change scores
Keller et al [52]	RCT	Wait-list	320	3 IGs: videos—7 minutes, 13 minutes, and 17 minutes	SSOSH[bj]; stigma survey	Only the 17-minute IG reduced stigma compared with CG

Study	Study design	Control	Sample size, N	Intervention	Measures	Significant outcomes
Hötzel et al [71]	RCT	Wait-list	212	WP	PO: SOCQ-ED[bk]; SO: P-CED[bl]; SES[bm]; RSES[bn]; EDE-Q[bo]	IG had higher motivation to change, self-esteem, and symptom improvement than CG. No differences in motivation to begin treatment

[a] TAU: treatment as usual.

[b] IG: intervention group.

[c] CG: control group.

[d] NRCT: nonrandomized controlled trial.

[e] NTCC: nonspecific treatment component controls.

[f] WP: web-based program.

[g] PO: primary outcomes.

[h] SO: secondary outcomes.

[i] FTND: Fagerström test for nicotine dependence.

[j] SCS: Stage of Change Scale.

[k] ATS: Attitudes Toward Smoking Scale

[l] TT: target treatment.

[m] SCL-90: Symptom Checklist-90-Revised.

[n] IIP-32: Inventory of Interpersonal problems-32.

[o] CASF-P: Combined Alliance Short Form-Patient version.

[p] GAF: Global Assessment of Functioning Scale.

[q] CASF-T: Combined Alliance Short Form-Therapist version.

[r] CIDIS: Composite International Diagnostic Interview Screening Scales.

[s] AUDIT: Alcohol Use Disorders Identification Test.

[t] CSSR: Columbia Suicidal Severity Rating Scale.

[u] SITBI: Self Injurious Thoughts and Behaviors Interview.

[v] CQ: Change Questionnaire.

[w] GAD-7: Generalized Anxiety Disorder 7-item.

[x] PHQ-9: Patient Health Questionnaire 9-item.

[y] K-10: Kessler 10-item.

[z] SDS: Sheehan Disability Scales.

[aa] W: website.

[ab] CES-D: Centre for Epidemiological Studies Depression Scale.

[ac] TSQ: Termination Status Questionnaire.

[ad] AHSQ: Actual Help Seeking Questionnaire.

[ae] GHSQ: General Help Seeking Questionnaire.

[af] WA: working alliance.

[ag] EA: experimental acceptance.

[ah] PANAS: Positive and Negative Affect Schedule.

[ai] WAI-S: Working Alliance Inventory.

[aj] A-Lit: Anxiety Literacy Scale.

[ak] D-Lit: Depression Literacy Scale.

[al] LSS: Literacy of Suicide Scale.

[am] DSS: Depression Stigma Scale.

[an] GASS: Generalised Anxiety Stigma Scale.

[ao] SOSS: Stigma of Suicide Scale short form.

[ap] ATSPPH-SF: Attitudes Toward Seeking Professional Help Short Form Scale.

[aq] SA-Lit: Social Anxiety Literacy Questionnaire.

[ar] SASS-I: Social Anxiety Stigma Scale.

[as] DDS: Discrimination-Devaluation Scale.

[at] SOPHS: Social Phobia Screener.

[au] AQoL-4D: Assessment of Quality of Life.

[av] IM: interactive message.

[aw] NM: noninteractive message.

[ax] BBGS: Brief Biosocial Gambling Screen.

[ay] GBQ: Gamblers' Beliefs Questionnaire.

[az] SIAS: Social Interaction Anxiety Scale.

[ba] SPS: Social Phobia Scale.

[bb] CEQ: Credibility/Expectancy Questionnaire.

[bc] CSQ-8: Client Satisfaction Questionnaire.

[bd] IWP: interactive web-based program.

[be] DEP: digital education pamphlet.

[bf] TFB: Timeline Follow-Back method.

[bg] PUEUS: Perceived Usefulness and Ease of Use Scale.

[bh] URICA: University of Rhode Island Change Assessment.

[bi] BSI-GSI: Global Severity Index of the Brief Symptom Inventory.

[bj] SSOSH: Self-Stigma of Seeking Help Scale.

[bk] SOCQ-ED: Stages of Change Questionnaire for Eating Disorders.

[bl] P-CED: Pros and Cons of Eating Disorders Scale.

[bm] SES: Self-Efficacy Scale.

[bn] RSES: Rosenberg Self-Esteem Scale.

[bo] EDE-Q: eating disorder symptomatology.

Discussion

Principal Findings

This scoping review explores digital interventions to enhance readiness for psychological therapy. These interventions are delivered most often as unsupported web-based programs designed for once-off use that takes <90 minutes. They are used to prepare clients for specific therapies or, more generally, to enhance readiness for professional treatment; they are provided to clients either via outreach methods for those who have not sought help, or they are inserted into the care pathway before the main treatment for those who have already reached out. Thus, these interventions appear to cater to clients across multiple stages of change, from those in precontemplation, who are not yet aware that they need help, to those in the preparation stage, who are taking initial steps toward change.

What is the most apparent from this review is the substantial variation not only in the types of digital readiness interventions that have been used but also in their development, delivery, and evaluation. When it comes to the feasibility of digitally delivering interventions, the included studies indicate that there is potential in this area. The current state of the literature, however, does not yet support the possibility of determining which components or types of interventions are effective or not effective; this is a complex undertaking with multiple factors to consider. For example, in some contexts, interactivity appears to be an important aspect of these interventions, which makes sense when considering the conversational nature of traditional FTF MI. However, many simple, noninteractive interventions such as videos and advertisements were also effective at improving variables related to further treatment seeking or engagement. Despite the variability among the studies included in this review, several common topics emerged: *tailoring to the stage of change*, *intervention pathways*, *risk*, and *evaluation*.

Tailoring to the Stage of Change

The existing literature indicates the effectiveness of tailoring psychosocial interventions to clients' stages of change [29]. Several studies in our review involved tailoring to the stage of change [58,69,73,84]. In 2 studies, tailoring involved a simple 2-way split, with different content for those who were interested in further treatment and those not interested [69,73]. In one of these studies, clients who were not yet interested in further treatment were given CBT coping techniques as a way to show them how treatment works and how effective it can be, rather than simply telling them this [73]. When clients are highly motivated, tailored interventions tend to focus on the practical aspects of engaging with further treatment (ie, choosing the right treatment, setting expectations, and planning).

The effective identification of a client's stage of change is a significant aspect of tailoring. This can be done by asking simple binary questions, such as those in the aforementioned studies (eg, Are you interested in treatment?) or more formally with readiness measures such as the General Help Seeking Questionnaire [93], Stage of Change Scale [94], or University of Rhode Island Change Assessment [95]. One interesting website intervention used the stages of change to frame the headings of the main website navigation (ie, "Do I have a problem?" "Should I get help?" "I want and need help?" "I have tried to get help"), giving the client agency in self-selecting their own stage and thus controlling and tailoring their own journey [84]. Outside the stages of change, information-based interventions can be tailored to the client's personal circumstances and needs at a broader level. For example, Dreier et al [50] provided suicide stigma interventions that were modified depending on whether clients had a suicide attempt in the past, had suicidal thoughts, had lost a close person by suicide, were fearing the loss of a close person by suicide, or were interested in the topic in general.

In all, 2 studies in this review illustrate the importance of effective stage identification and tailoring, with findings indicating negative or no effects of their inventions on those who already had high motivation or intentions to seek help [83,89]. Previous research also demonstrated that FTF MI is most beneficial for those who are not already motivated or engaged in treatment [96]. However, in contrast to this, Krampe et al [80] found that their brief digital intervention was as effective as FTF MI–based psychotherapy but only for clients who were already motivated. Tailored digital readiness interventions have the potential to bridge the divide between client and treatment, providing light-touch interactions for those who are already motivated as well as more detailed programs for those in earlier stages of change. For clients, these interventions could serve as stepping stones between information gathering and formal treatment, with layered interactions that support individuals on their journey through the stages of change [77].

Intervention Pathways

The implementation of digital readiness interventions involves both onboarding (ie, the uptake of the intervention itself) and off-boarding (ie, the link between the intervention and further treatment). In terms of onboarding, the first point of contact and framing of digital readiness interventions are crucial, as uptake issues can drastically impact their effectiveness in the real world. Denison-Day et al [54] found that although their intervention was highly effective for those who used it, only 34% of the IG actually used it (they offered participants the intervention but allowed for natural uptake). They noted that simply offering new interventions to address the problem of target treatment engagement may not be enough, and approaches focused on low engagement may need to be considered even earlier in the treatment pathway. An interesting aspect of their intervention (further detailed in the development process paper by Muir et al [58]) was that instead of aiming to prepare clients for the full extent of treatment, they framed it as preparation for the initial assessment appointment only. This removed some of the overwhelming aspects of thinking about full “recovery” and instead allowed clients to take their treatment journey 1 step at a time [58]. How the first step on a client’s journey is presented and by whom could have an impact on the client’s subsequent progress toward change.

Several studies included in this review were conducted in health care settings, where client trust has already been established [59,80,91]. Embedding readiness interventions within existing pathways, such as routine screening, general practitioner waiting rooms, or treatment waiting lists, can draw on this trust and help the client gain direct access to appropriate services. Regarding off-boarding, many studies in this review noted that access to the target treatment needs to be provided in a timely manner following the readiness intervention, as motivation wanes over time [66,74,78,82,86,90]. Moessner et al [76] included clinician monitoring of client deterioration as part of their intervention, allowing clients to take their time to become ready for treatment, while still being supported. In the intervention developed by Brown et al [55], the first session of the target treatment immediately followed the readiness intervention (if the client wanted to proceed), making the most of their motivation and removing any lag time between the interventions. Where digital readiness interventions fit within the wider context of client pathways appears to be an important consideration for both their development and evaluation.

Risk

An important aspect that surfaced while reviewing these studies was the potential risk of readiness interventions impairing treatment engagement, reducing help seeking, worsening symptoms, and increasing self-stigma. Batterham et al [75] found that tailored feedback on screening reduced help seeking for individuals with social anxiety compared with a control that was just generic information; the directive nature of this feedback may have come across as particularly confrontational to clients experiencing difficult emotions centered on their interactions with others. Similarly, Johansen et al [67] found that a video providing information on the working alliance between the client and therapist led to more negative emotions for the client and no improvement in working alliance ratings. Information designed simply to “prepare” clients for what is to come can potentially lead to negative emotions and apprehension, which can in turn affect readiness for treatment.

Stigma adds another layer of complexity to the help-seeking and treatment readiness process; Keller et al [52] found that informational videos on suicide prevention increased empathy, while simultaneously decreasing help seeking. Previous research shows that different types of stigma (eg, public stigma vs internalized self-stigma [97]) affect help seeking in different ways [98]. How we interpret the experiences and emotions of other people is distinct from how we perceive our own internal states. When addressing stigma with a digital readiness intervention, care should be taken as to which types of stigma are being targeted and the intricate relationships among them. Furthermore, stigma is not only complex, layered, and subjective, but even the act of measuring it can reproduce or reinforce stigmatizing attitudes [50]. Individual emotional responses to engaging in treatment for mental health difficulties are sensitive and differ from person to person; a delicate, cautious approach is clearly needed when developing and implementing readiness interventions.

Evaluation

The final discussion concerns the evaluation of readiness interventions and issues when conducting research in such a sensitive area. Several studies in this review found that clients in the control arms improved as much as those in the intervention arms [64,68,86]. Considering the large battery of measures used in several studies and the fact that screening was a core component in many of the included interventions, it is difficult to separate the effects of the interventions themselves from the overall effects of trial participation. Although this is often the case with research trials, the specific light-touch, preparatory nature of these interventions makes them more susceptible to this reactivity. Perhaps, in many cases, being included in a trial focused on help seeking constitutes a readiness intervention in itself.

In addition, the trial design has a significant influence on the “effectiveness” of a given intervention. Constrained processes that force engagement with an intervention may provide rigor in intervention effects but have little ecological validity. There is also potentially greater baseline motivation among people who are prepared to participate in clinical trials than among the general population [65]. The real-world uptake of digital readiness interventions is key to their effectiveness. Naturalistic studies could therefore be a more useful method of understanding how these interventions would function in practice.

Another aspect of evaluation involves the chosen research methodology, which not only has a fundamental impact on the outcomes of the study but also on how we come to understand complex social constructs such as stigma, motivation, and the stages of change. Using quantitative measures to isolate and examine phenomena such as attitudes and emotions is limited because these experiences are highly subjective and contextual [52]; we miss vital information when we detach these occurrences from what gives them meaning. Considering that only one-fourth of the studies in this review included qualitative data collection, there exists a significant gap in our understanding of the nuances of this process at the individual level. Furthermore, many of the studies in this review used proxy measures, such as intentions and attitudes, to infer potential future action although the attitude-intention-behavior models that underpin these inferences have been contested in research across several fields [99-101]. This review suggests that the measurement of digital readiness interventions requires careful consideration because of the many intricacies involved.

Limitations

There are several limitations to this study. First, we did not include help seeking as a search term (we decided to focus our search on the more general areas of readiness, preparation, and motivation); therefore, our coverage of help-seeking interventions was not comprehensive. Furthermore, our digital-only inclusion criteria excluded some interesting interventions that could easily be reproduced digitally (eg, a postal survey on implementation intentions [102] and an educational handout about the dose-effect relationship of therapy and expectation setting around treatment length [103]).

Implications for Research

Given the inconclusive nature of findings presented here, further research is needed to enhance our knowledge and shape the field of digital readiness interventions for psychological therapy. In-depth qualitative research is crucial to understanding individual differences in emotional responses to readiness interventions and how constructs such as self-stigma affect motivation. Longitudinal research could also provide insights into individual trajectories through the stages of change because the process of becoming ready for treatment can be a long-term one, involving many layers and influences [76]. Recent phenomenological research indicates that change is perhaps a more continuous, internal, and holistic process than the TTM allows [104], and therefore, mapping the process of change in relation to readiness for mental health treatment would add depth to our theoretical foundations. Naturalistic effectiveness studies that attempt to reduce confounding trial effects and examine intervention implementation would help us to ground our knowledge in ecologically valid data and thus improve the practical application of digital readiness interventions. In addition, few studies in this review reported on how the interventions in question were developed or the design strategies used; this is important information for advancing the field and building best practices for future development. Finally, to further understand the different types of readiness interventions being used, future reviews could use more specific search terms (eg, help seeking, screening, and wait-list) to explore these areas in more detail. They could also include quality assessments in their charting process; however, the methodological issues discussed earlier would need to be further unraveled to enable a useful discussion of quality.

Conclusions

Digital interventions to enhance readiness for psychological therapy are broad and varied. The interventions in question range from brief, simple videos and advertisements to supported web-based programs. They are used to help clients across multiple stages of change, from those in precontemplation who have not yet sought help to those already preparing to take action. Although these easily accessible digital approaches show potential as a means of preparing people for therapy and thus reducing the mental health treatment gap, they are not without risks. The complex nature of stigma, motivation, and individual emotional responses toward engaging in treatment for mental health difficulties suggests that a careful approach is needed when developing and measuring readiness interventions. The results of this review indicate that the implementation and uptake of these interventions are important elements to consider in design, delivery, and measurement and that further qualitative and longitudinal research is needed to deepen our knowledge of the process of change in relation to readiness for therapy. Overall, this review highlights the fact that the field of digital readiness interventions is an emerging one, and more research is needed in this area.

Acknowledgments

The work of JJ was supported by SilverCloud Health and the Irish Research Council (grant EBPPG/2020/53). The work of RB is supported by Microsoft Research through its PhD Scholarship Programme and the Science Foundation Ireland Centre for Research Training in Digitally Enhanced Reality (d-real; grant 18/CRT/6224). The work of GD was supported by Science Foundation Ireland (ADAPT, grant 13/RC/2106_P2 and LERO, grant 13/RC/2094_P2).

Conflicts of Interest

JJ is partly funded by SilverCloud Health, a company that develops digital mental health interventions.

Multimedia Appendix 1

PRISMA-Scr checklist. PRISMA-Scr: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews.

[PDF File (Adobe PDF File), 244 KB - jmir_v24i8e37851_app1.pdf]

References

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
2. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. J Clin Psychiatry 2013 Jun;74(6):595-602 [FREE Full text] [doi: 10.4088/JCP.12r07757] [Medline: 23842011]
3. Cuijpers P, van Straten A, van Oppen P, Andersson G. Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. J Clin Psychiatry 2008 Nov;69(11):1675-1841. [doi: 10.4088/jcp.v69n1102] [Medline: 18945396]
4. Timulak L. Identifying core categories of client-identified impact of helpful events in psychotherapy: a qualitative meta-analysis. Psychother Res 2007 May;17(3):305-314. [doi: 10.1080/10503300600608116]
5. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
6. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
7. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
8. Cooper AA, Conklin LR. Dropout from individual psychotherapy for major depression: a meta-analysis of randomized clinical trials. Clin Psychol Rev 2015 Aug;40:57-65. [doi: 10.1016/j.cpr.2015.05.001] [Medline: 26067572]
9. Doherty K, Doherty G. Engagement in HCI: conception, theory and measurement. ACM Comput Surv 2019 Sep 30;51(5):1-39. [doi: 10.1145/3234149]
10. Taylor CB, Ruzek JI, Fitzsimmons-Craft EE, Sadeh-Sharvit S, Topooco N, Weissman RS, et al. Using digital technology to reduce the prevalence of mental health disorders in populations: time for a new approach. J Med Internet Res 2020 Jul 24;22(7):e17493 [FREE Full text] [doi: 10.2196/17493] [Medline: 32706665]
11. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an Internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2019 Aug 01;21(8):e12775 [FREE Full text] [doi: 10.2196/12775] [Medline: 31373272]
12. Nordmo M, Monsen JT, Høglend PA, Solbakken OA. Investigating the dose-response effect in open-ended psychotherapy. Psychother Res 2021 Sep;31(7):859-869. [doi: 10.1080/10503307.2020.1861359] [Medline: 33331244]
13. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a Web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013 Aug 20;15(8):e172 [FREE Full text] [doi: 10.2196/jmir.2258] [Medline: 23963284]
14. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of Internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 01;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
15. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
16. Gersh E, Hallford DJ, Rice SM, Kazantzis N, Gersh H, Gersh B, et al. Systematic review and meta-analysis of dropout rates in individual psychotherapy for generalized anxiety disorder. J Anxiety Disord 2017 Dec;52:25-33. [doi: 10.1016/j.janxdis.2017.10.001] [Medline: 29028610]
17. Prochaska JO, Norcross JC, Saul SF. Generating psychotherapy breakthroughs: transtheoretical strategies from population health psychology. Am Psychol 2020 Oct;75(7):996-1010. [doi: 10.1037/amp0000568] [Medline: 31763861]
18. Thornicroft G. Most people with mental illness are not treated. Lancet 2007 Sep 08;370(9590):807-808. [doi: 10.1016/S0140-6736(07)61392-0] [Medline: 17826153]
19. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann Behav Med 2006 Dec;32(3):254-258. [doi: 10.1207/s15324796abm3203_12] [Medline: 17107299]
20. Gulliver A, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Banfield M, et al. Consumer-guided development of an engagement-facilitation intervention for increasing uptake and adherence for self-guided Web-based mental health programs: focus groups and online evaluation survey. JMIR Form Res 2020 Oct 29;4(10):e22528 [FREE Full text] [doi: 10.2196/22528] [Medline: 33118939]

21. Moskalenko MY, Hadjistavropoulos HD, Katapally TR. The complex association of barriers and interest in internet-delivered cognitive behavior therapy for depression and anxiety: informing e-health policies through exploratory path analysis. J Ment Health (forthcoming) 2020 Jul 25:1-10. [doi: 10.1080/09638237.2020.1793125] [Medline: 32715841]
22. 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]
23. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
24. Prins MA, Verhaak PF, Bensing JM, van der Meer K. Health beliefs and perceived need for mental health care of anxiety and depression--the patients' perspective explored. Clin Psychol Rev 2008 Jul;28(6):1038-1058. [doi: 10.1016/j.cpr.2008.02.009] [Medline: 18420323]
25. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
26. Marker I, Norton PJ. The efficacy of incorporating motivational interviewing to cognitive behavior therapy for anxiety disorders: a review and meta-analysis. Clin Psychol Rev 2018 Jun;62:1-10. [doi: 10.1016/j.cpr.2018.04.004] [Medline: 29727863]
27. Walitzer KS, Dermen KH, Connors GJ. Strategies for preparing clients for treatment. A review. Behav Modif 1999 Jan;23(1):129-151. [doi: 10.1177/0145445599231006] [Medline: 9926524]
28. Prochaska JO, Redding CA, Evers KE. The transtheoretical model and stages of change. In: Glanz K, Rimer BK, Viswanath KV, editors. Health Behavior: Theory, Research, and Practice. 5th edition. San Francisco, CA, USA: Jossey-Bass; 2015:125-148.
29. Krebs P, Norcross JC, Nicholson JM, Prochaska JO. Stages of change and psychotherapy outcomes: a review and meta-analysis. J Clin Psychol 2018 Nov;74(11):1964-1979. [doi: 10.1002/jclp.22683] [Medline: 30335193]
30. Norcross JC, Krebs PM, Prochaska JO. Stages of change. J Clin Psychol 2011 Feb;67(2):143-154. [doi: 10.1002/jclp.20758] [Medline: 21157930]
31. Jardine J, Earley C, Richards D, Timulak L, Palacios JE, Duffy D, et al. The experience of guided online therapy: a longitudinal, qualitative analysis of client feedback in a naturalistic RCT. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI, USA p. 1-15. [doi: 10.1145/3313831.3376254]
32. Westra HA, Dozois DJ. Preparing clients for cognitive behavioral therapy: a randomized pilot study of motivational interviewing for anxiety. Cogn Ther Res 2006 Aug 11;30(4):481-498. [doi: 10.1007/s10608-006-9016-y]
33. Lewis CC, Simons AD, Kim HK. The role of early symptom trajectories and pretreatment variables in predicting treatment response to cognitive behavioral therapy. J Consult Clin Psychol 2012 Aug;80(4):525-534. [doi: 10.1037/a0029131] [Medline: 22730951]
34. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 02;2:97-105 [FREE Full text] [Medline: 19920949]
35. Romano M, Peters L. Evaluating the mechanisms of change in motivational interviewing in the treatment of mental health problems: a review and meta-analysis. Clin Psychol Rev 2015 Jun;38:1-12. [doi: 10.1016/j.cpr.2015.02.008] [Medline: 25779459]
36. Miller WR, Rollnick S. Talking oneself into change: motivational interviewing, stages of change, and therapeutic process. J Cogn Psychother 2004 Oct 01;18(4):299-308. [doi: 10.1891/jcop.18.4.299.64003]
37. Shinitzky HE, Kub J. The art of motivating behavior change: the use of motivational interviewing to promote health. Public Health Nurs 2001;18(3):178-185. [doi: 10.1046/j.1525-1446.2001.00178.x] [Medline: 11359619]
38. Shingleton RM, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: a systematic review of the current research. Patient Educ Couns 2016 Jan;99(1):17-35 [FREE Full text] [doi: 10.1016/j.pec.2015.08.005] [Medline: 26298219]
39. Magill M, Gaume J, Apodaca TR, Walthers J, Mastroleo NR, Borsari B, et al. The technical hypothesis of motivational interviewing: a meta-analysis of MI's key causal model. J Consult Clin Psychol 2014 Dec;82(6):973-983 [FREE Full text] [doi: 10.1037/a0036833] [Medline: 24841862]
40. Batterham PJ, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Gulliver A. A brief intervention to increase uptake and adherence of an online program for depression and anxiety: protocol for the Enhancing Engagement with Psychosocial Interventions (EEPI) randomized controlled trial. Contemp Clin Trials 2019 Mar;78:107-115. [doi: 10.1016/j.cct.2019.01.015] [Medline: 30711664]
41. Doherty G, Coyle D, Matthews M. Design and evaluation guidelines for mental health technologies. Interact Comput 2010 Jul;22(4):243-252. [doi: 10.1016/j.intcom.2010.02.006]
42. Barry M, Doherty K, Bellisario JM, Car J, Morrison C, Doherty G. mHealth for maternal mental health: everyday wisdom in ethical design. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO, USA p. 2708-2756. [doi: 10.1145/3025453.3025918]

43. Jardine J, Doherty G, Bowman R. Digital interventions to enhance readiness for psychological treatment: a scoping review. Open Science Framework. 2021 Mar 26. URL: https://osf.io/qzkam [accessed 2022-02-20]
44. 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]
45. Smoktunowicz E, Barak A, Andersson G, Banos RM, Berger T, Botella C, et al. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. Internet Interv 2020 Sep;21:100331 [FREE Full text] [doi: 10.1016/j.invent.2020.100331] [Medline: 32577404]
46. 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]
47. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019 Jun 21;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]
48. Liverpool S, Mota CP, Sales CM, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
49. BinDhim NF, Alanazi EM, Aljadhey H, Basyouni MH, Kowalski SR, Pont LG, et al. Does a mobile phone depression-screening app motivate mobile phone users with high depressive symptoms to seek a health care professional's help? J Med Internet Res 2016 Jun 27;18(6):e156 [FREE Full text] [doi: 10.2196/jmir.5726] [Medline: 27349441]
50. Dreier M, Ludwig J, Härter M, von dem Knesebeck O, Baumgardt J, Bock T, et al. Development and evaluation of e-mental health interventions to reduce stigmatization of suicidality - a study protocol. BMC Psychiatry 2019 May 17;19(1):152 [FREE Full text] [doi: 10.1186/s12888-019-2137-0] [Medline: 31101103]
51. Beck CD, Soucy JN, Hadjistavropoulos HD. Mixed-method evaluation of an online motivational intervention as a pre-treatment to internet-delivered cognitive behaviour therapy: immediate benefits and user feedback. Internet Interv 2020 Apr;20:100311 [FREE Full text] [doi: 10.1016/j.invent.2020.100311] [Medline: 32154118]
52. Keller S, McNeill V, Tran T. The perceived stigma reduction expressed by young adults in response to suicide prevention videos. Int J Environ Res Public Health 2021 Jun 08;18(12):6180 [FREE Full text] [doi: 10.3390/ijerph18126180] [Medline: 34200986]
53. Duffy D, Enrique A, Connell S, Connolly C, Richards D. Internet-delivered cognitive behavior therapy as a prequel to face-to-face therapy for depression and anxiety: a naturalistic observation. Front Psychiatry 2019 Jan 9;10:902 [FREE Full text] [doi: 10.3389/fpsyt.2019.00902] [Medline: 31998149]
54. Denison-Day J, Muir S, Newell C, Appleton KM. A Web-based intervention (MOtivate) to increase attendance at an eating disorder service assessment appointment: zelen randomized controlled trial. J Med Internet Res 2019 Feb 27;21(2):e11874 [FREE Full text] [doi: 10.2196/11874] [Medline: 30810533]
55. Brown RA, Hecht J, Bloom EL, Minami H, Kahler CW, Abrantes AM, et al. Development and preliminary pilot evaluation of a brief tablet computer intervention to motivate tobacco quitline use among smokers in substance use treatment. Am J Addict 2017 Sep;26(6):587-594 [FREE Full text] [doi: 10.1111/ajad.12559] [Medline: 28800182]
56. Olson AL, Gaffney CA, Hedberg VA, Gladstone GR. Use of inexpensive technology to enhance adolescent health screening and counseling. Arch Pediatr Adolesc Med 2009 Feb;163(2):172-177. [doi: 10.1001/archpediatrics.2008.533] [Medline: 19188650]
57. Haas A, Koestner B, Rosenberg J, Moore D, Garlow SJ, Sedway J, et al. An interactive web-based method of outreach to college students at risk for suicide. J Am Coll Health 2008;57(1):15-22. [doi: 10.3200/JACH.57.1.15-22] [Medline: 18682341]
58. Muir S, Newell C, Griffiths J, Walker K, Hooper H, Thomas S, et al. MotivATE: a pretreatment Web-based program to improve attendance at UK outpatient services among adults with eating disorders. JMIR Res Protoc 2017 Jul 26;6(7):e146 [FREE Full text] [doi: 10.2196/resprot.7440] [Medline: 28747295]
59. Dannenberg MD, Bienvenida JC, Bruce ML, Nguyen T, Hinn M, Matthews J, et al. End-user views of an electronic encounter decision aid linked to routine depression screening. Patient Educ Couns 2019 Mar;102(3):555-563. [doi: 10.1016/j.pec.2018.10.002] [Medline: 30497800]
60. Liu Z, Warren J, Christie G. Design of a technology-delivered motivational interviewing intervention for youth empowerment. In: Proceedings of the Australasian Computer Science Week Multiconference. 2018 Presented at: ACSW '18; January 29-February 2, 2018; Brisbane, Australia p. 1-10. [doi: 10.1145/3167918.3167934]
61. Johansen AB, Helland PF, Wennesland DK, Henden E, Brendryen H. Exploring online problem gamblers' motivation to change. Addict Behav Rep 2019 Dec;10:100187 [FREE Full text] [doi: 10.1016/j.abrep.2019.100187] [Medline: 31297434]
62. Christensen H, Leach LS, Barney L, Mackinnon AJ, Griffiths KM. The effect of web based depression interventions on self reported help seeking: randomised controlled trial [ISRCTN77824516]. BMC Psychiatry 2006 Apr 05;6:13 [FREE Full text] [doi: 10.1186/1471-244X-6-13] [Medline: 16595018]

63. Reis BF, Brown LG. Preventing therapy dropout in the real world: the clinical utility of videotape preparation and client estimate of treatment duration. Prof Psychol Res Pr 2006;37(3):311-316. [doi: 10.1037/0735-7028.37.3.311]
64. Costin DL, Mackinnon AJ, Griffiths KM, Batterham PJ, Bennett AJ, Bennett K, et al. Health e-cards as a means of encouraging help seeking for depression among young adults: randomized controlled trial. J Med Internet Res 2009 Oct 22;11(4):e42 [FREE Full text] [doi: 10.2196/jmir.1294] [Medline: 19850549]
65. Titov N, Andrews G, Schwencke G, Robinson E, Peters L, Spence J. Randomized controlled trial of Internet cognitive behavioural treatment for social phobia with and without motivational enhancement strategies. Aust N Z J Psychiatry 2010 Oct;44(10):938-945. [doi: 10.3109/00048674.2010.493859] [Medline: 20932208]
66. Brunette MF, Ferron JC, McHugo GJ, Davis KE, Devitt TS, Wilkness SM, et al. An electronic decision support system to motivate people with severe mental illnesses to quit smoking. Psychiatr Serv 2011 Apr;62(4):360-366. [doi: 10.1176/ps.62.4.pss6204_0360] [Medline: 21459986]
67. Johansen AB, Lumley M, Cano A. Effects of video-based therapy preparation targeting experiential acceptance or the therapeutic alliance. Psychotherapy (Chic) 2011 Jun;48(2):163-169. [doi: 10.1037/a0022422] [Medline: 21639659]
68. Strassle CG, Borckardt JJ, Handler L, Nash M. Video-tape role induction for psychotherapy: moving forward. Psychotherapy (Chic) 2011 Jun;48(2):170-178. [doi: 10.1037/a0022702] [Medline: 21639660]
69. Ferron JC, Brunette MF, McGurk SR, Xie H, Frounfelker R, Cook JA, et al. Do symptoms and cognitive problems affect the use and efficacy of a Web-based decision support system for smokers with serious mental illness? J Dual Diagn 2012 Nov;8(4):315-325. [doi: 10.1080/15504263.2012.723316]
70. Reins JA, Ebert DD, Lehr D, Riper H, Cuijpers P, Berking M. Internet-based treatment of major depression for patients on a waiting list for inpatient psychotherapy: protocol for a multi-centre randomised controlled trial. BMC Psychiatry 2013 Nov 26;13:318 [FREE Full text] [doi: 10.1186/1471-244X-13-318] [Medline: 24279841]
71. Hötzel K, von Brachel R, Schmidt U, Rieger E, Kosfelder J, Hechler T, et al. An Internet-based program to enhance motivation to change in females with symptoms of an eating disorder: a randomized controlled trial. Psychol Med 2014 Jul;44(9):1947-1963. [doi: 10.1017/S0033291713002481] [Medline: 24128818]
72. Taylor-Rodgers E, Batterham PJ. Evaluation of an online psychoeducation intervention to promote mental health help seeking attitudes and intentions among young adults: randomised controlled trial. J Affect Disord 2014 Oct;168:65-71. [doi: 10.1016/j.jad.2014.06.047] [Medline: 25038293]
73. Ahmedani BK, Crotty N, Abdulhak MM, Ondersma SJ. Pilot feasibility study of a brief, tailored mobile health intervention for depression among patients with chronic pain. Behav Med 2015;41(1):25-32. [doi: 10.1080/08964289.2013.867827] [Medline: 24313728]
74. King CA, Eisenberg D, Zheng K, Czyz E, Kramer A, Horwitz A, et al. Online suicide risk screening and intervention with college students: a pilot randomized controlled trial. J Consult Clin Psychol 2015 Jun;83(3):630-636 [FREE Full text] [doi: 10.1037/a0038805] [Medline: 25688811]
75. Batterham PJ, Calear AL, Sunderland M, Carragher N, Brewer JL. Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial. BJPsych Open 2016 Jan;2(1):67-73 [FREE Full text] [doi: 10.1192/bjpo.bp.115.001552] [Medline: 27703756]
76. Moessner M, Minarik C, Özer F, Bauer S. Can an internet-based program for the prevention and early intervention in eating disorders facilitate access to conventional professional healthcare? J Ment Health 2016 Oct;25(5):441-447. [doi: 10.3109/09638237.2016.1139064] [Medline: 26850624]
77. Birnbaum ML, Garrett C, Baumel A, Scovel M, Rizvi AF, Muscat W, et al. Using digital media advertising in early psychosis intervention. Psychiatr Serv 2017 Nov 01;68(11):1144-1149. [doi: 10.1176/appi.ps.201600571] [Medline: 28712355]
78. Bommelé J, Schoenmakers TM, Kleinjan M, Peters GJ, Dijkstra A, van de Mheen D. Targeting hardcore smokers: the effects of an online tailored intervention, based on motivational interviewing techniques. Br J Health Psychol 2017 Sep;22(3):644-660. [doi: 10.1111/bjhp.12256] [Medline: 28703470]
79. Griffiths KM, Walker J, Batterham PJ. Help seeking for social anxiety: a pilot randomised controlled trial. Digit Health 2017 Jun 2;3:2055207617712047 [FREE Full text] [doi: 10.1177/2055207617712047] [Medline: 29942603]
80. Krampe H, Salz A, Kerper LF, Krannich A, Schnell T, Wernecke KD, et al. Readiness to change and therapy outcomes of an innovative psychotherapy program for surgical patients: results from a randomized controlled trial. BMC Psychiatry 2017 Dec 29;17(1):417 [FREE Full text] [doi: 10.1186/s12888-017-1579-5] [Medline: 29284443]
81. Metz MJ, Elfeddali I, Krol DG, Veerbeek MA, de Beurs E, Beekman AT, et al. A digital intake approach in specialized mental health care: study protocol of a cluster randomised controlled trial. BMC Psychiatry 2017 Mar 07;17(1):86 [FREE Full text] [doi: 10.1186/s12888-017-1247-9] [Medline: 28270129]
82. Suka M, Yamauchi T, Yanagisawa H. Comparing responses to differently framed and formatted persuasive messages to encourage help-seeking for depression in Japanese adults: a cross-sectional study with 2-month follow-up. BMJ Open 2018 Nov 12;8(11):e020823 [FREE Full text] [doi: 10.1136/bmjopen-2017-020823] [Medline: 30420341]
83. Ebert DD, Franke M, Kählke F, Küchler AM, Bruffaerts R, Mortier P, WHO World Mental Health - International College Student collaborators. Increasing intentions to use mental health services among university students. Results of a pilot randomized controlled trial within the World Health Organization's World Mental Health International College Student Initiative. Int J Methods Psychiatr Res 2019 Jun;28(2):e1754 [FREE Full text] [doi: 10.1002/mpr.1754] [Medline: 30456814]

84. McLean SA, Caldwell B, Roberton M. Reach Out and Recover: intentions to seek treatment in individuals using online support for eating disorders. Int J Eat Disord 2019 Oct;52(10):1137-1149. [doi: 10.1002/eat.23133] [Medline: 31298791]
85. Shand F, Mackinnon A, O'Moore K, Ridani R, Reda B, Hoy M, et al. The iBobbly Aboriginal and Torres Strait Islander app project: study protocol for a randomised controlled trial. Trials 2019 Apr 05;20(1):198 [FREE Full text] [doi: 10.1186/s13063-019-3262-2] [Medline: 30953556]
86. Brunette MF, Ferron JC, McGurk SR, Williams JM, Harrington A, Devitt T, et al. Brief, Web-based interventions to motivate smokers with schizophrenia: randomized trial. JMIR Ment Health 2020 Feb 08;7(2):e16524 [FREE Full text] [doi: 10.2196/16524] [Medline: 32039811]
87. Peter SC, Whelan JP, Pfund RA, Meyers AW. Can motivational messages engage individuals at-risk for gambling disorder in an online assessment? Psychol Addict Behav 2021 Feb;35(1):124-131. [doi: 10.1037/adb0000591] [Medline: 32525328]
88. Olthof MI, Blankers M, van Laar MW, Goudriaan AE. ICan, an Internet-based intervention to reduce cannabis use: study protocol for a randomized controlled trial. Trials 2021 Jan 06;22(1):28 [FREE Full text] [doi: 10.1186/s13063-020-04962-3] [Medline: 33407776]
89. Soucy JN, Hadjistavropoulos HD, Karin E, Dear BF, Titov N. Brief online motivational interviewing pre-treatment intervention for enhancing internet-delivered cognitive behaviour therapy: a randomized controlled trial. Internet Interv 2021 Sep;25:100394 [FREE Full text] [doi: 10.1016/j.invent.2021.100394] [Medline: 33996510]
90. Tobias MR, Landy LN, Levin ME, Arch JJ. A randomized trial of brief online interventions to facilitate treatment seeking for social anxiety. Behav Modif 2022 Sep;46(5):1137-1166. [doi: 10.1177/01454455211040051] [Medline: 34459271]
91. Yoon J, Fredua E, Davari SB, Ismail MH. The ATTAIN solution tested: initial pilot results of an automated, Web-based screening tool for unhealthy drinking behaviors. Perm J 2021 May;25(2):1-6 [FREE Full text] [doi: 10.7812/TPP/20.143] [Medline: 33970076]
92. Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. Psychother Psychosom 2009;78(5):275-284. [doi: 10.1159/000228248] [Medline: 19602916]
93. Wilson CJ, Deane FP, Ciarrochi JV, Rickwood D. Measuring help-seeking intentions: properties of the general help-seeking questionnaire. Can J Couns 2005;39(1):15-28 [FREE Full text] [doi: 10.1037/t42876-000]
94. Donovan RJ, Jones S, Holman CD, Corti B. Assessing the reliability of a stage of change scale. Health Educ Res 1998 Jun;13(2):285-291. [doi: 10.1093/her/13.2.285] [Medline: 10181026]
95. Dozois DJ, Westra HA, Collins KA, Fung TS, Garry JK. Stages of change in anxiety: psychometric properties of the University of Rhode Island Change Assessment (URICA) scale. Behav Res Ther 2004 Jun;42(6):711-729. [doi: 10.1016/S0005-7967(03)00193-1] [Medline: 15081886]
96. Lawrence P, Fulbrook P, Somerset S, Schulz P. Motivational interviewing to enhance treatment attendance in mental health settings: a systematic review and meta-analysis. J Psychiatr Ment Health Nurs 2017 Nov;24(9-10):699-718. [doi: 10.1111/jpm.12420] [Medline: 28816412]
97. Corrigan PW, Rao D. On the self-stigma of mental illness: stages, disclosure, and strategies for change. Can J Psychiatry 2012 Aug;57(8):464-469 [FREE Full text] [doi: 10.1177/070674371205700804] [Medline: 22854028]
98. Schnyder N, Panczak R, Groth N, Schultze-Lutter F. Association between mental health-related stigma and active help-seeking: systematic review and meta-analysis. Br J Psychiatry 2017 Apr;210(4):261-268. [doi: 10.1192/bjp.bp.116.189464] [Medline: 28153928]
99. Rhodes RE, de Bruijn GJ. How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. Br J Health Psychol 2013 May;18(2):296-309. [doi: 10.1111/bjhp.12032] [Medline: 23480428]
100. ElHaffar G, Durif F, Dubé L. Towards closing the attitude-intention-behavior gap in green consumption: a narrative review of the literature and an overview of future research directions. J Clean Prod 2020 Dec 1;275:122556. [doi: 10.1016/j.jclepro.2020.122556]
101. Baumann S, Gaertner B, Schnuerer I, Haberecht K, John U, Freyer-Adam J. Belief incongruence and the intention-behavior gap in persons with at-risk alcohol use. Addict Behav 2015 Sep;48:5-11. [doi: 10.1016/j.addbeh.2015.04.007] [Medline: 25930010]
102. Sheeran P, Aubrey R, Kellett S. Increasing attendance for psychotherapy: implementation intentions and the self-regulation of attendance-related negative affect. J Consult Clin Psychol 2007 Dec;75(6):853-863. [doi: 10.1037/0022-006X.75.6.853] [Medline: 18085903]
103. Swift JK, Callahan JL. Decreasing treatment dropout by addressing expectations for treatment length. Psychother Res 2011 Mar;21(2):193-200. [doi: 10.1080/10503307.2010.541294] [Medline: 21271460]
104. Rapp A, Tirassa M, Tirabeni L. Rethinking technologies for behavior change: a view from the inside of human change. ACM Trans Comput Hum Interact 2019 Aug 31;26(4):1-30. [doi: 10.1145/3318142]

Abbreviations

CBT: cognitive behavioral therapy
CG: control group

FTF: face-to-face
IG: intervention group
MI: motivational interviewing
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
TTM: transtheoretical model

Edited by A Mavragani; submitted 09.03.22; peer-reviewed by T Fleming, C Varsi; comments to author 22.05.22; revised version received 02.06.22; accepted 31.07.22; published 30.08.22.

Please cite as:
Jardine J, Bowman R, Doherty G
Digital Interventions to Enhance Readiness for Psychological Therapy: Scoping Review
J Med Internet Res 2022;24(8):e37851
URL: https://www.jmir.org/2022/8/e37851
doi:10.2196/37851
PMID:36040782

Review

The Effects of Internet-Based Acceptance and Commitment Therapy on Process Measures: Systematic Review and Meta-analysis

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

[1]Department of Occupational Therapy, University of Alabama at Birmingham, Birmingham, AL, United States
[2]Department of Psychiatry, Yonsei University Wonju College of Medicine, Wonju, Gangwon province, Republic of Korea
[3]Department of Psychiatry, Yonsei University Wonju Christian Hospital, Wonju, Gangwon province, Republic of Korea

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

Abstract

Background: Acceptance and commitment therapy (ACT) is based on a psychological flexibility model that encompasses 6 processes: acceptance, cognitive defusion, self-as-context, being present, values, and committed action.

Objective: This systematic review and meta-analysis of randomized controlled trials (RCTs) aimed to examine the effects of internet-based ACT (iACT) on process measures.

Methods: A comprehensive search was conducted using 4 databases. The quality of the included RCTs was assessed using the Cochrane Collaboration Risk of Bias Tool. A random-effects or fixed-effects model was used. Subgroup analyses for each outcome were conducted according to the type of control group, use of therapist guidance, delivery modes, and use of targeted participants, when applicable.

Results: A total of 34 RCTs met the inclusion criteria. This meta-analysis found that iACT had a medium effect on psychological flexibility and small effects on mindfulness, valued living, and cognitive defusion at the immediate posttest. In addition, iACT had a small effect on psychological flexibility at follow-up. The overall risk of bias across studies was unclear.

Conclusions: Relatively few studies have compared the effects of iACT with active control groups and measured the effects on mindfulness, valued living, and cognitive defusion. These findings support the processes of change in iACT, which mental health practitioners can use to support the use of iACT.

(J Med Internet Res 2022;24(8):e39182) doi:10.2196/39182

KEYWORDS

acceptance and commitment therapy; process measure; internet-based intervention; digital mental health; meta-analysis; mindfulness; systematic review

Introduction

Background

Acceptance and commitment therapy (ACT), which is a type of mindfulness-based intervention, is an empirically supported transdiagnostic approach [1,2]. Mindfulness is defined as “awareness that arises through paying attention, on purpose, in the present moment, and nonjudgmentally” [3]. ACT aims to develop greater psychological flexibility, that is, the ability to face challenging experiences in an open, conscious manner and change one’s behaviors to participate in valued activities rather than avoiding or suppressing uncomfortable or painful experiences, emotions, and thoughts [1,2]. ACT is based on a psychological flexibility model involving 6 processes [2]. These six processes include (1) acceptance (ie, being open to unwanted thoughts and emotions as they are), (2) cognitive defusion (ie, stepping back from unhelpful thoughts and emotions to reduce their dominance over behaviors), (3) being present (ie, maintaining voluntary and flexible contact with the present moment), (4) observing self (ie, flexible self-conceptualization and perspective taking), (5) values (ie, identifying and connecting values to behaviors for a meaningful life), and (6) committed action (ie, establishing patterns of behaviors to live a meaningful life aligned with values) [2]. The first 2 processes (1 and 2) are conceptualized as *acceptance and mindfulness processes*, and the last 2 processes (5 and 6) are conceptualized as *commitment and behavior change processes*; the remaining 2 processes (3 and 4) are conceptualized to be shared by the *acceptance and mindfulness processes* and *commitment and behavior change processes* [4].

Studies have used ACT process measures to assess the effects of ACT on these process measures and to understand the processes of change [5,6]. Measuring psychological flexibility and its interrelated processes (eg, cognitive defusion, mindfulness, and valued living) has been emphasized to better understand why ACT works and how it directly affects these process measures [7]. Improvement of psychological flexibility and its interrelated processes is theorized to foster improvements in mental health outcomes, and this theory is supported by empirical studies to some extent [8]. For example, studies have found a significant predictive role of psychological flexibility in predicting mental health outcomes; negative associations of psychological flexibility with depressive symptoms, anxiety, and overall psychological distress; and positive associations of psychological flexibility with quality of life, emotional well-being, and resilience [9,10]. Studies have also shown negative relationships between cognitive defusion and psychological distress and a significant predictive role of cognitive fusion (ie, the opposite of cognitive defusion) for depressive symptoms, anxiety, distress, and lowered quality of life [11-14]. In addition, the literature suggests that personal values and mindfulness have a significant effect on psychological distress and quality of life [15,16].

Although a growing body of evidence shows that ACT can improve health and well-being outcomes in various populations [17,18], the synthesized evidence of ACT on process measures is lacking. For example, previous meta-analysis studies did not consider ACT process measures except for psychological flexibility, possibly because of the limited number of published studies and because previous meta-analysis studies often limited their review questions to specific populations [19]. Internet-based psychological interventions are easy to access and inexpensive; therefore, it is important to determine whether internet-based ACT (iACT) is an effective alternative option [20]. In particular, Brown et al [21] conducted a meta-analysis to measure the effects of iACT on outcomes related to mental health and well-being in any population, but none of the ACT process measures were addressed, possibly because of the limited number of included studies for meta-analyses of these outcomes at that time. In addition, Thompson et al [22] conducted a similar meta-analysis that involved only meta-analysis for psychological flexibility, with no meta-analyses for any other process measures.

Objectives

This systematic review and meta-analysis aimed to assess the effects of iACT on different process measures (eg, psychological flexibility, valued living, mindfulness, and cognitive defusion) in any population. In addition, this meta-analysis aimed to conduct subgroup analyses for each outcome according to the type of control group to determine whether the effects of iACT differed when compared with active control groups provided with other comparable interventions and passive control groups provided with no intervention. In addition, other subgroup analyses related to the characteristics of the included studies may be possible and may provide useful information. For example, studies have found that iACT with therapist guidance showed larger effects on psychological flexibility compared with iACT without therapist guidance, and populations with psychological distress symptoms showed larger effects on mental health outcomes compared with nonclinical populations [22]. As studies have also found negative associations of psychological flexibility and cognitive defusion with psychological distress as well as a significant effect of personal values and mindfulness on psychological distress, studies that directly targeted people with psychological distress symptoms might show greater effects on ACT process measures [10-12,14-16]. Outcomes may differ based on how iACT was delivered (eg, web-based ACT modules, iACT accompanied by in-person ACT sessions, and videoconferencing ACT). Therefore, this study aimed to conduct additional subgroup analyses according to the use of therapist guidance, delivery modes, and the use of targeted participants with psychological distress symptoms, when applicable.

Methods

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [23] and the Cochrane Handbook for Systematic Reviews of Interventions (version 5.1.0) [24] were used as guides for conducting and reporting this systematic review and meta-analysis. This study was not preregistered.

Inclusion and Exclusion Criteria

Studies were selected based on the following inclusion criteria: (1) the study must be a randomized controlled trial (RCT), (2) ACT must be delivered mainly on the web (ie, iACT), (3) the study must have pre-post test results in ACT process measures (eg, psychological flexibility, mindfulness, valued living, and cognitive defusion), (4) the study must compare iACT with a non-ACT condition or with a control condition, and (5) the study must be written in English. Studies were excluded if they compared ACT groups only with different delivery modes without any other comparison or control condition (eg, ACT delivered on the web vs ACT delivered in person).

Search Strategy

Relevant articles were identified by searching 4 electronic databases from the date of inception of each database to June 5, 2021. The databases were PubMed (1966-2021), CINAHL (1981-2021), PsycINFO (1935-2021), and Scopus (1966-2021). Key search terms were combined to identify the relevant literature using keywords for iACT. To broaden the database search, keywords relevant to the outcomes were not entered as search terms. The full search strategies for all databases can be found in Multimedia Appendix 1. Articles were also searched manually using the reference lists of the identified articles and related article features in the databases.

Data Extraction and Quality Assessment

The characteristics of the included RCTs, such as the country of origin, characteristics of participants, description of intervention and control groups, outcome measures, and results (ie, between-group differences with *P* values), were extracted into a table. The mean and SDs at each data collection time point and sample sizes of the intervention and control groups in the included studies were entered into a Microsoft Excel file. The methodological quality of the included RCTs was assessed using the Cochrane Collaboration Risk of Bias Tool [24]. The domains in the tool include random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective reporting. Risk of bias in each of the domains was judged as *low risk* of bias, *high risk* of bias, or *unclear risk* of bias following the criteria provided in the Cochrane Collaboration’s handbook [24]. Summary assessments of the risk of bias within a study and across studies were also determined based on the handbook’s criteria [24]. One author with extensive experience in conducting systematic reviews and expertise in ACT completed the process for data extraction and quality assessment.

Meta-analysis

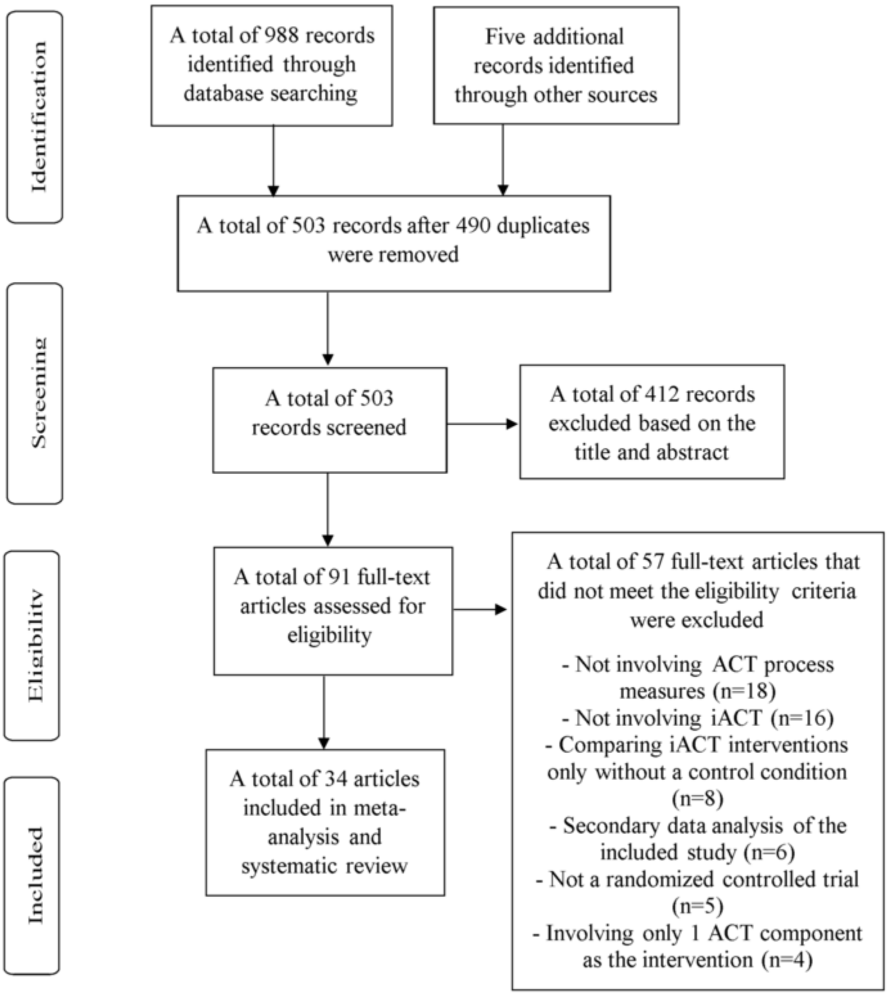
Means, SDs, and sample sizes of intervention and control groups in the included studies were entered into RevMan (version 5.4; Cochrane Collaboration) for meta-analyses and pooled for each outcome at the immediate posttest and at follow-up. The I^2 statistic was used to indicate statistical heterogeneity across studies, and $I^2>60\%$ might indicate substantial heterogeneity [24]. The decision to use either a random-effects model or a fixed-effects model with the inverse variance method was determined using I^2 statistic values for each outcome. In other words, a random-effects model was used when the I^2 statistic for each variable was >60%; otherwise, a fixed-effects model was used. The standardized mean difference (SMD) with 95% CIs was used as a summary statistic for the size of the intervention effect to account for outcomes measured using different assessment tools [24]. SMDs<0.4 indicate a small effect, SMDs between 0.4 and 0.7 indicate a medium effect, and SMDs>0.7 indicate a large effect [24]. The mean difference, rather than SMD, was used when the studies used the same assessment tool [24]. Subgroup analyses for each outcome were performed according to the type of control group, if applicable, to see whether the effects of iACT differed compared with active control groups provided with other comparable interventions and compared with passive control groups provided with no intervention (ie, treatment as usual control groups and wait-list control groups). Additional subgroup analyses for each outcome were conducted according to the use of therapist guidance, delivery modes, and the use of targeted participants with psychological distress symptoms, when applicable. Funnel plot analysis was used to test for possible publication bias (ie, studies with positive findings are more likely to be published) [24]. A possible publication bias was suggested if the inverted funnel shape was asymmetrical [24].

Results

Selection of Studies

Figure 1 illustrates the study selection process. A total of 988 articles were identified through database searching, and 5 additional articles were identified through hand searching. After removing 490 duplicates, 503 articles were screened based on their titles and abstracts. A total of 412 articles were excluded based on title and abstract screening, and 91 articles were assessed for eligibility by reading the full text. A total of 57 articles were excluded after reading the full text because of the following reasons: not involving any ACT process measures (18 studies), not involving iACT (16 studies), comparing iACT interventions delivered differently without a control condition (8 studies), involving secondary data analysis (6 studies), not an RCT (5 studies), and involving only 1 ACT component for the intervention (4 studies). A total of 34 articles met the eligibility criteria [8,25-57].

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



Characteristics of the Included Studies

The main characteristics of the included RCTs are summarized in Multimedia Appendix 2. The average number of ACT modules (sessions) in the included studies was 6.4 (SD 2.5), ranging from 2 to 12 modules in total. The average duration of ACT modules (sessions) delivered in the included studies was 7.9 (SD 3.2) weeks, ranging from 2 to 15 weeks in total. ACT was delivered on the web with therapist guidance in 26 studies (eg, via videoconferencing, phone calls, written feedback, and a mobile app) and without therapist guidance in 8 studies [25-32]. A total of 7 studies used a blended ACT program, which involved both iACT and in-person sessions [33-40]. A total of 3 studies involved videoconferencing ACT [41,42]. The remaining 24 RCTs used web-based ACT modules. Out of a total of 34 studies, 11 (32%) involved active control groups, including web-based cognitive behavioral therapy (CBT) [25], web-based smoking cessation interventions [26,27], web-based discussion forums [44,45,57], web-based mental health education [30], in-person behavioral support [35], web-based expressive writing [48,55], and in-person documentary discussion [38]. The population of the included studies varied widely, including college students [8,29-32,36,47,54], adults with chronic pain [37,39,52,55-57], family caregivers of people with chronic conditions [42,49,50,53], adults with insomnia [28,33,41], and smokers [26,27,35] (refer to Multimedia Appendix 2 for these different participant characteristics in the included studies). A total of 14 studies directly targeted people with certain types of psychological distress, such as depressive symptoms, anxiety, stress, and overall psychological distress [8,27,33,36-39,42,44-49].

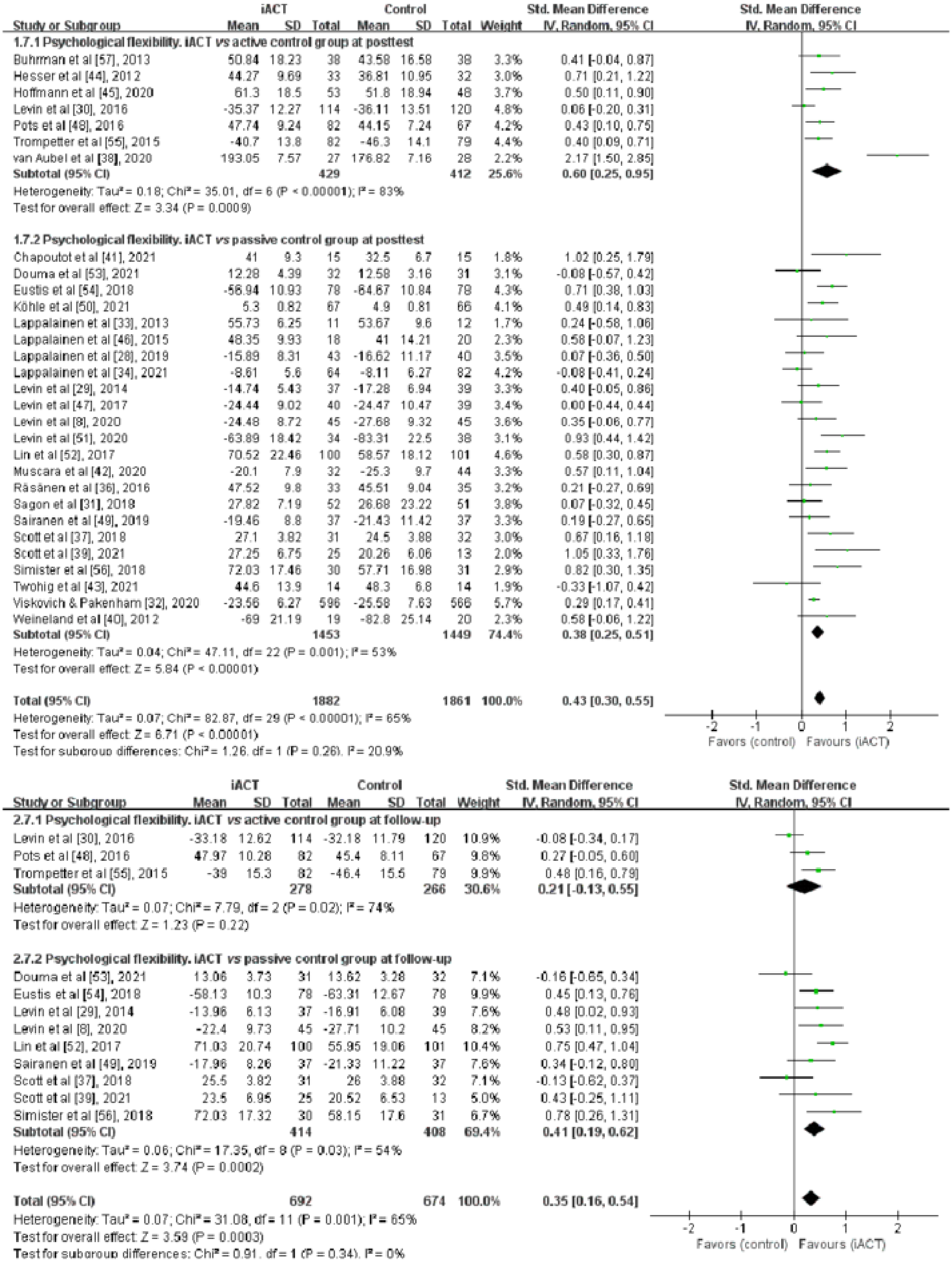
The average sample size of participants in the included RCTs was 141 (SD 196), ranging from 24 to 1162. The mean age of the participants was 37.8 (SD 13.2) years, ranging from 13.9 to 55.9 years, and the average percentage of female participants was 68.1% (SD 19.7%), ranging from 0% to 98.5%. The included RCTs were conducted in the United States (10 studies), Finland (5 studies), Sweden (4 studies), the Netherlands (4 studies), the United Kingdom (2 studies), Australia (2 studies), Ireland (2 studies), Canada (1 study), Belgium (1 study), France (1 study), Denmark (1 study), and Germany (1 study). Of the 34 included studies, 26 (76%) were published between 2016 and 2021 and the remaining 8 (24%) were published between 2012 and 2015.

The following sections describe the results of the meta-analysis of the efficacy of iACT for psychological flexibility, mindfulness, valued living, and cognitive defusion at the immediate posttest and at follow-up, with findings of subgroup analyses listed according to the type of control group (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups), when applicable.

Effects of iACT on Improving Psychological Flexibility at the Immediate Posttest

A meta-analysis of 30 RCTs (n=3743 participants) found that iACT had a medium effect on improving psychological flexibility at the immediate posttest compared with control groups overall (SMD=0.43, 95% CI 0.30-0.55; Figure 2). There was no significant subgroup difference at the immediate posttest (χ^2_1=1.26; *P*=.26), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at the immediate posttest were not statistically different from one another. The iACT had a medium effect on psychological flexibility compared with active control groups at the immediate posttest (7 studies that involved 841 participants; SMD=0.60, 95% CI 0.25-0.95), whereas iACT had a small effect compared with passive control groups (23 studies that involved 2902 participants; SMD=0.38, 95% CI 0.25-0.51).

Figure 2. Forest plots showing the effects of internet-based acceptance and commitment therapy on psychological flexibility at the immediate posttest and at follow-up. iACT: internet-based acceptance and commitment therapy.



Study or Subgroup	iACT Mean	iACT SD	iACT Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.7.1 Psychological flexibility. iACT *vs* active control group at posttest								
Buhrman et al [57], 2013	50.84	18.23	38	43.58	16.58	38	3.3%	0.41 [-0.04, 0.87]
Hesser et al [44], 2012	44.27	9.69	33	36.81	10.95	32	3.0%	0.71 [0.21, 1.22]
Hoffmann et al [45], 2020	61.3	18.5	53	51.8	18.94	48	3.7%	0.50 [0.11, 0.90]
Levin et al [30], 2016	-35.37	12.27	114	-36.11	13.51	120	4.8%	0.06 [-0.20, 0.31]
Pots et al [48], 2016	47.74	9.24	82	44.15	7.24	67	4.2%	0.43 [0.10, 0.75]
Trompetter et al [55], 2015	-40.7	13.8	82	-46.3	14.1	79	4.4%	0.40 [0.09, 0.71]
van Aubel et al [38], 2020	193.05	7.57	27	176.82	7.16	28	2.2%	2.17 [1.50, 2.85]
Subtotal (95% CI)			**429**			**412**	**25.6%**	**0.60 [0.25, 0.95]**
Heterogeneity: Tau^2 = 0.18; Chi^2 = 35.01, df = 6 (P < 0.00001); I^2 = 83%								
Test for overall effect: Z = 3.34 (P = 0.0009)								
1.7.2 Psychological flexibility. iACT *vs* passive control group at posttest								
Chapoutot et al [41], 2021	41	9.3	15	32.5	6.7	15	1.8%	1.02 [0.25, 1.79]
Douma et al [53], 2021	12.28	4.39	32	12.58	3.16	31	3.1%	-0.08 [-0.57, 0.42]
Eustis et al [54], 2018	-56.94	10.93	78	-64.67	10.84	78	4.3%	0.71 [0.38, 1.03]
Köhle et al [50], 2021	5.3	0.82	67	4.9	0.81	66	4.1%	0.49 [0.14, 0.83]
Lappalainen et al [33], 2013	55.73	6.25	11	53.67	9.6	12	1.7%	0.24 [-0.58, 1.06]
Lappalainen et al [46], 2015	48.35	9.93	18	41	14.21	20	2.3%	0.58 [-0.07, 1.23]
Lappalainen et al [28], 2019	-15.89	8.31	43	-16.62	11.17	40	3.5%	0.07 [-0.36, 0.50]
Lappalainen et al [34], 2021	-8.61	5.6	64	-8.11	6.27	82	4.2%	-0.08 [-0.41, 0.24]
Levin et al [29], 2014	-14.74	5.43	37	-17.28	6.94	39	3.3%	0.40 [-0.05, 0.86]
Levin et al [47], 2017	-24.44	9.02	40	-24.47	10.47	39	3.4%	0.00 [-0.44, 0.44]
Levin et al [8], 2020	-24.48	8.72	45	-27.68	9.32	45	3.6%	0.35 [-0.06, 0.77]
Levin et al [51], 2020	-63.89	18.42	34	-83.31	22.5	38	3.1%	0.93 [0.44, 1.42]
Lin et al [52], 2017	70.52	22.46	100	58.57	18.12	101	4.6%	0.58 [0.30, 0.87]
Muscara et al [42], 2020	-20.1	7.9	32	-25.3	9.7	44	3.3%	0.57 [0.11, 1.04]
Räsänen et al [36], 2016	47.52	9.8	33	45.51	9.04	35	3.2%	0.21 [-0.27, 0.69]
Sagon et al [31], 2018	27.82	7.19	52	26.68	23.22	51	3.8%	0.07 [-0.32, 0.45]
Sairanen et al [49], 2019	-19.46	8.8	37	-21.43	11.42	37	3.3%	0.19 [-0.27, 0.65]
Scott et al [37], 2018	27.1	3.82	31	24.5	3.88	32	3.0%	0.67 [0.16, 1.18]
Scott et al [39], 2021	27.25	6.75	25	20.26	6.06	13	2.0%	1.05 [0.33, 1.76]
Simister et al [56], 2018	72.03	17.46	30	57.71	16.98	31	2.9%	0.82 [0.30, 1.35]
Twohig et al [43], 2021	44.6	13.9	14	48.3	6.8	14	1.9%	-0.33 [-1.07, 0.42]
Viskovich & Pakenham [32], 2020	-23.56	6.27	596	-25.58	7.63	566	5.7%	0.29 [0.17, 0.41]
Weineland et al [40], 2012	-69	21.19	19	-82.8	25.14	20	2.3%	0.58 [-0.06, 1.22]
Subtotal (95% CI)			**1453**			**1449**	**74.4%**	**0.38 [0.25, 0.51]**
Heterogeneity: Tau^2 = 0.04; Chi^2 = 47.11, df = 22 (P = 0.001); I^2 = 53%								
Test for overall effect: Z = 5.84 (P < 0.00001)								
Total (95% CI)			**1882**			**1861**	**100.0%**	**0.43 [0.30, 0.55]**
Heterogeneity: Tau^2 = 0.07; Chi^2 = 82.87, df = 29 (P < 0.00001); I^2 = 65%								
Test for overall effect: Z = 6.71 (P < 0.00001)								
Test for subgroup differences: Chi^2 = 1.26, df = 1 (P = 0.26), I^2 = 20.9%								

Study or Subgroup	iACT Mean	iACT SD	iACT Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.7.1 Psychological flexibility. iACT *vs* active control group at follow-up								
Levin et al [30], 2016	-33.18	12.62	114	-32.18	11.79	120	10.9%	-0.08 [-0.34, 0.17]
Pots et al [48], 2016	47.97	10.28	82	45.4	8.11	67	9.8%	0.27 [-0.05, 0.60]
Trompetter et al [55], 2015	-39	15.3	82	-46.4	15.5	79	9.9%	0.48 [0.16, 0.79]
Subtotal (95% CI)			**278**			**266**	**30.6%**	**0.21 [-0.13, 0.55]**
Heterogeneity: Tau^2 = 0.07; Chi^2 = 7.79, df = 2 (P = 0.02); I^2 = 74%								
Test for overall effect: Z = 1.23 (P = 0.22)								
2.7.2 Psychological flexibility. iACT *vs* passive control group at follow-up								
Douma et al [53], 2021	13.06	3.73	31	13.62	3.28	32	7.1%	-0.16 [-0.65, 0.34]
Eustis et al [54], 2018	-58.13	10.3	78	-63.31	12.67	78	9.9%	0.45 [0.13, 0.76]
Levin et al [29], 2014	-13.96	6.13	37	-16.91	6.08	39	7.6%	0.48 [0.02, 0.93]
Levin et al [8], 2020	-22.4	9.73	45	-27.71	10.2	45	8.2%	0.53 [0.11, 0.95]
Lin et al [52], 2017	71.03	20.74	100	55.95	19.06	101	10.4%	0.75 [0.47, 1.04]
Sairanen et al [49], 2019	-17.96	8.26	37	-21.33	11.22	37	7.6%	0.34 [-0.12, 0.80]
Scott et al [37], 2018	25.5	3.82	31	26	3.88	32	7.1%	-0.13 [-0.62, 0.37]
Scott et al [39], 2021	23.5	6.95	25	20.52	6.53	13	5.0%	0.43 [-0.25, 1.11]
Simister et al [56], 2018	72.03	17.32	30	58.15	17.6	31	6.7%	0.78 [0.26, 1.31]
Subtotal (95% CI)			**414**			**408**	**69.4%**	**0.41 [0.19, 0.62]**
Heterogeneity: Tau^2 = 0.06; Chi^2 = 17.35, df = 8 (P = 0.03); I^2 = 54%								
Test for overall effect: Z = 3.74 (P = 0.0002)								
Total (95% CI)			**692**			**674**	**100.0%**	**0.35 [0.16, 0.54]**
Heterogeneity: Tau^2 = 0.07; Chi^2 = 31.08, df = 11 (P = 0.001); I^2 = 65%								
Test for overall effect: Z = 3.59 (P = 0.0003)								
Test for subgroup differences: Chi^2 = 0.91, df = 1 (P = 0.34), I^2 = 0%								

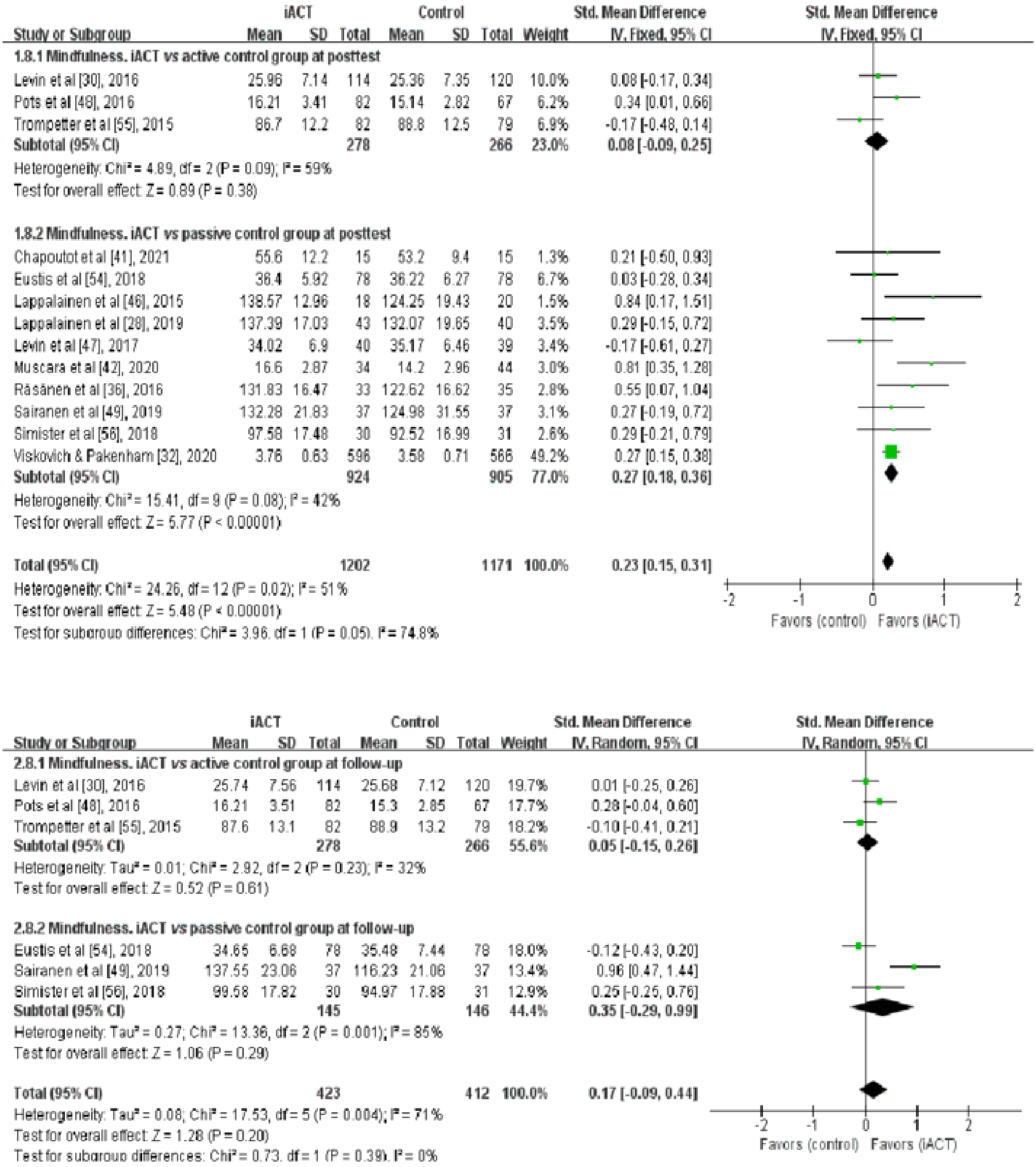
Effects of iACT on Improving Psychological Flexibility at Follow-up

A meta-analysis of 12 RCTs with follow-up data (n=1366 participants) revealed that iACT had a small effect on improving psychological flexibility at follow-up compared with control groups overall (SMD=0.35, 95% CI 0.16-0.54; Figure 2). There was no significant subgroup difference at follow-up (χ^2_1=0.91; *P*=.34), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at follow-up were not statistically different from one another. The iACT had a medium effect on improving psychological flexibility compared with passive control groups at follow-up (9 studies that involved 822 participants; SMD=0.41, 95% CI 0.19-0.62), but iACT was not significantly different from active control groups (3 studies that involved 544 participants; SMD=0.21, 95% CI −0.13 to 0.55).

Effects of iACT on Improving Mindfulness at the Immediate Posttest

A meta-analysis of 13 RCTs (n=2373 participants) showed that iACT had a small effect on improving mindfulness at the immediate posttest compared with control groups overall (SMD=0.23, 95% CI 0.15-0.31; Figure 3). There was a significant subgroup difference at the immediate posttest (χ^2_1=2.96; *P*=.05), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at the immediate posttest were statistically different from one another. The iACT had a small effect on improving mindfulness compared with passive control groups at the immediate posttest (10 studies that involved 1829 participants; SMD=0.27, 95% CI 0.18-0.36), but iACT was not significantly different from active control groups (3 studies that involved 544 participants; SMD=0.08, 95% CI −0.09 to 0.25).

Figure 3. Forest plots showing the effects of internet-based acceptance and commitment therapy on mindfulness at the immediate posttest and at follow-up. iACT: internet-based acceptance and commitment therapy.



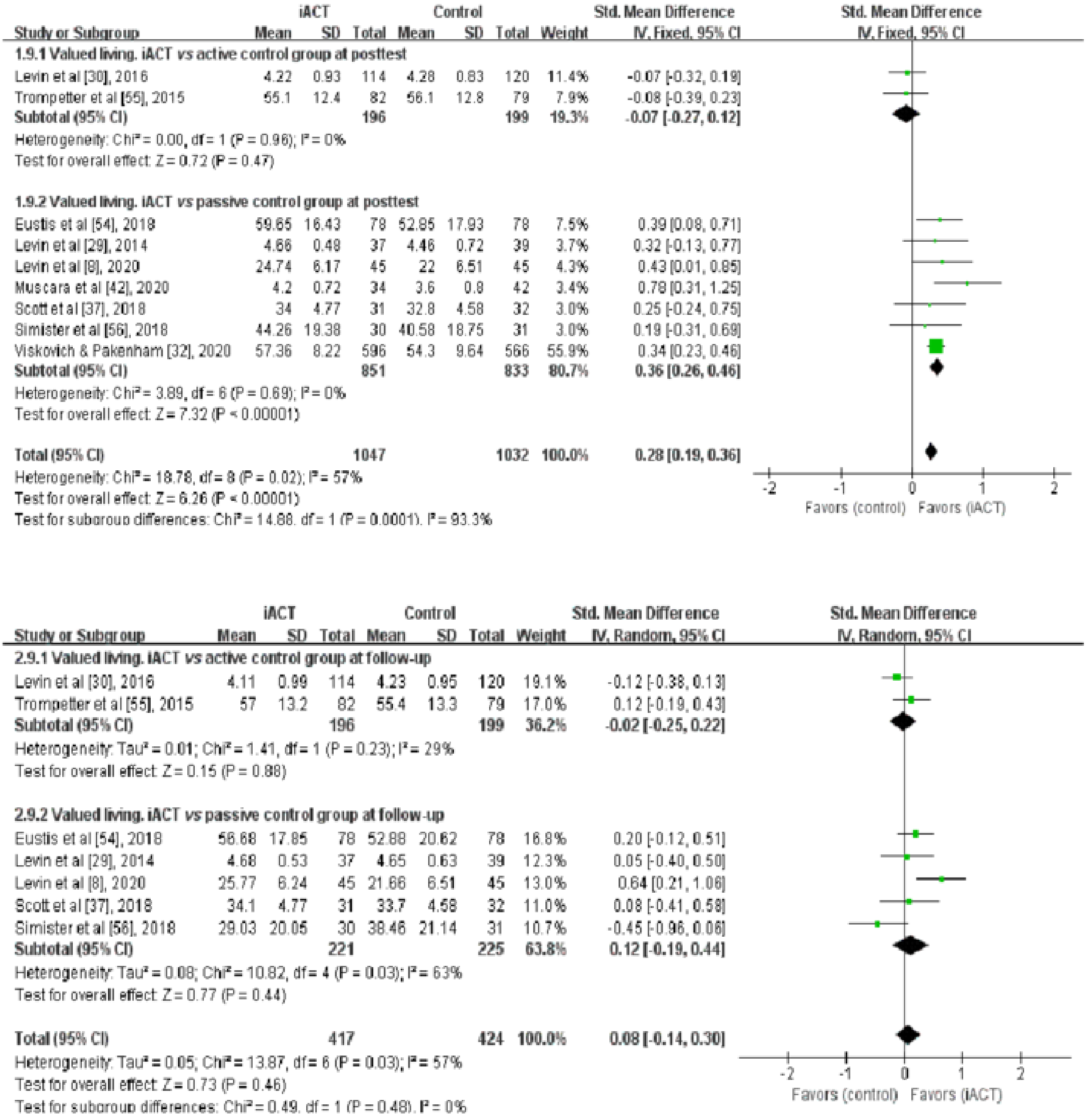
Effects of iACT on Improving Mindfulness at Follow-up

A meta-analysis of 6 RCTs with follow-up data (n=835 participants) found that iACT was not significantly different overall from control groups in improving mindfulness at follow-up (SMD=0.17, 95% CI −0.09 to 0.44; Figure 3). There was no significant subgroup difference at follow-up (χ^2_1=0.73; *P*=.39), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at follow-up were not statistically different from one another. No significant between-group difference in improving mindfulness was found at follow-up, regardless of control group conditions, including 3 studies (n=544 participants) that compared iACT with active control groups (SMD=0.05, 95% CI −0.15 to 0.26) and 3 studies (n=291) that compared iACT with passive control conditions (SMD=0.35, 95% CI −0.29 to 0.99).

Effects of iACT on Improving Valued Living at the Immediate Posttest

A meta-analysis of 9 RCTs (n=2079 participants) revealed that iACT had a small effect on improving valued living at the immediate posttest compared with control groups overall (SMD=0.28, 95% CI 0.19-0.36; Figure 4). There was a statistically significant subgroup difference at the immediate posttest (χ^2_1=14.88; P<.001), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at the immediate posttest were statistically different from one another. The iACT had a small effect on improving valued living compared with passive control groups at the immediate posttest (7 studies that involved 1684 participants; SMD=0.36, 95% CI 0.26-0.46), but iACT was not significantly different from active control groups (2 studies that involved 395 participants; SMD=−0.07, 95% CI −0.27 to 0.12).

Figure 4. Forest plots showing the effects of internet-based acceptance and commitment therapy on valued living at the immediate posttest and at follow-up. iACT: internet-based acceptance and commitment therapy.



	iACT			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
1.9.1 Valued living. iACT *vs* active control group at posttest								
Levin et al [30], 2016	4.22	0.93	114	4.28	0.83	120	11.4%	-0.07 [-0.32, 0.19]
Trompetter et al [55], 2015	55.1	12.4	82	56.1	12.8	79	7.9%	-0.08 [-0.39, 0.23]
Subtotal (95% CI)			**196**			**199**	**19.3%**	**-0.07 [-0.27, 0.12]**
Heterogeneity: Chi^2 = 0.00, df = 1 (P = 0.96); I^2 = 0%								
Test for overall effect: Z = 0.72 (P = 0.47)								
1.9.2 Valued living. iACT *vs* passive control group at posttest								
Eustis et al [54], 2018	59.65	16.43	78	52.85	17.93	78	7.5%	0.39 [0.08, 0.71]
Levin et al [29], 2014	4.66	0.48	37	4.46	0.72	39	3.7%	0.32 [-0.13, 0.77]
Levin et al [8], 2020	24.74	6.17	45	22	6.51	45	4.3%	0.43 [0.01, 0.85]
Muscara et al [42], 2020	4.2	0.72	34	3.6	0.8	42	3.4%	0.78 [0.31, 1.25]
Scott et al [37], 2018	34	4.77	31	32.8	4.58	32	3.0%	0.25 [-0.24, 0.75]
Simister et al [56], 2018	44.26	19.38	30	40.58	18.75	31	3.0%	0.19 [-0.31, 0.69]
Viskovich & Pakenham [32], 2020	57.36	8.22	596	54.3	9.64	566	55.9%	0.34 [0.23, 0.46]
Subtotal (95% CI)			**851**			**833**	**80.7%**	**0.36 [0.26, 0.46]**
Heterogeneity: Chi^2 = 3.89, df = 6 (P = 0.69); I^2 = 0%								
Test for overall effect: Z = 7.32 (P < 0.00001)								
Total (95% CI)			**1047**			**1032**	**100.0%**	**0.28 [0.19, 0.36]**
Heterogeneity: Chi^2 = 18.78, df = 8 (P = 0.02); I^2 = 57%								
Test for overall effect: Z = 6.26 (P < 0.00001)								
Test for subgroup differences: Chi^2 = 14.88, df = 1 (P = 0.0001), I^2 = 93.3%								

	iACT			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
2.9.1 Valued living. iACT *vs* active control group at follow-up								
Levin et al [30], 2016	4.11	0.99	114	4.23	0.95	120	19.1%	-0.12 [-0.38, 0.13]
Trompetter et al [55], 2015	57	13.2	82	55.4	13.3	79	17.0%	0.12 [-0.19, 0.43]
Subtotal (95% CI)			**196**			**199**	**36.2%**	**-0.02 [-0.25, 0.22]**
Heterogeneity: Tau^2 = 0.01; Chi^2 = 1.41, df = 1 (P = 0.23); I^2 = 29%								
Test for overall effect: Z = 0.15 (P = 0.88)								
2.9.2 Valued living. iACT *vs* passive control group at follow-up								
Eustis et al [54], 2018	56.68	17.85	78	52.88	20.62	78	16.8%	0.20 [-0.12, 0.51]
Levin et al [29], 2014	4.68	0.53	37	4.65	0.63	39	12.3%	0.05 [-0.40, 0.50]
Levin et al [8], 2020	25.77	6.24	45	21.66	6.51	45	13.0%	0.64 [0.21, 1.06]
Scott et al [37], 2018	34.1	4.77	31	33.7	4.58	32	11.0%	0.08 [-0.41, 0.58]
Simister et al [56], 2018	29.03	20.05	30	38.46	21.14	31	10.7%	-0.45 [-0.96, 0.06]
Subtotal (95% CI)			**221**			**225**	**63.8%**	**0.12 [-0.19, 0.44]**
Heterogeneity: Tau^2 = 0.08; Chi^2 = 10.82, df = 4 (P = 0.03); I^2 = 63%								
Test for overall effect: Z = 0.77 (P = 0.44)								
Total (95% CI)			**417**			**424**	**100.0%**	**0.08 [-0.14, 0.30]**
Heterogeneity: Tau^2 = 0.05; Chi^2 = 13.87, df = 6 (P = 0.03); I^2 = 57%								
Test for overall effect: Z = 0.73 (P = 0.46)								
Test for subgroup differences: Chi^2 = 0.49, df = 1 (P = 0.48), I^2 = 0%								

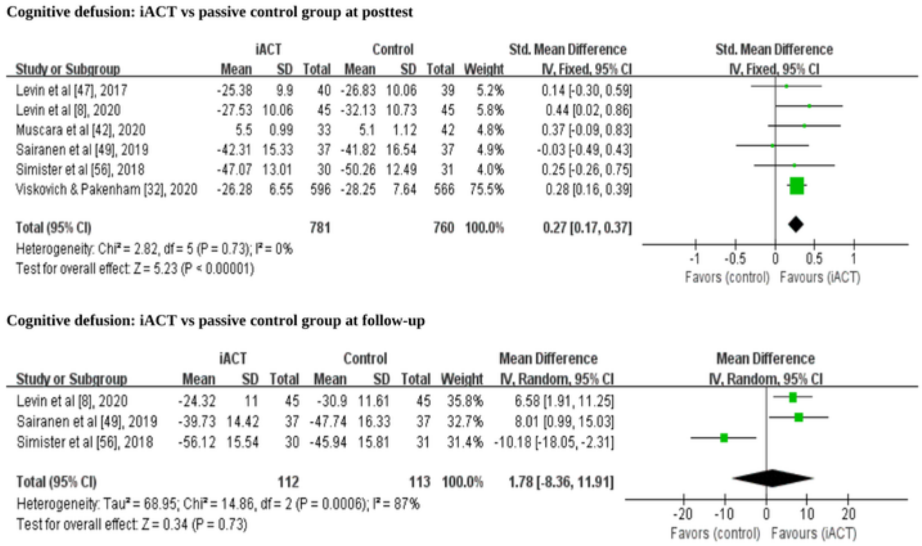
Effects of iACT on Improving Valued Living at Follow-up

A meta-analysis of 7 RCTs with follow-up data (n=841 participants) showed that iACT did not differ from control groups in improving valued living at follow-up (SMD=0.08, 95% CI −0.14 to 0.30; Figure 4). There was no statistically significant subgroup difference at follow-up (χ^2_1=0.49; *P*=.48), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at follow-up were not statistically different from one another. No significant between-group difference in improving valued living was found at follow-up, regardless of control group conditions, including 2 studies (n=395 participants) that compared iACT with active control groups (SMD=−0.02, 95% CI −0.25 to 0.22) and 5 studies (n=446 participants) that compared iACT with passive control conditions (SMD=0.12, 95% CI −0.19 to 0.44).

Effects of iACT on Improving Cognitive Defusion at the Immediate Posttest

All studies that measured cognitive defusion compared iACT with passive control groups only; therefore, a subgroup analysis was not conducted. A meta-analysis of 6 RCTs (n=1541 participants) found that iACT had a small effect on improving cognitive defusion at the immediate posttest compared with passive control groups (SMD=0.27, 95% CI 0.17-0.37; Figure 5). A meta-analysis of 3 RCTs with follow-up data (n=225 participants) revealed that iACT was not different from passive control groups in improving cognitive defusion at follow-up (SMD=1.78, 95% CI −8.36 to 11.91; Figure 5).

Figure 5. Forest plots showing the effects of internet-based acceptance and commitment therapy on cognitive defusion at the immediate posttest and at follow-up. iACT: internet-based acceptance and commitment therapy.



Subgroup Analyses According to the Use of Therapist Guidance

Subgroup analyses showed medium effects of iACT with therapist guidance on psychological flexibility at the immediate posttest (25 studies that involved 2085 participants; SMD=0.50, 95% CI 0.34-0.65) and at follow-up (10 studies that involved 1056 participants; SMD=0.40, 95% CI 0.21-0.58) and small effects of iACT with therapist guidance on mindfulness (10 studies that involved 894 participants; SMD=0.20, 95% CI 0.07-0.34) and valued living (7 studies that involved 686 participants; SMD=0.26, 95% CI 0.04-0.48) at the immediate posttest compared with control groups. The iACT with therapist guidance was not significantly different from the control groups in valued living at follow-up (5 studies that involved 531 participants, SMD=0.14, 95% CI −0.15 to 0.43). Conversely, subgroup analyses revealed small effects of iACT without therapist guidance on psychological flexibility (6 studies that involved 1700 participants, SMD=0.18, 95% CI 0.03-0.32) and mindfulness (3 studies that involved 1479 participants, SMD=0.24, 95% CI 0.14-0.34) at the immediate posttest only. The iACT without therapist guidance was not significantly different from control groups in psychological flexibility at follow-up (2 studies that involved 310 participants; SMD=0.17, 95% CI −0.38 to 0.71) and valued living at the immediate posttest (3 studies that involved 1472 participants; SMD=0.20, 95% CI −0.10 to 0.49) and follow-up (2 studies that involved 310 participants; SMD=−0.08, 95% CI −0.30 to 0.14). In particular, the pooled effect size of studies that involved iACT with therapist guidance was greater than that of studies that involved iACT without therapist guidance in psychological flexibility at the immediate posttest. In addition, although there was a statistically significant pooled effect of iACT with therapist guidance on valued living at the immediate posttest, iACT without therapist guidance was not significantly different from the control groups. There was no statistically significant subgroup difference in any of the outcomes (*P*>.05), except for psychological flexibility at the immediate posttest (*P*=.003). These findings indicate that there was no statistically significant difference among studies according to the use of therapist guidance in all outcomes, except for psychological flexibility at the immediate posttest, in which a statistically significant larger effect of iACT was found when iACT studies involved therapist guidance (ie, SMD=0.50 vs SMD=0.18). Forest plots of these subgroup analyses are illustrated in Multimedia Appendices 3-7.

Subgroup Analyses According to the Delivery Modes

Subgroup analyses showed small effects of web-based ACT on psychological flexibility (22 studies that involved 3257 participants; SMD=0.38, 95% CI 0.26-0.49) and mindfulness at the immediate posttest (10 studies that involved 2197 participants; SMD=0.20, 95% CI 0.11-0.28) compared with control groups. Subgroup analyses revealed medium effects of iACT accompanied by in-person ACT sessions on psychological flexibility (6 studies that involved 394 participants; SMD=0.61, 95% CI 0.02-1.20) and mindfulness (1 study that involved 68 participants; SMD=0.55, 95% CI 0.07-1.04) at the immediate posttest compared with control groups. Although videoconferencing ACT was not significantly different from control groups in psychological flexibility at the immediate posttest (3 studies that involved 134 participants; SMD=0.43, 95% CI −0.25 to 1.11), subgroup analyses found medium effects of videoconferencing ACT on mindfulness at the immediate posttest compared with control groups (2 studies that involved 108 participants; SMD=0.64, 95% CI 0.24-1.03). There was no statistically significant subgroup difference in psychological flexibility at the immediate posttest (*P*>.05), indicating that there was no statistically significant difference among studies according to delivery modes. However, there was a statistically significant subgroup difference in mindfulness at the immediate posttest (*P*=.04), suggesting a statistically significant difference among studies according to their delivery modes, in which medium effects of iACT with in-person sessions and videoconferencing ACT were found, whereas web-based ACT showed small effects. Subgroup analyses according to the delivery modes were not conducted for the other outcomes because of the lack of studies on any other outcomes. Forest plots of these subgroup analyses are illustrated in Multimedia Appendices 8 and 9.

Subgroup Analyses According to the Targeted Participants With Psychological Distress Symptoms

Subgroup analyses found a medium effect of iACT on psychological flexibility at the immediate posttest (13 studies that involved 919 participants; SMD=0.55, 95% CI 0.31-0.79) and small effects of iACT on psychological flexibility at follow-up (5 studies that involved 414 participants; SMD=0.29, 95% CI 0.09-0.49) and on mindfulness (6 studies that involved 486 participants; SMD=0.38, 95% CI 0.20-0.56), valued living (4 studies that involved 308 participants; SMD=0.35, 95% CI 0.02-0.69), and cognitive defusion (4 studies that involved 318 participants; SMD=0.24, 95% CI 0.02-0.46) at the immediate posttest compared with control groups when studies directly targeted participants with some type of psychological distress. Subgroup analyses, however, showed small effects of iACT on psychological flexibility at the immediate posttest (18 studies that involved 2866 participants; SMD=0.33, 95% CI 0.19-0.48) and at follow-up (7 studies that involved 952 participants; SMD=0.38, 95% CI 0.10-0.67), mindfulness at the immediate posttest (7 studies that involved 1887 participants; SMD=0.19, 95% CI 0.10-0.28), and cognitive defusion at the immediate posttest (2 studies that involved 1223 participants; SMD=0.28, 95% CI 0.16-0.39) compared with control groups in studies that did not involve targeted participants with some type of psychological distress. However, there was no statistically significant difference in iACT among control groups in mindfulness and valued living at follow-up when studies did or did not directly target participants with some type of psychological distress, or in valued living at the immediate posttest when studies did not involve targeted participants with some type of psychological distress. There were no statistically significant subgroup differences in any of the outcomes (*P*>.05), indicating that there was no statistically significant difference among studies according to the use of targeted participants with some type of psychological distress in all the outcomes. Forest plots of these subgroup analyses are illustrated in Multimedia Appendices 10-16.

Risk of Bias and Publications Bias of the Included Studies

Out of the 34 included studies, 18 (53%) had an unclear risk of bias, 10 (29%) had a low risk of bias, and 6 (18%) had a high overall risk of bias (Multimedia Appendix 17). A domain regarding blinding of participants and personnel was not regarded as the key domain for the overall risk of bias within a study because studies that involved passive control conditions were less able to conceal the group allocation from participants. The overall risk of bias across the 34 studies was interpreted as unclear because most information was from studies with an unclear risk of bias [24]. The risk of bias in each domain for each study is reported in Multimedia Appendix 17, and the risk of bias summary graph and chart are reported in Multimedia Appendix 18 [8,25-57] and Multimedia Appendix 19.

As at least 10 studies have suggested using the funnel plot asymmetry analysis, possible publication bias was tested for psychological flexibility and mindfulness at posttest only [24]. Overall, the funnel plots tend to be symmetrical, although more studies are needed to better interpret such visual inspection, especially for mindfulness. Funnel plots are reported in Multimedia Appendix 20.

Discussion

This systematic review and meta-analysis identified 34 RCTs that assessed the efficacy of iACT for process measures (ie, psychological flexibility, mindfulness, valued living, and cognitive defusion). This meta-analysis found that iACT had a medium effect on improving psychological flexibility and small effects on improving mindfulness, valued living, and cognitive defusion at the immediate posttest. A small effect of iACT on psychological flexibility was also observed at follow-up.

Previous meta-analysis studies that involved ACT (ie, not iACT in particular) found similar findings to this meta-analysis study that included a medium effect of ACT on psychological flexibility in family caregivers [6] and a small effect of self-help ACT on psychological flexibility in adult populations [58]. A total of 26 studies that assessed the effects of iACT on process measures were conducted from 2016 to the end date of the search for this meta-analysis (June 5, 2021), whereas only 8 studies were published before 2016. This may explain why a previous meta-analysis study of iACT in particular did not include process measures [21]. Thompson et al [22] conducted a meta-analysis for psychological flexibility only among process measures and found small effects of iACT on psychological flexibility at the immediate posttest and at follow-up. These findings by Thompson et al [22] were based on 23 studies at immediate posttest and 11 studies at follow-up. This meta-analysis found a medium effect at the immediate posttest (based on 30 studies) and a small effect at follow-up (based on 12 studies). The larger effect size of iACT on psychological flexibility at the immediate posttest found in this study may be because of the inclusion of 7 more studies than the meta-analysis by Thompson et al [22]. Unlike previous meta-analysis studies for iACT [21,22], this meta-analysis study conducted meta-analyses in other process measures, such as mindfulness, valued living, and cognitive defusion, and found small effects on these measures at the immediate posttest. Thus, the findings of this meta-analysis contribute to the body of evidence on the efficacy of iACT for different process measures, which could be applicable to both clinical and nonclinical populations, as ACT is a transdiagnostic approach [1].

Subgroup analyses for each outcome were conducted in this meta-analysis study according to the type of control group, unlike previous iACT meta-analysis studies [21,22]. Subgroup analyses showed small effects of iACT on psychological flexibility, mindfulness, valued living, and cognitive defusion at the immediate posttest compared with passive control groups. In contrast, subgroup analyses that compared iACT with active control groups found no significant between-group differences in mindfulness and valued living at the immediate posttest and follow-up and in psychological flexibility at follow-up. Such findings indicate that iACT was not significantly more effective than active control conditions (eg, CBT and mental health education). However, relatively few studies have compared the effects of iACT with active control conditions. There were 3 times more studies that compared iACT with passive control conditions than those that compared iACT with active control conditions in all the process measures, except for mindfulness at follow-up. Such a gap in the literature suggests a need for studies that compare iACT with active control conditions, such as CBT, psychoeducational interventions, and support groups, to better understand whether iACT is comparable or superior to other evidence-based treatments in improving process measures such as mindfulness and cognitive defusion.

This study conducted subgroup analyses according to the use of therapist guidance, delivery modes, and targeted participants with symptoms of psychological distress. This study found no statistically significant subgroup difference among studies of these 3 characteristics in all the outcomes, except for the subgroup difference among studies according to the use of therapist guidance for psychological flexibility at the immediate posttest (ie, a larger effect of iACT when iACT was provided with therapist guidance compared with iACT without therapist guidance) and for the subgroup difference among studies according to the delivery modes for mindfulness at the immediate posttest (ie, a larger effect of iACT when iACT was delivered with in-person sessions or videoconferencing ACT compared with web-based ACT modules). Thompson et al [22] also found larger effects of iACT with therapist guidance on psychological flexibility compared with iACT without therapist guidance. However, more studies are needed to confirm these findings, especially in subgroup analyses according to delivery modes, because only a few studies have involved delivery modes other than web-based ACT modules.

This review had several limitations that should be considered when interpreting the findings. A total of 4 electronic databases were used to search the literature, and some relevant articles could have been missed if they were published only in other databases. Only studies written in English were searched and included in this review, which could create a publication bias. One author with extensive experience in comprehensive literature reviews and expertise in ACT searched the literature; therefore, this review did not include 2 independent reviewers in the search process. A recent systematic review found that single screening for study selection in systematic reviews conducted by experienced reviewers had no impact on the findings of the meta-analysis [59]. The overall risk of bias across the included RCTs was interpreted as unclear, indicating the need for high-quality studies to better determine the effects of iACT on process measures.

According to this meta-analysis study, there were 2 to 5 times more studies that assessed psychological flexibility than studies that assessed mindfulness, valued living, and cognitive defusion. Investigators of further studies should consider assessing diverse ACT process measures to better understand the processes of change. In addition, relatively few studies have been conducted to compare the effects of iACT with active control groups. Future high-quality studies that compare iACT with active control conditions are needed to better understand whether iACT is comparable or superior to other evidence-based treatments in process measures. The findings of this review contribute to the literature by showing the direct effects of iACT on ACT processes, which are theorized to foster improvements in mental health outcomes. These synthesized findings support the processes of change in iACT, which mental health practitioners can use to support the use of iACT.

Acknowledgments

This research was supported by a grant from the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute and funded by the Ministry of Health and Welfare, Republic of Korea (grant number HR21C0885040021). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

AH was responsible for the conceptualization, methodology, formal analysis, investigation, resources, data curation, writing the original draft, visualization, and project administration. THK was responsible for the conceptualization, validation, resources, writing the review and editing, supervision, project administration, and funding acquisition. All the authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms used in database searches.
[DOCX File , 15 KB - jmir_v24i8e39182_app1.docx]

Multimedia Appendix 2
Characteristics of the included studies.
[DOCX File , 41 KB - jmir_v24i8e39182_app2.docx]

Multimedia Appendix 3
Forest plots showing effects of internet-based acceptance and commitment therapy on psychological flexibility according to the use of therapist guidance at the immediate posttest.
[DOCX File , 46 KB - jmir_v24i8e39182_app3.docx]

Multimedia Appendix 4
Forest plots showing effects of internet-based acceptance and commitment therapy on psychological flexibility according to the use of therapist guidance at follow-up.
[DOCX File , 36 KB - jmir_v24i8e39182_app4.docx]

Multimedia Appendix 5
Forest plots showing effects of internet-based acceptance and commitment therapy on mindfulness according to the use of therapist guidance at the immediate posttest.
[DOCX File , 36 KB - jmir_v24i8e39182_app5.docx]

Multimedia Appendix 6
Forest plots showing effects of internet-based acceptance and commitment therapy on valued living according to the use of therapist guidance at the immediate posttest.
[DOCX File , 34 KB - jmir_v24i8e39182_app6.docx]

Multimedia Appendix 7
Forest plots showing effects of internet-based acceptance and commitment therapy on valued living according to the use of therapist guidance at follow-up.
[DOCX File , 33 KB - jmir_v24i8e39182_app7.docx]

Multimedia Appendix 8
Forest plots showing effects of internet-based acceptance and commitment therapy on psychological flexibility according to delivery modes at the immediate posttest.
[DOCX File , 51 KB - jmir_v24i8e39182_app8.docx]

Multimedia Appendix 9
Forest plots showing effects of internet-based acceptance and commitment therapy on mindfulness according to delivery modes at the immediate posttest.
[DOCX File , 40 KB - jmir_v24i8e39182_app9.docx]

Multimedia Appendix 10
Forest plots showing effects of internet-based acceptance and commitment therapy on psychological flexibility according to the use of targeted distressed participants at the immediate posttest.
[DOCX File , 48 KB - jmir_v24i8e39182_app10.docx]

Multimedia Appendix 11
Forest plots showing effects of internet-based acceptance and commitment therapy on psychological flexibility according to the use of targeted distressed participants at follow-up.
[DOCX File , 37 KB - jmir_v24i8e39182_app11.docx]

Multimedia Appendix 12
Forest plots showing effects of internet-based acceptance and commitment therapy on mindfulness according to the use of targeted distressed participants at the immediate posttest.
[DOCX File , 37 KB - jmir_v24i8e39182_app12.docx]

Multimedia Appendix 13
Forest plots showing effects of internet-based acceptance and commitment therapy on mindfulness according to the use of targeted distressed participants at follow-up.
[DOCX File , 34 KB - jmir_v24i8e39182_app13.docx]

Multimedia Appendix 14
Forest plots showing effects of internet-based acceptance and commitment therapy on valued living according to the use of targeted distressed participants at the immediate posttest.
[DOCX File , 36 KB - jmir_v24i8e39182_app14.docx]

Multimedia Appendix 15
Forest plots showing effects of internet-based acceptance and commitment therapy on valued living according to the use of targeted distressed participants at follow-up.
[DOCX File , 34 KB - jmir_v24i8e39182_app15.docx]

Multimedia Appendix 16
Forest plots showing effects of internet-based acceptance and commitment therapy on cognitive defusion according to the use of targeted distressed participants at the immediate posttest.
[DOCX File , 34 KB - jmir_v24i8e39182_app16.docx]

Multimedia Appendix 17
Risk of bias of the included studies.
[DOCX File , 20 KB - jmir_v24i8e39182_app17.docx]

Multimedia Appendix 18
Risk of bias summary chart.
[DOCX File , 41 KB - jmir_v24i8e39182_app18.docx]

Multimedia Appendix 19
Risk of bias graph.
[DOCX File , 20 KB - jmir_v24i8e39182_app19.docx]

Multimedia Appendix 20
Funnel plot: Psychological flexibility and Mindfulness.
[DOCX File , 31 KB - jmir_v24i8e39182_app20.docx]

References

1. Dindo L, Van Liew JR, Arch JJ. Acceptance and commitment therapy: a transdiagnostic behavioral intervention for mental health and medical conditions. Neurotherapeutics 2017 Jul;14(3):546-553 [FREE Full text] [doi: 10.1007/s13311-017-0521-3] [Medline: 28271287]
2. Hayes SC, Pistorello J, Levin ME. Acceptance and commitment therapy as a unified model of behavior change. Couns Psychol 2012 Sep 26;40(7):976-1002 [FREE Full text] [doi: 10.1177/0011000012460836]
3. Kabat-Zinn J. Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life. New York, NY, USA: Hyperion; 2009.

4. Ii T, Sato H, Watanabe N, Kondo M, Masuda A, Hayes SC, et al. Psychological flexibility-based interventions versus first-line psychosocial interventions for substance use disorders: systematic review and meta-analyses of randomized controlled trials. J Context Behav Sci 2019 Jul;13:109-120 [FREE Full text] [doi: 10.1016/j.jcbs.2019.07.003]
5. Akbari M, Mohamadkhani S, Zarghami F. The mediating role of cognitive fusion in explaining the relationship between emotional dysregulation with anxiety and depression: a transdiagnostic factor. Iran J Psychiatry Behav Sci 2016;22(1):17-29 [FREE Full text]
6. Han A, Yuen HK, Lee HY, Zhou X. Effects of acceptance and commitment therapy on process measures of family caregivers: a systematic review and meta-analysis. J Context Behav Sci 2020 Oct;18:201-213 [FREE Full text] [doi: 10.1016/j.jcbs.2020.10.004]
7. Fledderus M, Bohlmeijer ET, Fox JP, Schreurs KM, Spinhoven P. The role of psychological flexibility in a self-help acceptance and commitment therapy intervention for psychological distress in a randomized controlled trial. Behav Res Ther 2013 Mar;51(3):142-151. [doi: 10.1016/j.brat.2012.11.007] [Medline: 23337183]
8. Levin ME, Krafft J, Hicks ET, Pierce B, Twohig MP. A randomized dismantling trial of the open and engaged components of acceptance and commitment therapy in an online intervention for distressed college students. Behav Res Ther 2020 Mar;126:103557. [doi: 10.1016/j.brat.2020.103557] [Medline: 32014692]
9. Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, et al. Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. Behav Ther 2011 Dec;42(4):676-688. [doi: 10.1016/j.beth.2011.03.007] [Medline: 22035996]
10. Butler J, Ciarrochi J. Psychological acceptance and quality of life in the elderly. Qual Life Res 2007 May;16(4):607-615. [doi: 10.1007/s11136-006-9149-1] [Medline: 17268930]
11. Bardeen JR, Fergus TA. The interactive effect of cognitive fusion and experiential avoidance on anxiety, depression, stress and posttraumatic stress symptoms. J Context Behav Sci 2016 Jan;5(1):1-6. [doi: 10.1016/j.jcbs.2016.02.002]
12. Gillanders DT, Sinclair AK, MacLean M, Jardine K. Illness cognitions, cognitive fusion, avoidance and self-compassion as predictors of distress and quality of life in a heterogeneous sample of adults, after cancer. J Context Behav Sci 2015 Oct;4(4):300-311. [doi: 10.1016/j.jcbs.2015.07.003]
13. Krafft J, Haeger JA, Levin ME. Comparing cognitive fusion and cognitive reappraisal as predictors of college student mental health. Cogn Behav Ther 2019 May;48(3):241-252. [doi: 10.1080/16506073.2018.1513556] [Medline: 30230428]
14. Pinto-Gouveia J, Dinis A, Gregório S, Pinto AM. Concurrent effects of different psychological processes in the prediction of depressive symptoms – the role of cognitive fusion. Curr Psychol 2018 Jan 8;39(2):528-539. [doi: 10.1007/s12144-017-9767-5]
15. Pagnini F, Cavalera C, Rovaris M, Mendozzi L, Molinari E, Phillips D, et al. Longitudinal associations between mindfulness and well-being in people with multiple sclerosis. Int J Clin Health Psychol 2019 Jan;19(1):22-30 [FREE Full text] [doi: 10.1016/j.ijchp.2018.11.003] [Medline: 30619494]
16. Romero-Moreno R, Gallego-Alberto L, Márquez-González M, Losada A. Psychometric properties of the Valued Living Questionnaire Adapted to Dementia Caregiving. Aging Ment Health 2017 Sep;21(9):983-990. [doi: 10.1080/13607863.2016.1191055] [Medline: 27253426]
17. Bai Z, Luo S, Zhang L, Wu S, Chi I. Acceptance and Commitment Therapy (ACT) to reduce depression: a systematic review and meta-analysis. J Affect Disord 2020 Jan 01;260:728-737. [doi: 10.1016/j.jad.2019.09.040] [Medline: 31563072]
18. Han A, Yuen HK, Jenkins J. Acceptance and commitment therapy for family caregivers: a systematic review and meta-analysis. J Health Psychol 2021 Jan;26(1):82-102. [doi: 10.1177/1359105320941217] [Medline: 32659142]
19. Hughes LS, Clark J, Colclough JA, Dale E, McMillan D. Acceptance and commitment therapy (ACT) for chronic pain: a systematic review and meta-analyses. Clin J Pain 2017 Jun;33(6):552-568. [doi: 10.1097/AJP.0000000000000425] [Medline: 27479642]
20. 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]
21. Brown M, Glendenning A, Hoon AE, John A. Effectiveness of web-delivered acceptance and commitment therapy in relation to mental health and well-being: a systematic review and meta-analysis. J Med Internet Res 2016 Aug 24;18(8):e221 [FREE Full text] [doi: 10.2196/jmir.6200] [Medline: 27558740]
22. Thompson EM, Destree L, Albertella L, Fontenelle LF. Internet-based acceptance and commitment therapy: a transdiagnostic systematic review and meta-analysis for mental health outcomes. Behav Ther 2021 Mar;52(2):492-507. [doi: 10.1016/j.beth.2020.07.002] [Medline: 33622516]
23. 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]
24. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. London, UK: The Cochrane Collaboration; 2008.

25. Barrett K, Stewart I. A preliminary comparison of the efficacy of online Acceptance and Commitment Therapy (ACT) and Cognitive Behavioural Therapy (CBT) stress management interventions for social and healthcare workers. Health Soc Care Community 2021 Jan;29(1):113-126. [doi: 10.1111/hsc.13074] [Medline: 32613644]
26. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. Nicotine Tob Res 2013 Oct;15(10):1756-1764 [FREE Full text] [doi: 10.1093/ntr/ntt056] [Medline: 23703730]
27. Heffner JL, Kelly MM, Waxmonsky J, Mattocks K, Serfozo E, Bricker JB, et al. Pilot randomized controlled trial of Web-delivered acceptance and commitment therapy versus smokefree.gov for smokers with bipolar disorder. Nicotine Tob Res 2020 Aug 24;22(9):1543-1552 [FREE Full text] [doi: 10.1093/ntr/ntz242] [Medline: 31883336]
28. Lappalainen P, Langrial S, Oinas-Kukkonen H, Muotka J, Lappalainen R. ACT for sleep - Internet-delivered self-help ACT for sub-clinical and clinical insomnia: a randomized controlled trial. J Context Behav Sci 2019 Apr;12:119-127. [doi: 10.1016/j.jcbs.2019.04.001]
29. Levin ME, Pistorello J, Seeley JR, Hayes SC. Feasibility of a prototype web-based acceptance and commitment therapy prevention program for college students. J Am Coll Health 2014;62(1):20-30 [FREE Full text] [doi: 10.1080/07448481.2013.843533] [Medline: 24313693]
30. Levin ME, Hayes SC, Pistorello J, Seeley JR. Web-based self-help for preventing mental health problems in universities: comparing acceptance and commitment training to mental health education. J Clin Psychol 2016 Mar;72(3):207-225. [doi: 10.1002/jclp.22254] [Medline: 26784010]
31. Sagon AL, Danitz SB, Suvak MK, Orsillo SM. The Mindful Way through the Semester: evaluating the feasibility of delivering an acceptance-based behavioral program online. J Context Behav Sci 2018 Jul;9:36-44 [FREE Full text] [doi: 10.1016/j.jcbs.2018.06.004]
32. Viskovich S, Pakenham KI. Randomized controlled trial of a web-based Acceptance and Commitment Therapy (ACT) program to promote mental health in university students. J Clin Psychol 2020 Jun;76(6):929-951. [doi: 10.1002/jclp.22848] [Medline: 31468528]
33. Lappalainen P, Kaipainen K, Lappalainen R, Hoffrén H, Myllymäki T, Kinnunen ML, et al. Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. JMIR Res Protoc 2013 Jan 09;2(1):e1 [FREE Full text] [doi: 10.2196/resprot.2389] [Medline: 23611946]
34. Lappalainen R, Lappalainen P, Puolakanaho A, Hirvonen R, Eklund K, Ahonen T, et al. The Youth Compass -the effectiveness of an online acceptance and commitment therapy program to promote adolescent mental health: a randomized controlled trial. J Context Behav Sci 2021 Apr;20:1-12 [FREE Full text] [doi: 10.1016/j.jcbs.2021.01.007]
35. O'Connor M, Whelan R, Bricker J, McHugh L. Randomized controlled trial of a smartphone application as an adjunct to acceptance and commitment therapy for smoking cessation. Behav Ther 2020 Jan;51(1):162-177. [doi: 10.1016/j.beth.2019.06.003] [Medline: 32005334]
36. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. Behav Res Ther 2016 Mar;78:30-42. [doi: 10.1016/j.brat.2016.01.001] [Medline: 26848517]
37. Scott W, Chilcot J, Guildford B, Daly-Eichenhardt A, McCracken L. Feasibility randomized-controlled trial of online Acceptance and Commitment Therapy for patients with complex chronic pain in the United Kingdom. Eur J Pain 2018 Apr 28:1473-1484. [doi: 10.1002/ejp.1236] [Medline: 29704880]
38. van Aubel E, Bakker JM, Batink T, Michielse S, Goossens L, Lange I, et al. Blended care in the treatment of subthreshold symptoms of depression and psychosis in emerging adults: a randomised controlled trial of Acceptance and Commitment Therapy in Daily-Life (ACT-DL). Behav Res Ther 2020 May;128:103592. [doi: 10.1016/j.brat.2020.103592] [Medline: 32146218]
39. Scott W, Guildford BJ, Badenoch J, Driscoll E, Chilcot J, Norton S, et al. Feasibility randomized-controlled trial of online acceptance and commitment therapy for painful peripheral neuropathy in people living with HIV: the OPEN study. Eur J Pain 2021 Aug;25(7):1493-1507. [doi: 10.1002/ejp.1762] [Medline: 33711209]
40. Weineland S, Arvidsson D, Kakoulidis TP, Dahl J. Acceptance and commitment therapy for bariatric surgery patients, a pilot RCT. Obes Res Clin Pract 2012;6(1):e1-90. [doi: 10.1016/j.orcp.2011.04.004] [Medline: 24331170]
41. Chapoutot M, Peter-Derex L, Schoendorff B, Faivre T, Bastuji H, Putois B. Telehealth-delivered CBT-I programme enhanced by acceptance and commitment therapy for insomnia and hypnotic dependence: a pilot randomized controlled trial. J Sleep Res 2021 Feb;30(1):e13199 [FREE Full text] [doi: 10.1111/jsr.13199] [Medline: 33020985]
42. Muscara F, McCarthy MC, Rayner M, Nicholson JM, Dimovski A, McMillan L, et al. Effect of a videoconference-based online group intervention for traumatic stress in parents of children with life-threatening illness: a randomized clinical trial. JAMA Netw Open 2020 Jul 01;3(7):e208507 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8507] [Medline: 32735335]
43. Twohig MP, Petersen JM, Fruge J, Ong CW, Barney JL, Krafft J, et al. A pilot randomized controlled trial of online-delivered ACT-enhanced behavior therapy for trichotillomania in adolescents. Cogn Behav Pract 2021 Nov;28(4):653-668 [FREE Full text] [doi: 10.1016/j.cbpra.2021.01.004]

44. Hesser H, Gustafsson T, Lundén C, Henrikson O, Fattahi K, Johnsson E, et al. A randomized controlled trial of Internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. J Consult Clin Psychol 2012 Aug;80(4):649-661. [doi: 10.1037/a0027021] [Medline: 22250855]
45. Hoffmann D, Rask CU, Hedman-Lagerlöf E, Jensen JS, Frostholm L. Efficacy of internet-delivered acceptance and commitment therapy for severe health anxiety: results from a randomized, controlled trial. Psychol Med 2021 Nov;51(15):2685-2695 [FREE Full text] [doi: 10.1017/S0033291720001312] [Medline: 32404226]
46. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-based acceptance and commitment therapy for depressive symptoms with minimal support: a randomized controlled trial. Behav Modif 2015 Nov;39(6):805-834. [doi: 10.1177/0145445515598142] [Medline: 26253644]
47. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017 Jan 01;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
48. Pots WT, Fledderus M, Meulenbeek PA, ten Klooster PM, Schreurs KM, Bohlmeijer ET. Acceptance and commitment therapy as a Web-based intervention for depressive symptoms: randomised controlled trial. Br J Psychiatry 2016 Jan;208(1):69-77. [doi: 10.1192/bjp.bp.114.146068] [Medline: 26250745]
49. Sairanen E, Lappalainen R, Lappalainen P, Kaipainen K, Carlstedt F, Anclair M, et al. Effectiveness of a web-based Acceptance and Commitment Therapy intervention for wellbeing of parents whose children have chronic conditions: a randomized controlled trial. J Context Behav Sci 2019 Jul;13:94-102. [doi: 10.1016/j.jcbs.2019.07.004]
50. Köhle N, Drossaert CH, Ten Klooster PM, Schreurs KM, Hagedoorn M, Van Uden-Kraan CF, et al. Web-based self-help intervention for partners of cancer patients based on acceptance and commitment therapy and self-compassion training: a randomized controlled trial with automated versus personal feedback. Support Care Cancer 2021 Sep;29(9):5115-5125 [FREE Full text] [doi: 10.1007/s00520-021-06051-w] [Medline: 33608762]
51. Levin ME, Petersen JM, Durward C, Bingeman B, Davis E, Nelson C, et al. A randomized controlled trial of online acceptance and commitment therapy to improve diet and physical activity among adults who are overweight/obese. Transl Behav Med 2021 Jun 17;11(6):1216-1225. [doi: 10.1093/tbm/ibaa123] [Medline: 33289785]
52. Lin J, Paganini S, Sander L, Lüking M, Ebert DD, Buhrman M, et al. An Internet-based intervention for chronic pain. Dtsch Arztebl Int 2017 Oct 13;114(41):681-688 [FREE Full text] [doi: 10.3238/arztebl.2017.0681] [Medline: 29082858]
53. Douma M, Maurice-Stam H, Gorter B, Krol Y, Verkleij M, Wiltink L, et al. Online psychosocial group intervention for parents: positive effects on anxiety and depression. J Pediatr Psychol 2021 Feb 19;46(2):123-134 [FREE Full text] [doi: 10.1093/jpepsy/jsaa102] [Medline: 33230541]
54. Eustis EH, Hayes-Skelton SA, Orsillo SM, Roemer L. Surviving and thriving during stress: a randomized clinical trial comparing a brief web-based therapist-assisted acceptance-based behavioral intervention versus waitlist control for college students. Behav Ther 2018 Nov;49(6):889-903. [doi: 10.1016/j.beth.2018.05.009] [Medline: 30316488]
55. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KM. Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
56. Simister HD, Tkachuk GA, Shay BL, Vincent N, Pear JJ, Skrabek RQ. Randomized controlled trial of online acceptance and commitment therapy for fibromyalgia. J Pain 2018 Jul;19(7):741-753. [doi: 10.1016/j.jpain.2018.02.004] [Medline: 29481976]
57. Buhrman M, Skoglund A, Husell J, Bergström K, Gordh T, Hursti T, et al. Guided internet-delivered acceptance and commitment therapy for chronic pain patients: a randomized controlled trial. Behav Res Ther 2013 Jun;51(6):307-315. [doi: 10.1016/j.brat.2013.02.010] [Medline: 23548250]
58. French K, Golijani-Moghaddam N, Schröder T. What is the evidence for the efficacy of self-help acceptance and commitment therapy? A systematic review and meta-analysis. J Context Behav Sci 2017 Oct;6(4):360-374 [FREE Full text] [doi: 10.1016/j.jcbs.2017.08.002]
59. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol 2019 Jun 28;19(1):132 [FREE Full text] [doi: 10.1186/s12874-019-0782-0] [Medline: 31253092]

Abbreviations

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

Edited by R Kukafka, G Eysenbach; submitted 02.05.22; peer-reviewed by M Brown, Z Dai; comments to author 17.07.22; revised version received 22.07.22; accepted 29.07.22; published 30.08.22.

Please cite as:
Han A, Kim TH
The Effects of Internet-Based Acceptance and Commitment Therapy on Process Measures: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(8):e39182
URL: https://www.jmir.org/2022/8/e39182
doi:10.2196/39182
PMID:36040783

Review

Publishing Identifiable Patient Photographs in Scientific Journals: Scoping Review of Policies and Practices

Marija Roguljić[1], DMD, PhD; Dina Šimunović[2], DMD; Tina Poklepović Peričić[3,4], DMD, PhD; Marin Viđak[4], MD, PhD; Ana Utrobičić[5], MA; Matko Marušić[4], MD, PhD; Ana Marušić[4], MD, PhD

[1]Department of Oral Medicine and Periodontology, University of Split School of Medicine, Split, Croatia
[2]Private Dental Practice, Split, Croatia
[3]Department of Prosthodontics, Study of Dental Medicine, School of Medicine, University of Split Library, Split, Croatia
[4]Department of Research in Biomedicine and Health, Center for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia
[5]University of Split Library, Split, Croatia

Corresponding Author:
Ana Marušić, MD, PhD
Department of Research in Biomedicine and Health
Center for Evidence-based Medicine
University of Split School of Medicine
Soltanska 2
Split, 21000
Croatia
Phone: 385 21 557 812
Fax: 385 21 557 811
Email: ana.marusic@mefst.hr

Abstract

Background: Publishing identifiable patient data in scientific journals may jeopardize patient privacy and confidentiality if best ethical practices are not followed. Current journal practices show considerable diversity in the publication of identifiable patient photographs, and different stakeholders may have different opinions of and practices in publishing patient photographs.

Objective: This scoping review aimed to identify existing evidence and map knowledge gaps in medical research on the policies and practices of publishing identifiable photographs in scientific articles.

Methods: We performed a comprehensive search of the Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, CINAHL with Full Text, Database of Abstracts of Reviews of Effects, Ovid MEDLINE, and Scopus. The Open Science Framework, PROSPERO, BASE, Google Scholar, OpenGrey, ClinicalTrials.gov, the Campbell Collaboration Library, and Science.gov were also searched.

Results: After screening the initial 15,949 titles and abstracts, 98 (0.61%) publications were assessed for eligibility at the full-text level, and 30 (0.19%) publications were included in this review. The studies were published between 1994 and 2020; most had a cross-sectional design and were published in journals covering different medical disciplines. We identified 3 main topics. The first included ethical aspects of the use of facial photographs in publications. In different clinical settings, the consent process was not conducted properly, and health professionals did not recognize the importance of obtaining written patient consent for taking and using patient medical photographs. They often considered verbal consent sufficient or even used the photographs without consent. The second topic included studies that investigated the practices and use of medical photography in publishing. Both patients and doctors asked for confidential storage and maintenance of medical photographs. Patients preferred to be photographed by their physicians using an institutional camera and preferred nonidentifiable medical photographs not only for publication but also in general. Conventional methods of deidentification of facial photographs concealing the eye area were recognized as unsuccessful in protecting patient privacy. The third topic emerged from studies investigating medical photography in journal articles. These studies showed great diversity in publishing practices regarding consent for publication of medical photographs. Journal policies regarding the consent process and consent forms were insufficient, and existing ethical professional guidelines were not fully implemented in actual practices. Patients' photographs from open-access medical journals were found on public web-based platforms.

Conclusions: This scoping review showed a diversity of practices in publishing identifiable patient photographs and an unsatisfactory level of knowledge of this issue among different stakeholders despite existing standards. Emerging issues include the availability of patients' photographs from open-access journals or preprints in the digital environment. There is a need to improve standards and processes to obtain proper consent to fully protect the privacy of patients in published articles.

(J Med Internet Res 2022;24(8):e37594) doi:10.2196/37594

KEYWORDS

identifiable patient photographs; medical photography; data protection; patient privacy; confidentiality; informed consent; ethical publishing; scientific journals; open access; scoping review; mobile phone

Introduction

Background

Scientific publications are considered to be the most important formal elements of scholarly communication, translating new evidence in practice and increasing relevant stakeholders' knowledge. Publishing identifiable patient photographs in journals, such as photographs of the face, is a challenging ethical issue, not only because it requires consent but also because many research participants as well as researchers are not aware of what happens when identifying photographs of individual persons are published. This is of particular importance in digital publishing, especially when open publishing licenses such as the Creative Commons license CC BY are used [1]. Many medical journals that publish articles identifying patient photographs are in open access and under open licenses [2], which means that identifiable patient data are widely available and can be easily abused [3].

It is not always clear how patient data are classified as identifiable, nonidentifiable, or anonymized. The scoping review by Chevrier et al [4] on the use and understanding of terms of anonymization and deidentification in biomedical literature showed that there is large variability in the use of these as well as the need for clearer definitions and better education. Current publication standards on the use of patient identifiable data proposed by the International Committee of Medical Journal Editors (ICMJE) recommend avoiding publication of identifiable photographs or marks and the necessity of obtaining written consent from the patient [5]. However, in some clinical disciplines such as those involving the head and neck, it is necessary to show the patients' faces to illustrate study findings. The most used methods of deidentification to protect patient identity—covering the eye area in facial photographs—have been recognized as insufficient and should not be published without the patient's written consent [6-8]. Consent for the publication of an identifiable photograph should be obtained after the patient is informed of all the potential consequences of a publication even if the publication results from routine health care and is written up as a case report [9-12]. Patients should also be aware of the impossibility of withdrawing or controlling any future use of photographs once they have been published on the web [1,13]. This means that consent for the publication of an identifying patient photograph is separate from and additional to the general consent for research [2,8,14-16].

Objectives

Despite these recommendations, there is diversity among medical journals in their policies on patient consent for the publication of identifying photographs and their implementation in practice [2,17]. There are also varying opinions and practices among different stakeholders—patients, professionals, journals, and professional societies. To identify the existing evidence and map knowledge gaps in research on the policies and practices of publishing identifiable photographs in medicine, we performed a scoping review of the published literature on this topic. The research question of this scoping review was as follows: what are the opinions, standards, and practices of different stakeholders (patients, health professionals, policy makers, journals, editors, and publishers) regarding consent for publishing potentially identifiable medical photographs?

Methods

Methodological Approach

We used the methodology for scoping reviews from the Joanna Briggs Institute [18]. The protocol of this scoping review was registered at the Open Science Framework [19]. Study results are presented following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [20].

Study Selection (Eligibility Criteria)

We performed a sensitive search without language, time, or geographical limitations to identify studies that investigated the conditions of publication of patient facial photographs regardless of whether they were identifiable or not; articles that addressed only body parts other than the face were excluded. Publications that reported the results of conducted studies were included in the analysis. All other types of publications, such as editorials, opinion letters, reviews, and book chapters, were excluded.

Information Sources and Search

Search strategies for bibliographical databases were designed by an experienced librarian (AU; Multimedia Appendix 2). We searched the Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, CINAHL with Full Text, Database of Abstracts of Reviews of Effects, Ovid MEDLINE, and Scopus in September 2018 and updated the search in December 2020. In January 2021, we also searched registries and gray literature sources: ClinicalTrials.gov, Campbell Collaboration Library, Open Science Framework, PROSPERO, BASE, Google Scholar, OpenGrey, and Science.gov. These sources were searched using variations of the terms “medical” and “photography.” The reference lists of all studies included in the full-text assessment were also searched.

Screening

The retrieved articles were exported to and deduplicated in EndNote (Clarivate Analytics). Owing to the large number of articles for screening using a sensitive strategy that retrieved many nonrelevant articles, 2 authors (MR and TPP) screened the titles and abstracts of separate sets of articles. Articles identified in the screening were then jointly assessed by 2 authors (DŠ and MR), who discussed each article. Independent assessment by the 2 authors was not performed, and agreement indexes were not calculated as reporting in many articles was not always clear and significant disagreement was expected. The 2 assessors reached a joint conclusion on the inclusion of an article during their discussion based on the article’s full text.

Data Charting Process

Two authors (MR and MV) created the charting form for the variables to be extracted. The form was reviewed by another author (AM) and tested by 2 authors (DŠ and MR), who extracted the data for the first 10 articles and discussed the coding for each variable. They confirmed the inclusion and exclusion criteria and then each independently extracted the data for half of the articles in the final sample. The data from 2 articles authored by some of the authors of this review were collected by the author who did not participate in the study (MV). AM checked the quality of data extraction.

Data Items

Data were collected for the following variables: authors, article title, year of publication, source origin and country of origin, World Bank ranking of the country, publication type, journal title, journal access status, study design, study population, setting, sample size and response rate, age of the participants, gender, aim of the study, methodology, key outcomes, philosophical approach, key findings, limitations, future study ideas, and recommendations.

Summarizing Data and Reporting Results

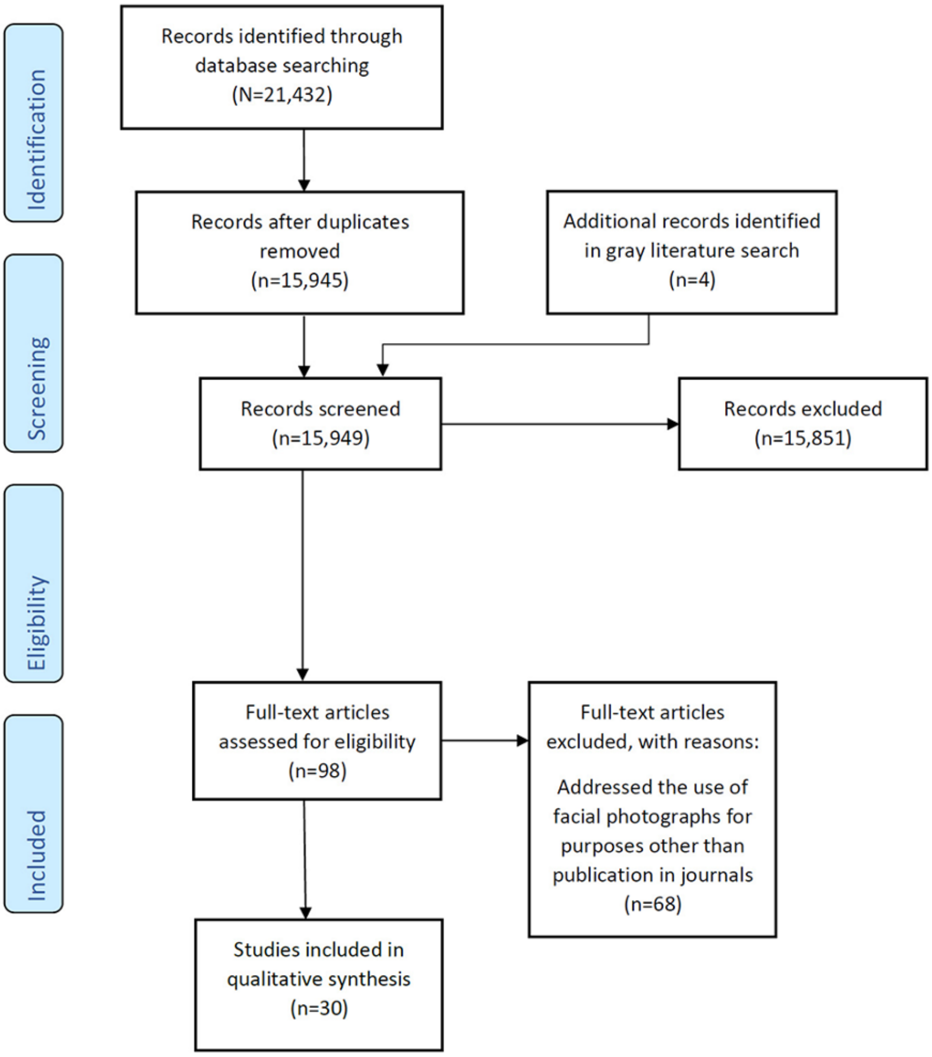
We summarized the data quantitatively for the description of the included studies. In the qualitative analysis, we grouped the results of the studies into main themes. According to the PRISMA-ScR guidance, we did not formally assess the methodological quality, including risk of bias, of the studies from which data were extracted as the scoping review method is not intended to be used to appraise the risk of bias of a cumulative body of evidence [20].

Results

Selection of Sources of Evidence

The search of bibliographical database literature retrieved a total of 21,432 published items, leaving 15,945 (74.4%) items after deduplication. The search of registries and gray literature yielded 4 additional items. After screening titles and abstracts, of the 15,949 items, 98 (0.61%) references were screened at the full-text level. We excluded 69% (68/98) of the studies as they addressed uses of facial photographs other than for publication in journals. This left 31% (30/98) of articles for analysis [2,3,6-12,14-16,21-38]. The flow diagram of the literature review is shown in Figure 1.

Figure 1. Flowchart of the literature review.



Characteristics of the Sources of Evidence

Of the 30 articles included in the analysis, 6 (20%) investigated the publication of facial photographs in scientific journals as the main topic [2,3,7,32,35,36]. Other studies (24/30, 80%) had the publication of medical photography as one of the topics, so only those results were included in this review.

The included studies were published between 1994 and 2020. Almost half of the studies (13/30, 43%) were published in open access [3,6,7,11,15,21-23,25,30,31,36]. The studies were performed in countries from all continents, from both high– and low–research-intensive countries. Of the 30 studies, 29 (97%) were published in English, and 1 (3%) was published in French. Most of the studies (5/30, 17%) were conducted internationally in scientific journals, followed by studies from the United Kingdom (5/30, 17%); the United States (4/30, 13%); Australia (3/30, 10%); Brazil, Canada, France, and India (2/30, 7% each); and China, Croatia, Ireland, Nigeria, and Saudi Arabia (1/30, 3% each). The studies were published in journals from different disciplines: ethics (4/30, 13%), surgery (10/30, 33%), general medicine and education (3/30, 10%), and clinical dermatology (13/30, 43%). Most of the studies (26/30, 87%) had an observational or cross-sectional design (Table 1), 3% (1/30) were randomized controlled trials [6], 3% (1/30) used another experimental design [28], and 7% (2/30) used qualitative approaches [23,25]. Reported limitations were the small number of participants [8,12,26,27,34], pilot studies [22], a single type of health professional included [6,8,9,25,29], years of clinical practice for specialists [9], poor response rate [26,27,33,34], possibility of socially desirable answers [29,35], and a single study setting [21,22,25,29]. The limitations of the studies performed on data or journals were the small number of journals [7,36], small number of analyzed journal issues [2], filters for web-based image search, and the fluctuating number of available web-based images from academic journals [3].

There were 25 studies involving different stakeholders (Table 1): patients (n=11, 44%), legal representatives (parents) of minors (n=1, 4%), undergraduate and postgraduate students of medicine and dental medicine (n=5, 20%), nurses and other hospital health professionals (n=1, 4%), and medical doctors and doctors of dental medicine (n=11, 44%; Table 1). Another 17% (5/30) of studies involved editorial policies or published articles in journals (Table 1). The studies that involved human participants were conducted in clinical settings (17/30, 57%), at universities (3/30, 10%), and on the web (5/30, 17%). The median number of study participants was 153 (range 12-945), and the median response rate was 70% (range 17%-90%). All participants were adults, with a median age of 40 (range 27-57) years. The median percentage of women involved in the studies was 55% (range 33%-84%). The remaining 17% (5/30) of studies analyzed data on journals and articles [2,3,7,32,36].

The studies included in this review addressed three general topics: (1) ethical aspects of the use of medical photography in publications (Table 2), (2) practices and use of medical photography in journals (Table 3), and (3) characteristics of medical photographs in published articles (Table 4).

Table 1. Description of the studies included in the scoping review (N=30).

Study, year	Country where the study was performed	Study design	Study population	Setting or data sources	Sample size	Response rate	Age (years)
Jones [14], 1994	United Kingdom	Cross-sectional	Patients	Teaching and university hospitals or large district general hospitals (44 medical illustration departments)	35	79%	NR[a]
Cheung et al [22], 2005	United Kingdom	Cross-sectional	Patients	Emergency department	100	N/A[b]	>18
Windsor et al [38], 2006	Australia	Cross-sectional	Digital images (photographs and video clips)	Adult emergency department in Australia	493	N/A	N/A
Taylor et al [8], 2007	United Kingdom	Cross-sectional	All surgical staff	Plastic surgery units in the hospital	42	70%	NR
Clover et al [6], 2010	Ireland	Randomized controlled trial	Medical students	Medical school	126	96%	NR
Lau et al [15], 2010	United Kingdom	Cross-sectional	Patients	Department of plastic and reconstructive surgery	205	NR	>18
Engelstad et al [28], 2011	United States	Experimental study	Dental students	School of Dentistry	12	NR	Median 27 (range 24-29)
Adeyemo et al [10], 2012	Nigeria	Cross-sectional	Patients	Oral, maxillofacial, and plastic surgery clinics	338	NR	Mean 32.5 (SD 12.2; range 16-79)
Shintani and Williams [36], 2012	International	Cross-sectional	Journals	Medical journals in oral surgery	3	N/A	N/A
Devakumar et al [25], 2013	United Kingdom	Qualitative (focus groups)	Pediatricians	Teleconference via Skype	13	NR	NR
Hacard et al [29], 2013	France	Cross-sectional	Patients	Department of dermatology	272	NR	Adults: mean 57.5 (SD 17.6), children: median 1.5 (IQR 0.6-7.0), and accompanying parents: mean 35.0 (SD 6.8)
Kunde et al [31], 2013	Australia	Cross-sectional	Dermatology registrars and insurance providers	N/A	13	65%	NR
De Runz et al [24], 2014	France	Cross-sectional	Plastic surgeons and patients	Department of maxillofacial, plastic, and esthetic surgery in the hospital	176 surgeons and 93 patients	Surgeons: 42% and patients: NR	NR
Leger et al [12], 2014	United States	Cross-sectional	Patients	Hospital	398	NR	>18
Caires et al [11], 2015	Brazil	Cross-sectional	Nurses, nursing technicians, residents working at inpatient units, and physical therapists	Teaching hospital	360	Nurses: 31.4% and residents: 43.9%; regarding the place of work in the hospital, 3% worked in inpatient units	>40
Indu et al [30], 2015	India	Cross-sectional	Postgraduate students and teaching staff	Oral pathology departments	60	44%	NR

Study, year	Country where the study was performed	Study design	Study population	Setting or data sources	Sample size	Response rate	Age (years)
Rimoin et al [34], 2016	United States	Cross-sectional	Surgeons	Members of the American College of Mohs Surgery	158	17%	NR
Roberts et al [7], 2016	International	Cross-sectional	Journals	Medical journals that frequently publish facial photographs	13	N/A	N/A
Abbott et al [9], 2017	Australia	Cross-sectional	Dermatologists and dermatologic trainers	Australian College of Dermatologists	101	96%	NR
Dumestre and Fraulin [26], 2017	Canada	Cross-sectional	Patients, plastic surgeons, and residents	Section of plastic surgery	86 patients, 3 plastic surgeons, and 12 residents	57% of patients, 67% of surgeons, and 92% of residents	NR
Wang et al [37], 2017	China	Cross-sectional	Patients	Dermatology clinic	474	89%	Mean 31.9 (SD 11.7)
Marshall et al [3], 2018	International	Cross-sectional	Journals	Google Images and open-access articles	94	N/A	N/A
Milam and Leger [33], 2018	United States	Cross-sectional	Dermatologists	Board-certified dermatologists practicing in the United States	107	69%	Mean 47.2 (SD 11.7)
Nair et al [16], 2018	India	Cross-sectional	Patients	Ophthalmic plastic surgery clinic	280	NR	Mean 40.2 (range 18-82)
Dumestre and Fraulin [27], 2020	Canada	Cross-sectional	Plastic surgeons, residents, and patients	Section of plastic surgery	16 plastic surgeons, 24 residents, and 84 patients and parents	51% of surgeons and residents and 56% of patients	NR
Lessing et al [32], 2019	International	Cross-sectional	Journals	Top 10 impact factor general medical journals	10	N/A	N/A
Abouzeid et al [21], 2020	Saudi Arabia	Cross-sectional	Dental students	School of Dentistry	233	86%	NR
Costa et al [23], 2020	Brazil	Qualitative study (semistructured interviews)	Dentists	Unclear (clinical setting)	52	NR	Mean 30.4
Roguljić et al [35], 2020	Croatia	Cross-sectional	Patients, students of medicine and dentistry, and doctors of medicine and dental medicine	Dental outpatient clinics	292 patients, 281 students, and 281 doctors	Patients: NR, physicians: 85%, medical students: 72%, and dental students: 58%	Patients: median 55 (IQR 22), students: median 23 (IQR 1), and physicians: median 40 (IQR 18)
Roguljić et al [2], 2022	International	Cross-sectional	Journals	Medical journals in dentistry and otolaryngology	103	N/A	N/A

[a] NR: not reported.

[b] N/A: not applicable.

Table 2. Ethical aspects of medical photography for publications.

Study, year	Study aim	Key findings
Jones [14], 1994	To determine common practices and attitudes toward medical photography among hospital patients	Most departments insist on written informed consent. When releasing clinical slides for publication, most departments insist that patient consent is obtained.
Cheung et al [22], 2005	To investigate patients' attitudes toward medical photography and consent use at an emergency department	Most participants gave consent for publication of images in a medical journal or books but were more likely to refuse consent for use of images on internet medical sites.
Taylor et al [8], 2007	To investigate awareness of and compliance with present regulations regarding clinicians taking digital photographs of patients	Less than half of the surgeons reported always obtaining consent, more often verbal than written for different purposes. The process of consent must include the option that consent may be withdrawn at any time before the information has passed irretrievably into the public domain.
Lau et al [15], 2010	To explore patient perception of digital photography	Approximately half of the patients would consent for each purpose of use.
Adeyemo et al [10], 2012	To determine acceptance and perception of medical photography among Nigerian patients	Most respondents indicated that their consent should be sought for each purpose.
Kunde et al [31], 2012	To review ethical and legal considerations of clinical photography in dermatology and present a hypothetical medicolegal scenario	Verbal consent would be commonly obtained for different purposes, including publication.
Devakumar et al [25], 2013	To explore the issues around photography in low-resource settings by conducting discussions with medical doctors and researchers who are currently working or have recently worked in low-resource settings with children	Participants considered that informed consent is required, but its form may vary depending on the context. Protection of the rights of children is especially important in relation to photographs.
Hacard et al [29], 2013	To evaluate patients' perceptions of medical photographs	Written consent was considered necessary for adult and pediatric patients.
De Runz et al [24], 2014	To analyze the use of photography by plastic surgeons, the perception of this use by the patients, and medicolegal and ethical consequences	Most of the surgeons considered that verbal consent or no consent is sufficient for taking patient photographs.
Leger et al [12], 2014	To investigate patient opinions on clinical photography	Respondents preferred permission for photographs to be obtained in written form.
Caires et al [11], 2015	To evaluate the knowledge of health care professionals regarding taking medical photographs within the hospital environment among hospital staff	Verbal and written consent for taking the photographs was lacking.
Indu et al [30], 2015	To assess the awareness of oral pathologists regarding various aspects of medical photographs	Most students and faculty members informed the patients of the purpose of the photograph and took verbal consent. Most of them mentioned to the patient their right to withdraw consent.
Rimoin et al [34], 2016	To elucidate the nature of use, storage, and informed consent for digital photography among Mohs surgeons	A very small number of responders pursued some form of consent before taking photographs, with most preferring verbal consent over written consent. They considered that consent should be obtained for different purposes.
Abbott et al [9], 2017	To evaluate the understanding of the use of smartphones in clinical practice regarding professional and legal risks	Patient consent was not often documented regarding different uses of patient photographs; respondents mostly did not receive information on relevant guidelines.
Dumestre and Fraulin [26], 2017	To evaluate a smartphone app for clinical photography regarding patient security among plastic surgeons, plastic surgery residents, and patients who had undergone plastic surgery	The app ensured adequate consent for educational and research purposes but was inadequate for publication and disclosure to the public.
Wang et al [37], 2017	To assess the perception and acceptability of medical photography in patients of dermatology	Almost half of the respondents considered that oral consent only should be obtained before taking medical photographs, whereas the other half of the respondents answered that written consent should be obtained. Most of the respondents argued that all possible image uses should be detailed in the consent form.
Milam and Leger [33], 2018	To examine dermatologists' current practices in medical photography	Most respondents agreed that patients should be allowed to withdraw consent and should be informed of the use of their photographs on each occasion, including publication.

Study, year	Study aim	Key findings
Dumestre and Fraulin [27], 2020	To evaluate a smartphone app for clinical photography that prioritizes and facilitates patient security	Patients considered the consent process acceptable in the app. Surgeons and residents felt that the consent process was superior or equivalent to previous methods.
Costa et al [23], 2020	To evaluate the behavior of dentists on the use of patients' images	Participants considered that informed consent for sharing patients' images, including in publications, can be verbal or absent when the patient cannot be identified.
Roguljić et al [35], 2020	To explore opinions of patients, students, and doctors on the acceptability of different levels of deidentification and the informed consent needed for publication in academic journals	All respondents reported increased preference for more stringent forms of permission as the level of identifiability in photographs increased.

Table 3. Practices and use of medical photography for research publications.

Study, year	Study aim	Key findings
Jones [14], 1994	To determine common practices and attitudes toward medical photography among hospital patients	Most respondents felt that, even though the patient was consenting to treatment by being in hospital, they still had a right to refuse to be photographed.
Windsor et al [38], 2006	To summarize 3 months of digital photography taking in an adult emergency department	The use of digital photographs and video clips in clinical settings is very useful in creating a database of confidential medical records that can be used for medical teaching and publication.
Taylor et al [8], 2007	To investigate awareness of and compliance with present regulations regarding clinicians taking digital photographs of patients	Patients considered themselves insufficiently informed of their right to withdraw consent. Surgeons used methods of deidentification for patient photographs for teaching and publication purposes. They stored password-protected photographs in PCs and personal cameras.
Clover et al [6], 2010	To analyze the effectiveness of blacking out the eyes in facial photographs through alternative techniques	Deidentification failed most in the group with a covered eye area in a photograph, followed by covering the eye and nose and covering the eyes, nose, and mouth.
Lau et al [15], 2010	To explore patients' perception of digital photography	Patients preferred the use of hospital cameras and nonidentifiable photographs for all purposes.
Engelstad et al [28], 2011	To test the hypothesis that unaltered features from an original full-face patient image could be blended with other facial images to create a unique facial composite that deidentifies the patient	Facial composites were more effective at deidentification than traditional methods.
Adeyemo et al [10], 2012	To determine acceptance and perception of medical photography among Nigerian patients	Patients had high acceptance of medical photography, especially of nonidentifiable photographs. The use of institutional cameras operated and stored by the patients' physicians was the preferred method.
Kunde et al [31], 2012	To explore ethical and legal considerations of clinical photography in dermatology and present a hypothetical medicolegal scenario	Dermatologic registrars used personal smartphones for taking photographs for different purposes, such as to obtain advice from peers, teaching, sharing with colleagues, treatment and disease monitoring, and publication.
Devakumar et al [25], 2013	To explore the issues around photography in low-resource settings	Photographs of children in medical and research settings are useful as they enrich teaching, research, and advocacy.
Hacard et al [29], 2013	To evaluate patients' perceptions of medical photography	Low acceptability of the use of the images in professional emails, health magazines, television health programs, and medical websites. Publication in medical scientific articles was significantly more acceptable for adults than for children.
De Runz et al [24], 2014	To analyze the use of medical photography by plastic surgeons and perception of this use by the patients	Patients and surgeons had high acceptance of taking medical photographs for diagnosis and treatment follow-up and lower acceptance for publication purposes.
Leger et al [12], 2014	To investigate patients' opinions of clinical photography	Nonidentifiable photographs taken by their physician with clinic-owned cameras within the institution for all purposes were preferred. Race and ethnicity, income level, and age influenced the patients' answers.
Abbott et al [9], 2017	To evaluate the understanding of the use of smartphones in clinical practice regarding professional and legal risks	Most respondents had and used smartphones for taking medical photographs for different purposes.
Dumestre and Fraulin [26], 2017	To evaluate a smartphone app for clinical photography regarding patient security among plastic surgeons, plastic surgery residents, and patients who had undergone plastic surgery	Patients: high acceptance of use for educational, research, communication, and medical documentation purposes and less acceptance for publication in a public medium; surgeons and residents: the app will be suitable for use when certain issues regarding consent and protection of confidentiality are overcome
Wang et al [37], 2017	To assess the perception and acceptability of medical photography in patients of dermatology	Patients' physicians using clinic-owned cameras were the most accepted as photographers. Low acceptability of use was reported for medical websites and televised health programs.
Milam and Leger [33], 2018	To examine the current medical photography practices of dermatologists	Respondents reported the use of medical photographs for different purposes, including research and publication. They used digital cameras, personal phones, and electronic medical record applications. Photographs were stored in the office computer with various security measures and shared via email with colleagues and patients.
Nair et al [16], 2018	To assess patient perceptions regarding medical photography and smart devices	Most patients accepted the use of smartphones for medical photography, but only a third approved the use of medical photographs in presentations and medical journals. Patients preferred to be photographed by their physician with their own camera or an institutional camera at the institution.

Study, year	Study aim	Key findings
Dumestre and Fraulin [27], 2020	To evaluate a smartphone app for clinical photography that prioritizes and facilitates patient security	Patients: the purpose of the app was well explained, and it was perceived as safe; surgeons and residents: respondents believed the app was suitable for broad implementation to receive and send patient photographs
Abouzeid et al [21], 2020	To evaluate the awareness of practice, opportunity, and morals of dental photography among undergraduate dental students	Almost all students take photographs on a regular basis. Phone cameras were the most commonly used device, followed by digital single-lens reflex cameras. Verbal consent was obtained before taking photographs. For research publication, they edited the photographs using specific software or by covering the eye area. More training in photography techniques was perceived to be necessary.
Costa et al [23], 2020	To evaluate the behavior of dentists in using patients' images	The most common purposes of the use of photographs were didactic or academic. Discussion groups on social media may increase the knowledge of the use of patient photographs.

Table 4. Medical photography in research publications.

Study, year	Study aim	Key findings
Shintani and Williams [36], 2012	To investigate how guidelines on the protection of patient anonymity are actually implemented and how effective such methods of protection are in 3 oral surgery journals	Most of the published photographs were of the entire face or a part of the face. Masking the eye area was observed in half of the facial photographs, and deidentification failed.
Marshall et al [3], 2018	To analyze current practices used in patient facial photograph deidentification	Sensitive medical photographs from articles freely available were found on Google Images. A small number of articles reported obtaining written informed consent for publication of medical images from patients undergoing transgender surgery.
Roberts et al [7], 2016	To analyze the current practices used in patient facial photograph deidentification and set forth standardized guidelines for improving patient autonomy that are congruent with medical ethics and health insurance	Facial image anonymization guidelines varied across journals. When anonymization was attempted, 87% of the images were inadequately concealed. The most common technique used was masking the eyes alone with a black box.
Lessing et al [32], 2019	To assess consent requirements in a sample of 10 top impact factor general medicine journals that publish clinical images	All journals had web-based information regarding clinical image consent requirements. Written consent was required for all identifiable photographs. No journals were fully compliant with ICMJE[a] consent recommendations.
Roguljić et al [2], 2022	To analyze policies of journals that publish research and their implementation regarding patient consent for facial image publication	Only approximately half of the analyzed journals had a specific policy on clinical images. A small number of articles that published recognizable patient facial images included a statement on consent for image publication.

[a] ICMJE: International Committee of Medical Journal Editors.

Ethical Aspects of Medical Photography for Research Publications

Almost all studies that analyzed the ethical aspects of medical photography in research publications (19/20, 95%; Table 2) reported that the consent process was not conducted properly for different uses of patients' photographs, including for journal publications. A total of 5% (1/20) of the studies stated that they addressed medicolegal issues [31], but the study findings were not put in the context of privacy protection legal regulations. Studies that included patients and health professionals (13/20, 65%) were affirmative of the practice of obtaining informed consent for the use of patient medical photographs. However, relevant stakeholders recognized different levels of potential risks if patient medical photographs were used for different purposes, from treatment planning and follow-up in medical documentation to education and different forms of publication.

Studies that investigated patients' perspectives on the importance of informed consent (8/20, 40%) showed that patients were aware of the increased risks of being recognized after the publication of their medical photographs by anyone who has access to the publication [10,12,14-16,22,35,37]. Patients in an emergency medicine department were more likely to refuse consent for the use of images on internet sites, but they would provide consent for the purposes of medical education, medical books, or journals [22]. Patients in that study were not aware that medical books or journals could also be accessed on internet sites [22]. Patients in a plastic surgery department were more likely to approve the use of medical photography for diagnosis and treatment follow-up but were also less likely to consent for publication purposes [24]. A total of 10% (2/20) of the studies showed that patients preferred to be offered consent for a specific purpose and not a general consent for any type of use of their medical photographs, including identifiable and nonidentifiable photographs [10,14]. Furthermore, patients preferred being offered a written consent form rather than being offered oral consent [12,14,29].

Studies that involved plastic surgeons and dentists showed diversity in written informed consent for taking photographs of patients. A total of 17% (5/30) of the studies showed that patient consent was not always obtained for taking and using patient medical photographs and that the prevalent opinion was that verbal consent was sufficient [3,8,23,24,26]. Studies that involved dermatologists (2/20, 10%) [9,33] showed that most of them did obtain consent for patients' photographs, but they did not consider it necessary. In addition, dermatologists emphasized the need for better education on this issue and the need to create more realistic and practical policies for everyday practice. They also asked for better policies and tools for patients to exercise their right to withdraw their consent at a later time.

A small number of studies (4/20, 20%) investigated other health professionals' opinions regarding ethical publishing of medical photography, involving residents, students, nurses, nursing technicians, and physical therapists [11,30,31,35]. These studies reported that health professionals in general had a lack of knowledge regarding the need to obtain patient written consent and the use of patient medical photographs in general. A survey of nursing staff, physical therapists, and physicians reported a lack of knowledge of both verbal and written consent for taking patient photographs [11], whereas undergraduate and postgraduate students considered that verbal consent was sufficient for medical image publication [30,35]. Similarly, in the study by Kunde et al [31], 4 out of 13 dermatology registrars reported that they used verbal consent for taking photographs of patients for publishing purposes.

In total, 7% (2/30) of the studies investigated the issues of taking and using medical photographs of children [25,29]. In the focus group study by Devakumar et al [25], pediatricians emphasized that, although photographs are valuable resources, they might be potentially harmful. Thus, written informed consent was considered mandatory. In addition, they thought that the publication of photographs from this patient population required more stringent forms of informed consent to protect children. Similarly, a questionnaire survey by Hacard et al [29] included patients from a dermatology department and parents or legal guardians from the pediatric department and showed that acceptance of medical photographs was high among both groups. They considered that written informed consent was required for each purpose of use, with participants from the pediatric department being stricter in this aspect.

Practices and Use of Medical Photography for Research Publications

Studies addressing practices for taking medical photographs (20/30, 67%; Table 3) were conducted among different stakeholders: patients, medical staff, graduate and postgraduate students of medicine and dental medicine, residents, dermatologists, dentists, and plastic surgeons. The devices used for taking medical photographs included institutional cameras, personal cameras, and smartphones [9,29,38]. The device most often used for taking patient photographs was a personal camera (smartphone) [8,9,16,21,29,33], but 10% (3/30) of the studies showed that patients preferred to be photographed by their physicians using institutional cameras and in an institutional setting [10,15,37]. Patients considered that the use of personal devices, particularly smartphones, for taking medical photographs constituted a potential breach of patient-physician confidentiality [16].

Patients and physicians showed a high level of acceptance of medical photography for different purposes, such as medical documentation, research, communication, and education, but less for their publication in a public medium such as medical websites, professional emails, health magazines, and television health programs [24,26,29,37]. Furthermore, they preferred nonidentifiable over identifiable photographs for all types of use [10,12,15,35]. However, the studies also showed that conventional methods of deidentification of facial photographs concealing the eye area are not sufficient to achieve nonidentifiability [6,7]. The exception was the study by Engelstad et al [28], which demonstrated successful deidentification using a blended facial composite technique. This technique combined significant components of the original patient's photograph with cropped parts of the head and neck from other photographs using a computer software program to create nonidentifiable photographs that still presented patient details relevant to the clinical findings.

Several studies (8/20, 40%) reported that both patients and physicians considered important to ensure secure data storage, maintenance of privacy, and controlled access to the images [8,9,14,16,26,29,30,37]. A total of 7% (2/30) of the studies reported that dental students and dentists lacked knowledge and training regarding the techniques of taking and managing patient photographs, including for the purpose of publishing [21,23].

Medical Photography in Journal Articles

Studies addressing practices of publishing medical photographs in medical journals (5/30, 17%; Table 4) demonstrated a large diversity in publishing practices regarding consent for publication. Studies that analyzed high-impact general medical journals [32] or journals publishing dentistry and otorhinolaryngology research [2] showed that journal policies regarding the consent process and consent forms were insufficient and that existing ethical professional guidelines were not fully implemented in actual practices. A total of 40% (2/5) of the studies analyzed the deidentification of facial photographs published in medical journals and showed that the most common techniques, such as concealing the eye area, were not sufficient to protect the patient's identity [7,36]. The authors of these studies emphasized the importance of improving the policies regarding consent for publication of patient facial photographs. All facial photographs, with or without the applied deidentification technique, should require separate written consent for publication from the patient.

The only study that analyzed web access to patient-sensitive data published in open-access journals and platforms was the study by Marshall et al [3]. This study showed that patient medical photographs, including the face (8.1% of the photographs), were published in open-access formats and could be accessed easily via a Google Image browser, indicating a serious lack of protection of patient-sensitive data.

Discussion

Principal Findings

Our scoping review identified 30 studies that investigated different aspects of publishing identifiable patient photographs in research journals over a period of >25 years. It seems that, despite existing legal and professional guidelines regarding the use of patient photographs, obtaining informed consent properly is a challenge for many health professionals not only for scientific publications but also for other purposes. Relevant stakeholders did not always consider that only written informed consent was necessary for the publication of a patient photograph, and they also considered that it was possible to use patient photographs after obtaining oral consent or even without consent. Although relevant stakeholders were aware of the potential issues of using patients' medical photographs in terms of violating privacy and confidentiality, they did not have a satisfactory level of knowledge, skills, or tools to put existing guidance on medical photography in research into practice. Finally, there was little awareness of the current challenges, such as the protection of patients' clinical images and their permanence and availability in a digital environment.

Our study had some limitations. Most of the studies included in this scoping review (22/30, 73%) did not investigate medical photography publishing as the primary topic but as one of several purposes and aspects of medical photography. In addition, the studies did not define clear criteria for the deidentification of facial photographs. Although our search strategy did not have language restrictions, 97% (29/30) of the included studies were in English, and it is possible that there are studies in other languages that were not captured by the search of standard databases, registries, and gray literature sources. The methodological issues of the studies, such as questionnaire survey designs and insufficient reporting of methods and results, which might influence the validity of the studies, also limit the comparisons and generalizability of the findings.

We also did not assess the compliance of practices with legal standards as they varied in the countries in which the studies were performed. We presumed that the studies considered the contemporary legal regulations and investigated the compliance, practices, and opinions of the participants in relation to these regulations. The protection of patients' personal data is ensured by strict legal regulations such as the Health Insurance Portability and Accountability Act in the United States [39]. In Europe, the General Data Protection Regulation (GDPR) provides the strictest protection of patient personal data, including for research purposes [40,41]. According to the GDPR, consent is one of the legal bases for lawful processing of personal data and is most relevant for publications in scientific journals. Accordingly, a research participant consents to have the right to be informed, withdraw consent, have access to data, rectify the data or erase them, restrict data processing, and obtain and reuse their data (data portability) [40,41].

The problem of following procedures for obtaining valid patient consent in clinical practice is not limited to the purposes of publication but is a part of the general challenge in medical practice. A recent updated systematic review by Glaser et al [42] regarding interventions to improve patient comprehension in obtaining informed consent for medical procedures showed that, although progress was achieved, the consent process does not always meet 4 key elements of valid informed consent—decision capacity, documentation of consent, disclosure, and competency. The findings of our scoping review focused on the ethical publication of patient photographs in scientific journals and also showed that there are problems not only with the existing guidance, such as successful deidentification of published photographs, but even more with implementing existing guidance in practice, particularly in relation to proper and adequate informed consent. For example, despite the existence of ethical guidelines created by relevant professional or governmental organizations on the importance of obtaining patient consent in written form [5,43], studies that investigated the practices of obtaining informed consent for different uses of medical photographs (7/30, 23%) showed that even high-profile academic clinicians still considered verbal consent sufficient [8,9,23,26,33,35,37]. In another study, dermatologists reported that they did not always document patient consent for a specific purpose [9], indicating the need for better education at all career stages as well as the creation of more practical guidance for the implementation of standards for consent procedures in everyday practice. An unsatisfactory level of knowledge was also present among other health professionals as well as among patients [9,11,24,30,35]. Inasmuch as most of the analyzed studies in our scoping review used cross-sectional designs (26/30, 87%), future studies should have an interventional or qualitative design to investigate possible solutions for increasing the level of knowledge of all relevant stakeholders. We also did not identify studies that investigated whether consent for a research study included consent for publication and whether that consent provided sufficient information to the participants regarding how their photographs would be published and under which publishing license.

As open-access publishing has become a common format for medical research, patients' informed consent for identifiable photographs in open-access journals deserves special attention. Considering that patients can be a very heterogeneous group regarding educational level and socioeconomic background, health professionals should be able to explain to them that scientific articles published in open-access journals could be as accessible as any other information on the internet. It seems that patients do not always perceive that scientific journals are available on the web in the same way as any other content on the internet. For example, patients in an emergency department were more likely to consent to the use of their medical photographs in medical publications than on websites [22]. Thus, it is important that patients are fully informed of the implications of publishing photographs in a web-based medium before signing the consent form. They should be warned that open-access formats allow access to their photographs without any safeguards, leaving no possibility to withdraw or control their future use.

Generally, both patients and health professionals had high acceptance of medical photography and found it useful for many purposes, but patients preferred the use of nonidentifiable photographs [10,12,15]. Several studies that analyzed methods of deidentification of facial photographs (4/30, 13%) showed that conventional techniques were insufficient, and such photographs cannot be considered as nonidentifiable [6,7,35,36]. Such photographs should be considered as potentially reidentifiable, which leads to the conclusion that, in many situations, it is not possible to determine whether the photograph is identifiable. Furthermore, different computer programs have been developed to identify a person from a photograph (eg, DeepFace, Visual Search, Social Mapper, and Amazon Rekognition) with high levels of accuracy [44]. Patients may not be aware of this issue, but physicians should anticipate such situations and protect patients by providing proper informed consent. Although health care professionals commonly use medical photography in medical documentation, patient privacy becomes jeopardized when such photographs are used for other purposes such as communication with colleagues, lectures, presentations, or publications [43]. Patients were more likely to allow the use of their photographs for medical documentation, treatment follow-up, and education than for publications, websites, social media, and televised programs [26,29,37]. These findings suggest that patients recognize the increased risk of violating their privacy in a public environment regardless of their general affirmative attitudes toward medical photographs.

The studies included in this review addressed not only fully identifiable photographs but also potentially reidentifiable photographs as well as those that were considered to be nonidentifiable. The distinction between these types of photographs is very difficult [6-8], and it has been shown that patients and their families or social environment can recognize them even if the photograph that was published was considered to be fully nonidentifiable [45]. It would be safe to consider that all photographs of a patient's body are potentially identifiable or reidentifiable and that consent for the publication of such photographs should be sought.

The full maintenance of medical photography integrity requires practical protocols that should be in accordance with current guidelines and best ethical practices [5,46]. However, there are still different practices for taking and storing photographs, so it seems that the process of taking patient photographs has not yet been standardized and might be one of the reasons why the consent processes for their different uses are not often performed and reported in line with best ethical practices. Studies that analyzed the clinical practices of taking medical photographs (20/30, 67%) showed that patients were more likely to be photographed by their physicians than by other health personnel. In addition, patients were more consenting of being photographed with institutional cameras than with personal devices. These findings indicate that taking an identifying photograph is a sensitive procedure in which patients expect a high level of confidentiality and privacy. Following established guidelines such as those from the ICMJE would be a good beginning toward the responsible publishing of medical photography.

The analysis of medical journals also demonstrated the problem of an insufficient consent process for the publication of patient photographs. As journals have been shown not to be fully compliant with ICMJE consent recommendations, it was recommended that standard consent forms for the publication of identifiable images in medical journals should be developed [32]. Studies that investigated the ethical publication of identifiable photographs of patients (5/30, 17%) came to a very similar conclusion: there is a lack of consensus from journal editors and publishers, and uniform publishing policies are needed. The aforementioned recommendation seems reasonable and actionable as editorial organizations have created similar standards for other declarations in published studies, such as competing interest declarations from the ICMJE [47]. As previously mentioned, the digital environment represents a new challenge for publishing practices, especially with the growing trend of open-access publications. Journal editors and publishers should make clear what their publishing practices involve with regard to the use and sharing of published patient photographs, develop appropriate procedures for adequate and responsible declaration of obtaining informed consent for photograph publication that are separate from declaring and describing informed consent for research, and incorporate the submission of relevant declarations in web-based manuscript submission systems. They also have to protect the identity of the patients and not receive or publish consent forms from patients but rather ensure that authors provide declarations that appropriate procedures were followed and that patients gave informed consent for publishing their (identifying) photographs. If journals advise authors to provide proof of consent for persons mentioned in the acknowledgments [5], then they have to ensure the integrity of publishing patient photographs.

Gaps in Knowledge

This scoping review provided information about the attitudes, opinions, and practices regarding medical photography among relevant stakeholders and showed that they recognized the issues of privacy protection when medical photographs are used, particularly in publications. The impact of recent legal regulations related to personal data protection, such as the GDPR [40,41], on the publication of potentially identifiable photographs of research participants needs also to be further investigated. As it was shown that all stakeholders lack knowledge regarding the ethical publication of patient medical photographs, interventional studies are needed to address effective education and training. In addition, there is no evidence in the literature of the knowledge of stakeholders regarding published medical photographs in freely available web-based formats, particularly those published under licenses for wide use. An emerging issue that has not yet been addressed is the publication of medical photographs in preprints. Preprints, as "complete and public drafts of scientific documents, not yet certified by peer review" [48], do not pass the same scrutiny as regular journal publications, but their number and importance have enormously increased [49,50]. The latest update of the ICMJE Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals [5] in 2021 emphasized the need for appropriate declarations regarding published articles in preprint archives, such as disclosure of funding sources and disclosure of interest, but ethical issues about consent for patients' photographs were not mentioned. As one of the main aims of preprints is to increase the discoverability of research, the openness of such publications may be a facilitator for the research community but a threat and concern for patients whose photographs may be published in a way that will hinder the protection of their privacy. A recent study of editorial policies in preprint archives did not report on patient privacy protection, and only 20% of the archives in health sciences stated that they followed the ICMJE recommendations [51]. Future studies should investigate the practices of publishing patient photographs in preprints.

Conclusions

This scoping review of opinions, standards, and practices in publishing identifiable patient photographs in almost a 30-year period leads to the conclusion that all stakeholders in this issue have not fully developed and implemented best-practice standards for publishing medical images, particularly identifiable photographs of individuals. They are also not ready for the challenges of new developments in how we communicate research. In a digital environment, the protection of patient privacy is especially difficult because of how research information is shared on the web and on social media. Furthermore, newly developed digital tools for the deidentification of photographs are not commonly used, although it is clear that a standard black tape across the eyes on a photo does not make the person nonidentifiable. Despite the existence of legal, governmental, and professional policies and guidelines, the consent process and obtaining informed consent for publication are often not properly conducted or adequately reported in scientific literature. Relevant professional and ethics organizations, as well as journals and publishers, should address the emerging challenges in privacy protection by developing and updating guidance, protocols, and tools to ensure best practices in publishing patient photographs in medical literature.

Acknowledgments

This study was funded by the Croatian Science Foundation (*Professionalism in Health—Decision making in practice and research, ProDeM*) under grant agreement IP-2019-04-4882. The funder had no role in the design of the study; the collection, analysis, and interpretation of the data; or the decision to approve the publication of the finished manuscript.

Data Availability

The data sets used and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

MR, MM, and AM generated the idea and designed the study. MR, AU, and AM created the search strategy, and AU executed the search. MR, DŠ, TPP, MV, and AM participated in the screening or extraction and analysis of the data. All authors contributed to the interpretation of the results. MR and AM drafted the manuscript, and all authors critically revised it for important intellectual content. All authors approved the submitted version. All authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 100 KB - jmir_v24i8e37594_app1.pdf]

Multimedia Appendix 2
Search strategies.
[PDF File (Adobe PDF File), 82 KB - jmir_v24i8e37594_app2.pdf]

References

1. About The Licenses. Creative Commons. 2021. URL: https://creativecommons.org/licenses/ [accessed 2021-12-06]
2. Roguljić M, Buljan I, Veček N, Dragun R, Marušić M, Wager E, et al. Deidentification of facial photographs: a survey of editorial policies and practices. J Med Ethics 2022 Jan;48(1):56-60. [doi: 10.1136/medethics-2019-105823] [Medline: 32253363]
3. Marshall Z, Brunger F, Welch V, Asghari S, Kaposy C. Open availability of patient medical photographs in Google Images search results: cross-sectional study of transgender research. J Med Internet Res 2018 Feb 26;20(2):e70 [FREE Full text] [doi: 10.2196/jmir.8787] [Medline: 29483069]
4. Chevrier R, Foufi V, Gaudet-Blavignac C, Robert A, Lovis C. Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. J Med Internet Res 2019 May 31;21(5):e13484 [FREE Full text] [doi: 10.2196/13484] [Medline: 31152528]
5. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. International Committee of Medical Journal Editors. 2021. URL: http://www.icmje.org/icmje-recommendations.pdf [accessed 2021-12-06]
6. Clover AJ, Fitzpatrick E, Healy C. Analysis of methods of providing anonymity in facial photographs; a randomised controlled study. Ir Med J 2010 Sep;103(8):243-245. [Medline: 21046866]

7. Roberts EA, Troiano C, Spiegel JH. Standardization of guidelines for patient photograph deidentification. Ann Plast Surg 2016 Jun;76(6):611-614. [doi: 10.1097/SAP.0000000000000817] [Medline: 27015333]
8. McG Taylor D, Foster E, Dunkin CS, Fitzgerald AM. A study of the personal use of digital photography within plastic surgery. J Plast Reconstr Aesthet Surg 2008;61(1):37-40. [doi: 10.1016/j.bjps.2007.02.033] [Medline: 17766207]
9. Abbott LM, Magnusson RS, Gibbs E, Smith SD. Smartphone use in dermatology for clinical photography and consultation: current practice and the law. Australas J Dermatol 2018 May;59(2):101-107. [doi: 10.1111/ajd.12583] [Medline: 28247404]
10. Adeyemo WL, Mofikoya BO, Akadiri OA, James O, Fashina AA. Acceptance and perception of Nigerian patients to medical photography. Dev World Bioeth 2013 Dec;13(3):105-110. [doi: 10.1111/j.1471-8847.2012.00328.x] [Medline: 22708667]
11. Caires BR, Lopes MC, Okuno MF, Vancini-Campanharo CR, Batista RE. Knowledge of healthcare professionals about rights of patient's images. Einstein (Sao Paulo) 2015;13(2):255-259 [FREE Full text] [doi: 10.1590/S1679-45082015AO3207] [Medline: 26267838]
12. Leger MC, Wu T, Haimovic A, Kaplan R, Sanchez M, Cohen D, et al. Patient perspectives on medical photography in dermatology. Dermatol Surg 2014 Sep;40(9):1028-1037. [doi: 10.1097/01.DSS.0000452632.22081.79] [Medline: 25099296]
13. Copyright in images and photographs. Intellectual Property Office. 2021 Jan 4. URL: https://www.gov.uk/government/publications/copyright-notice-digital-images-photographs-and-the-internet/copyright-notice-digital-images-photographs-and-the-internet#copyright-in-images-and-photographs [accessed 2021-12-06]
14. Jones B. Ethics, morals and patient photography. J Audiov Media Med 1994 Apr;17(2):71-76. [doi: 10.3109/17453059409063661] [Medline: 8034908]
15. Lau CK, Schumacher HH, Irwin MS. Patients' perception of medical photography. J Plast Reconstr Aesthet Surg 2010 Jun;63(6):e507-e511. [doi: 10.1016/j.bjps.2009.11.005] [Medline: 19959408]
16. Nair AG, Potdar NA, Dadia S, Aulakh S, Ali MJ, Shinde CA. Patient perceptions regarding the use of smart devices for medical photography: results of a patient-based survey. Int Ophthalmol 2019 Apr;39(4):783-789. [doi: 10.1007/s10792-018-0878-2] [Medline: 29511959]
17. Bennett KG, Bonawitz SC, Vercler CJ. Guidelines for the ethical publication of facial photographs and review of the literature. Cleft Palate Craniofac J 2019 Jan;56(1):7-14. [doi: 10.1177/1055665618774026] [Medline: 29715061]
18. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
19. Viđak M, Marusic A. Consent for Publishing Facial Photographs in Academic Journals: a Scoping Review of the Quantitative and Qualitative Evidence. Open Science Framework. 2021 Jan 19. URL: https://osf.io/8cndt/ [accessed 2022-08-05]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Abouzeid HL, Chaturvedi S, Ahmed Alzahrani F, Alqahtani NM, Salim AlQarni AA, Alaajam WH, et al. A cross-sectional survey to evaluate acquaintance about dental photography among dental students in daily clinical practice. J Public Health Res 2020 Jul 28;9(3):1866 [FREE Full text] [doi: 10.4081/jphr.2020.1866] [Medline: 33014914]
22. Cheung A, Al-Ausi M, Hathorn I, Hyam J, Jaye P. Patients' attitudes toward medical photography in the emergency department. Emerg Med J 2005 Aug;22(8):609 [FREE Full text] [doi: 10.1136/emj.2004.019893] [Medline: 16046785]
23. Costa ED, Martins LA, Cral WG, Peroni LV, Freitas DQ, Oliveira ML. Assessment of dentists' behaviour on the use of patients' images. Eur J Dent Educ 2020 Aug;24(3):513-517. [doi: 10.1111/eje.12530] [Medline: 32304263]
24. de Runz A, Simon E, Brix M, Sorin T, Brengard-Bresler T, Pineau V, et al. [Photography in plastic surgery: practices, uses and legislation]. Ann Chir Plast Esthet 2015 Feb;60(1):12-18. [doi: 10.1016/j.anplas.2014.07.021] [Medline: 25179862]
25. Devakumar D, Brotherton H, Halbert J, Clarke A, Prost A, Hall J. Taking ethical photos of children for medical and research purposes in low-resource settings: an exploratory qualitative study. BMC Med Ethics 2013 Jul 09;14:27 [FREE Full text] [doi: 10.1186/1472-6939-14-27] [Medline: 23835013]
26. Dumestre DO, Fraulin FO. Balancing the need for clinical photography with patient privacy issues: the search for a secure smartphone application to take and store clinical photographs. Plast Surg (Oakv) 2017 Nov;25(4):255-260 [FREE Full text] [doi: 10.1177/2292550317731761] [Medline: 29619348]
27. Dumestre DO, Fraulin F. Avoiding breach of patient confidentiality: trial of a smartphone application that enables secure clinical photography and communication. Plast Surg (Oakv) 2020 Feb;28(1):12-18 [FREE Full text] [doi: 10.1177/2292550319880910] [Medline: 32110641]
28. Engelstad ME, McClellan M, Jacko JA, Melton GB. Deidentification of facial images using composites. J Oral Maxillofac Surg 2011 Dec;69(12):3026-3031. [doi: 10.1016/j.joms.2011.01.011] [Medline: 21601340]
29. Hacard F, Maruani A, Delaplace M, Caille A, Machet L, Lorette G, et al. Patients' acceptance of medical photography in a French adult and paediatric dermatology department: a questionnaire survey. Br J Dermatol 2013 Aug;169(2):298-305. [doi: 10.1111/bjd.12345] [Medline: 23551168]
30. Indu M, Sunil S, Rathy R, Binu MP. Imaging and image management: a survey on current outlook and awareness in pathology practice. J Oral Maxillofac Pathol 2015;19(2):153-157 [FREE Full text] [doi: 10.4103/0973-029X.164525] [Medline: 26604489]

31. Kunde L, McMeniman E, Parker M. Clinical photography in dermatology: ethical and medico-legal considerations in the age of digital and smartphone technology. Australas J Dermatol 2013 Aug;54(3):192-197. [doi: 10.1111/ajd.12063] [Medline: 23713892]
32. Lessing JN, Mark NM, Wynia MK, Cumbler E. Clinical image consent requirements: variability among top ten medical journals. J Acad Ethics 2019 Jun 21;17(4):423-427. [doi: 10.1007/s10805-019-09331-7]
33. Milam EC, Leger MC. Use of medical photography among dermatologists: a nationwide online survey study. J Eur Acad Dermatol Venereol 2018 Oct;32(10):1804-1809. [doi: 10.1111/jdv.14839] [Medline: 29405432]
34. Rimoin L, Haberle S, DeLong Aspey L, Grant-Kels JM, Stoff B. Informed consent, use, and storage of digital photography among Mohs surgeons in the United States. Dermatol Surg 2016 Mar;42(3):305-309. [doi: 10.1097/DSS.0000000000000634] [Medline: 26863597]
35. Roguljić M, Peričić TP, Gelemanović A, Jukić A, Šimunović D, Buljan I, et al. What patients, students and doctors think about permission to publish patient photographs in academic journals: a cross-sectional survey in Croatia. Sci Eng Ethics 2020 Jun;26(3):1229-1247. [doi: 10.1007/s11948-019-00134-y] [Medline: 31541413]
36. Shintani M, Williams J. Need for greater consensus on protection of patient anonymity and rights in facial photographs: a survey of international and domestic oral surgery journals. Bull Tokyo Dent Coll 2012;53(4):207-212 [FREE Full text] [doi: 10.2209/tdcpublication.53.207] [Medline: 23318927]
37. Wang Y, Tan H, Yang X. Perception and acceptability of medical photography in Chinese dermatologic patients: a questionnaire survey. Dermatol Surg 2017 Mar;43(3):437-442. [doi: 10.1097/DSS.0000000000000984] [Medline: 28099200]
38. Windsor JS, Rodway GW, Middleton PM, McCarthy S. Digital photography. Postgrad Med J 2006 Oct;82(972):688-692 [FREE Full text] [doi: 10.1136/pgmj.2006.044990] [Medline: 17068281]
39. Summary of the HIPAA privacy rule. U.S. Department of Health & Human Services. 2020. URL: https://www.hhs.gov/sites/default/files/privacysummary.pdf [accessed 2020-09-15]
40. General Data Protection Regulation (GDPR). General Data Protection Regulation, European Union. 2018. URL: https://gdpr.eu/tag/gdpr [accessed 2022-05-05]
41. Chassang G. The impact of the EU general data protection regulation on scientific research. Ecancermedicalscience 2017 Jan 3;11:709 [FREE Full text] [doi: 10.3332/ecancer.2017.709] [Medline: 28144283]
42. Glaser J, Nouri S, Fernandez A, Sudore RL, Schillinger D, Klein-Fedyshin M, et al. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: an updated systematic review. Med Decis Making 2020 Feb;40(2):119-143 [FREE Full text] [doi: 10.1177/0272989X19896348] [Medline: 31948345]
43. Stieber JC, Nelson T, Huebner CE. Considerations for use of dental photography and electronic media in dental education and clinical practice. J Dent Educ 2015 Apr;79(4):432-438. [Medline: 25838015]
44. Tegze J. How to Identify an Unknown Person using a Photograph. Medium. 2019 Mar 7. URL: https://medium.com/@jantegze/how-to-find-out-someones-name-from-a-picture-5472fcc60fe2 [accessed 2022-05-20]
45. BMJ journal pulls case report after UK tabloids publish graphic photos. Retraction Watch. URL: https://retractionwatch.com/2018/02/22/bmj-journal-pulls-case-report-after-uk-tabloids-publish-graphic-photos/ [accessed 2022-05-20]
46. Journals´ best practices for ensuring consent for publishing medical case reports: Guidance from COPE. Committee on Publication Ethics (COPE). 2021. URL: https://publicationethics.org/resources/guidelines/journals%E2%80%99-best-practices-ensuring-consent-publishing-medical-case-reports [accessed 2021-12-06]
47. Drazen JM, de Leeuw PW, Laine C, Mulrow C, DeAngelis CD, Frizelle FA, et al. Toward more uniform conflict disclosures--the updated ICMJE conflict of interest reporting form. N Engl J Med 2010 Jul 08;363(2):188-189. [doi: 10.1056/NEJMe1006030] [Medline: 20627859]
48. NIH Preprint Pilot. National Institutes of Health (NIH). 2021. URL: https://www.ncbi.nlm.nih.gov/pmc/about/nihpreprints/ [accessed 2022-01-12]
49. Watching preprints evolve. Nat Rev Immunol 2021 Feb;21(2):65-66 [FREE Full text] [doi: 10.1038/s41577-020-00489-5] [Medline: 33432126]
50. Bonnechère B. Preprints in medicine: useful or harmful? Front Med (Lausanne) 2020 Sep 22;7:579100 [FREE Full text] [doi: 10.3389/fmed.2020.579100] [Medline: 33072789]
51. Malicki M, Jeroncic A, Ter Riet G, Bouter LM, Ioannidis JP, Goodman SN, et al. Preprint servers' policies, submission requirements, and transparency in reporting and research integrity recommendations. JAMA 2020 Nov 10;324(18):1901-1903 [FREE Full text] [doi: 10.1001/jama.2020.17195] [Medline: 33170231]

Abbreviations

GDPR: General Data Protection Regulation
ICMJE: International Committee of Medical Journal Editors
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by R Kukafka; submitted 26.02.22; peer-reviewed by M Prictor, Y Joly, M Mars; comments to author 08.04.22; revised version received 27.05.22; accepted 23.06.22; published 31.08.22.

Please cite as:
Roguljić M, Šimunović D, Poklepović Peričić T, Viđak M, Utrobičić A, Marušić M, Marušić A
Publishing Identifiable Patient Photographs in Scientific Journals: Scoping Review of Policies and Practices
J Med Internet Res 2022;24(8):e37594
URL: https://www.jmir.org/2022/8/e37594
doi:10.2196/37594
PMID:36044262

Original Paper

Digital Health Care Industry Ecosystem: Network Analysis

Yoonseo Park[1*], MSc; Sewon Park[2*], PhD; Munjae Lee[2], PhD

[1]Department of BioMedical Engineering, Ajou University, Suwon, Republic of Korea

[2]Department of Medical Humanities and Social Medicine, Ajou University School of Medicine, Suwon, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Munjae Lee, PhD
Department of Medical Humanities and Social Medicine
Ajou University School of Medicine
206 World Cup-Ro
Yeongtong-gu, Suwon-si, Gyeonggi-do
Suwon, 16499
Republic of Korea
Phone: 82 31 219 5289
Fax: 82 31 219 4567
Email: emunjae@ajou.ac.kr

Abstract

Background: As the need for digital health care based on mobile devices is increasing, with the rapid development of digital technologies, especially in the face of the COVID-19 pandemic, gaining a better understanding of the industrial structure is needed to activate the use of digital health care.

Objective: The aim of this study was to suggest measures to revitalize the digital health care industry by deriving the stakeholders and major issues with respect to the ecosystem of the industry.

Methods: A total of 1822 newspaper articles were collected using Big Kings, a big data system for news, for a limited period from 2016 to August 2021, when the mobile health care project was promoted in Korea centered on public health centers. The R and NetMiner programs were used for network analysis.

Results: The Korean government and the Ministry of Health and Welfare showed the highest centrality and appeared as major stakeholders, and their common major issues were "reviewing the introduction of telemedicine," "concerns about bankruptcy of local clinics," and "building an integrated platform for precision medicine." In addition, the major stakeholders of medical institutions and companies were Seoul National University Hospital, Kangbuk Samsung Hospital, Ajou University Hospital, Samsung, and Vuno Inc.

Conclusions: This analysis confirmed that the issues related to digital health care are largely composed of telemedicine, data, and health care business. For digital health care to develop as a national innovative growth engine and to be institutionalized, the development of a digital health care fee model that can improve the regulatory system and the cost-effectiveness of patient care, centering on the Ministry of Health and Welfare as a key stakeholder, is essential.

(J Med Internet Res 2022;24(8):e37622) doi:10.2196/37622

KEYWORDS

digital health care; industrial ecosystem; network analysis; topic modeling; South Korea

Introduction

In the 21st century, epidemics, including Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome, spread rampantly, causing enormous social and economic losses and casualties. Above all else, pandemics such as COVID-19 are directly threatening our health and life, and modern society is entering a situation in which forecasting is more difficult than ever before [1,2]. Due to the prolonged COVID-19 pandemic, the amount of physical activity has decreased by 2%-30%. Measures such as quarantine and social distancing have only highlighted the importance of mental well-being and health care. The disconnection of communication in daily life due to social distancing has a great impact on the happiness and health of individuals. Social and economic factors such as unemployment and loss of income, depression, anxiety, and lack of social communication due to COVID-19 and associated restrictions on physical contact seem to have a negative impact on health. Accordingly, digital health care is being used to compensate for the collapse of the medical system due to the increase in infection and to manage health during disconnection. Digital health care is accessible regardless of location, enabling communication between users. Therefore, it is judged that more digital solutions will be essential not only for health care but also to solve the continuing COVID-19 pandemic and future pandemics.

As digital smart technologies such as smartphones, Internet of Things (IoT), wearable devices, and cloud computing, which were previously outside of the existing medical system area, are rapidly being grafted into the medical field, the utilization of digital health care has been emerging, through which health care can be received anytime and anywhere via various advanced information and communication technologies (ICTs) [3,4]. Digital health care technology is attracting attention as an effective untact (ie, without face-to-face encounters) treatment method in the COVID-19 crisis, which is recognized as an alternative that can be applied to patient evaluation and management [5]. Amid such changes, the medical environment is also expected to undergo a major change; the perception of telemedicine is changing due to the expansion of untact services and treatment caused by self-quarantine [6,7].

Digital health care was first mentioned by Seth R Frank in 2000 [8], which is defined as a service formed by the convergence of the internet and health care, combining health care with the core technologies of the so-called “4th industrial revolution,” including ICT, IoT, cloud computing, big data, and artificial intelligence (AI) [9,10]. In addition, digital health care provides a personalized health management process based on information, including health information, biorhythms, and health behaviors, collected through personal devices and health-related apps [11]. Restricted medical access and other unresolved problems in the medical field have traditionally limited access to medical data; however, a movement has recently emerged to foster the digital health care industry due to the revision of the 3 Data Acts and others [12,13]. A great feature of digital health care is that it enables the treatment and prediction of diseases by utilizing patient big data. Toward this end, it is necessary to effectively utilize big data while maintaining the security of personal information.

In Korea, regulations on data use have prevented digital health care companies from becoming active. However, with revision of the 3 Data Act, big data analysis using pseudonymized data became possible, and this is expected to be a turning point in digital health care innovation using AI and big data. Accordingly, as the development speed of digital technology has become rapidly high and various types of health care services based on mobile devices have been increasing, research is being actively conducted to expand the application of digital health care.

To date, studies on digital health care have mainly focused on user acceptance, intention, and willingness for continuous use. Becker [14] investigated the acceptance intention of mobile health (mHealth) apps, targeting German adults aged 18-35 years, finding that the leakage and loss of personal information in the medical sector is the most sensitive matter to users. In addition, Kim et al [15] analyzed the determinants of the intention to use wearable device products in a middle-aged and elderly population, with the goal of suggesting alternatives to the increase in health care demand and costs due to aging of the population. Moreover, studies on smart health care systems to expand the use of digital health care through big data case studies and providing patient-oriented services are being actively conducted. However, digital health care products involve many stakeholders before they are delivered to users. In the existing health care system, hospitals, pharmaceutical and medical device companies, and patients represent the main stakeholders, whereas the primary stakeholders in the digital health care industry are insurance companies, health care professional services, telecommunication companies, manufacturers of wearable devices (eg, biosensors), and health care app solution providers; thus, a new ecosystem is being formed.

In this changing paradigm, to activate and promote the utilization of digital health care, it is necessary to first understand the industry structure; however, the research in this field remains insufficient. In particular, digital health care products are closely associated with regulations and related policies because they affect the human body. For digital health care products to be practically used, it is necessary to establish future policy directions by grasping the relationship between stakeholders and issues regarding the industry.

Therefore, the aim of this study was to analyze the network of the digital health care industry using newspaper articles that form the basis of social debates on specific issues. Through this approach, we intend to suggest measures to revitalize the industry by deriving the stakeholders and major issues with respect to the ecosystem of the digital health care industry. In other words, by analyzing the ecosystem surrounding the current digital health care industry and identifying key stakeholders and major issues, we can propose policy alternatives for the future development of the industry. The ultimate goal is to provide personalized medical services through digital health care, thereby contributing to reducing social costs through preventive treatment.

Methods

Korean Digital Health Care

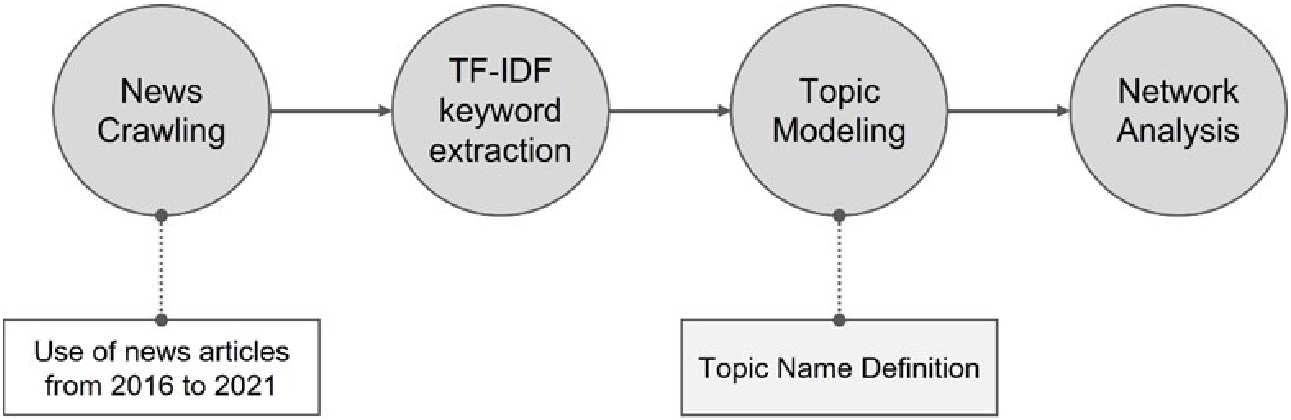
The major subject of digital health care in South Korea can be classified into a telemedicine pilot project and a mobile health care project. The former started with a remote image diagnosis pilot project as a collaboration between Seoul National University Hospital and Yeoncheon Public Health Center in 1988, which was promoted until the mid-1990s, but could not be activated due to factors such as limitations in ICT, an insufficient socioeconomic environment, and no supporting laws and systems [16]. Subsequently, after revision of the Medical Act to allow telemedicine between doctors and patients in 2003, digital health care has been used for health service accessibility, chronic disease management, and other health care services, centered on both local and public health centers [17]. In the telemedicine pilot that ran from 2014 to 2017, various types of pilot projects were applied for patients with chronic illnesses and visiting nursing systems to offer digital health care services for the disabled, elderly, and other vulnerable groups with access difficulty. A social consensus was not reached due to the risk of misdiagnosis and legal disputes in telemedicine. From this negative perspective, there were limits to the provision of telemedicine, which also limits the use of digital health care. Meanwhile, with the emergence of COVID-19, South Korea has temporarily allowed telemedicine, and has thus been able to provide patient-oriented medical services based on data obtained through digital health care. Through telemedicine, medical services were provided to patients who had difficulty visiting hospitals, improving health service accessibility. In the midst of such changes, even citizens who previously expressed negative views on telemedicine also empathize with the need for telemedicine, resulting in a trend of expansion in the utilization of digital health care in the country [18].

To overcome the limitations of the existing health care service and verify the effectiveness of mHealth, a mobile health care project has been promoted since 2016, centered on public health centers. For this purpose, an mHealth platform was developed, providing health care services by interlocking mobile apps and devices. To meet the demand for preventive health care, customized health care services are provided by utilizing digital health care based on ICT, big data, and other applications in health centers belonging to the public sector. For those with health risk factors for metabolic syndrome, doctors, nurses, nutritionists, exercise specialists, and other professionals are providing customized health counseling for 6 months. Digital health care is being utilized to reduce medical expenses and effectively help national health promotion by continuously managing and preventing chronic diseases and other conditions [19]. However, a health care service delivery system through digital health care has not yet been established owing to limitations in data utilization and related factors. Accordingly, to overcome such limitations and build an effective health management system utilizing digital health care, we performed an analysis on the stakeholders related to the digital health care industry.

Research Model

The purpose of this study was to derive measures to activate digital health care through network analysis. To this end, we used newspaper articles to identify trends in the related industries and analyze associated issues. Newspaper articles reflect public perceptions and industrial opinions on specific topics, enabling broader ecosystem analysis than possible with academic papers. In addition, as newspaper articles can reflect expert opinions, there is an added advantage of including the opinions of more professional stakeholders than possible when analyzing data from social networking sites (SNSs) such as Twitter and Google. The analysis period was set from 2016 to August 2021, when the mobile health care project was promoted, being centered on public health centers. For data collection, we used the news big data system BIG Kinds [20]. BIG Kinds is the largest search engine in Korea, providing articles from 11 central, 8 economic, 28 regional, 5 broadcasting, and 2 specialized magazines. To ensure data reliability, articles were collected using 11 domestic metropolitan newspapers, excluding economic magazines, regional comprehensive magazines, broadcasting companies, and specialized magazines. The search keyword was set to “digital health care,” and a total of 1822 articles were used for analysis. The study flow then followed the order term frequency-inverse document frequency (TF-IDF) keyword extraction, topic modeling, and network analysis (Figure 1).

Figure 1. Research model. TF-IDF: term frequency-inverse document frequency.



Analysis Method

The collected text data were analyzed using programs of R 4.1.1 (RStudio, Inc, Boston, MA, USA) and NetMiner 4.4.1 (Cyram Inc, Seoul, Korea). First, for data preprocessing, we used the “KoNLP” package, a Korean morpheme analysis function of R, and the “koRpus” package, which can extract language samples with a specific purpose for natural language research. Second, TF-IDF weights were extracted from the extracted text, and then major topics were derived using topic modeling. Third, network analysis was performed with NetMiner. Major stakeholders in the digital health care industry and related issues were analyzed through 2-mode network analysis, which can analyze the dual structure of data, such as the relationship between organizations and the relationship between an organization and its members. In other words, the 2-mode network was used because it enabled elucidating the relationships between stakeholders in the digital health care industry, between issues, and between stakeholders and issues.

The TF-IDF weight model is a statistical index to investigate the importance of keywords, which is the weight generated by multiplying the reciprocals of keyword and document frequencies. The larger the TF-IDF value, the more likely it is to determine the meaning or topic of the document to which the relevant word belongs; thus, it is utilized as a measure to extract major keywords from text data [21,22]. TF-IDF analysis is useful for analyzing issues frequently mentioned in SNS or news articles, through which we can discover the main content, information, network relationships, and other aspects regarding the issues of interest [23]. With topic modeling, all keywords included in text are organized by topic, where values within topics are automatically arranged in descending order. Since a large number of keywords are summarized into analysis units referred to as “topics,” the arranged topic can be considered a cluster representing the keywords that constitute the topic. Through this approach, it is possible to understand what topics the keywords are composed of, the importance between keywords, and similar characteristics [24,25]. Since topic modeling has the advantage of considering and identifying all of the multiple topics included in a single document, it has recently become more widely used in studies related to the fields of management, policy, and industry [26,27].

Network analysis is a method used to quantitatively analyze the relationship of individualized nodes, and is also a technique to identify the regular and stylized patterns that are consequently induced by interactions between actors [28]. Network analysis can find various hidden relationship types that are not normally recognized, and can structurally grasp the expression of specific relationship types [29,30]. Network analysis is divided into a 1-mode and 2-mode network according to the analysis target. One-mode network analysis is performed to analyze the relationship between objects when there is an n number of objects of the same nature, and the associative relation is analyzed after organizing the objects into an $n \times n$ matrix. When two objects with different properties exist, 2-mode network analysis is used to analyze the relation between them [31].

In addition, to interpret the network analysis, the degree of centrality is measured by using concepts such as the node, link, and connection degree [32]. Centrality is an index expressing the degree to which an actor is centrally located in the entire network; through centrality analysis, it is possible to identify key actors in the network and to determine how close each actor is to the center, along with similar metrics [33]. By showing the position each actor (nodes, keywords) occupies in the overall network and mathematically presenting their size, the actors can be separated into the core part and the periphery of the network. This approach can therefore enable searching for actors playing a central role in the network [34].

There are various types of centrality indices, including degree centrality, betweenness centrality, closeness centrality, and others. In this study, we used degree centrality, which is the most commonly used index in network analysis. Degree centrality refers to the degree of how many neighboring words are connected to a specific keyword; a higher number of connected words indicates higher centrality [35-37]. In degree centrality, the connection degree with other nodes is emphasized, which enables identifying how many relationships the node is involved in [38]. In other words, when using nodes with a high degree of connection, it will be easier to obtain information through their relationships with other nodes. If the nodes disappear in the network, interactions with connected neighboring nodes will be lost, resulting in loss of network function; thus, they are more likely to be key nodes in the network [39,40]. Therefore, in this study, we derived stakeholders and issues regarding the digital health care industry through 2-mode network analysis, also aiming to identify key stakeholders and corresponding major issues with respect to their relationship through evaluation of the degree centrality.

Results

TF-IDF Analysis

TF-IDF analysis to identify keywords that have been frequently utilized in the articles analyzed showed that the word with the highest importance among the top 25 keywords (Table 1) was “Medical Care,” indicating that discussions related to the application of digital health care in the medical field have been continuously ongoing. This highest-ranking word was followed by “Government,” “Hospitals,” “COVID-19,” “Care,” and “Insurance,” as the terms with the highest weight values. As a state of public health emergency has been declared in accordance with the spread of COVID-19 and telemedicine services have consequently expanded, it can be seen that the government is promoting precision medicine, smart hospital construction, and so forth. Conversely, keywords with relatively lower weight values were found to be “Regulation,” “Innovation,” “Samsung,” “Seoul National University,” and “Venture” (Table 1). These terms with lower weight demonstrate that discussion is insufficient on the regulatory policy that could be the basis for the activation of digital health care, and that cooperation between companies and universities for product development is not yet active.

Table 1. Top 25 ranking keywords based on term frequency-inverse document frequency (TF-IDF) analysis.

Rank	Keywords	TF-IDF
1	Medical care	60.75
2	Government	55.69
3	Hospital	48.76
4	COVID-19	46.52
5	Care	45.61
6	Insurance	44.98
7	Health	42.01
8	Information	40.43
9	Bio	39.92
10	Remote	37.15
11	Data	36.60
12	Company	36.29
13	USA	34.58
14	Seoul	33.84
15	Startup	33.23
16	Technology	31.70
17	Health	31.49
18	Digital	30.22
19	Communication	30.02
20	Economy	27.54
21	Regulation	27.41
22	Innovation	27.13
23	Samsung	26.48
24	Seoul National University	24.31
25	Venture	23.33

Topic Modeling Analysis

Topic modeling analysis identified a total of 7 topics; we directly assigned the topic name after identifying the correlation between keywords in the topic. Topics 2, 3, and 4 were assigned to "Government"; Topics 1 and 6 were assigned to "Medical Institution"; and Topics 5 and 7 were assigned the topic name "Company." Keywords belonging to the "Government" topic included Digital, Data, Support, Government, Regulation, COVID-19, and Platform, reflecting the recent movement to invigorate the digital health care industry through the government's vitalization of the usage of medical data, including the My Data project and the Bio-Health promotion policy. With COVID-19, the demand for untact medical care will further increase continuously in the future; thus, it seems that the necessity to vitalize related industries will be further highlighted. In addition, the government is preparing a foothold for nurturing the digital health care industry via related projects such as establishment of a big data platform, which has limitations in the private capacity.

The keywords belonging to the topic "Medical Institution" included Medical Treatment, Patient, Remote, Smart, Medical Care, and Policy. In the case of medical institutions, it is judged that since the number of outpatients has decreased due to COVID-19, the importance of telemedicine is expanding. In addition, this finding seems to reflect attempts to realize the effective management of patients with chronic diseases and the improvement of health service accessibility by building smart medical care using digital health care products. Conversely, the top keywords belonging to the topic "Company" were Service, Innovation, Device, Startup, Company, and Insurance. In the case of digital health care, those who have access to digital technologies such as AI, ICT, and IoT, even if not belonging to the existing digital health care industry, can readily enter the industry; accordingly, startups are actively being established. Furthermore, to develop digital health care products, cooperation with doctors and hospitals is required; thus, it seems necessary to establish a cooperative company-hospital-doctor system (Table 2).

Table 2. Topic modeling analysis.

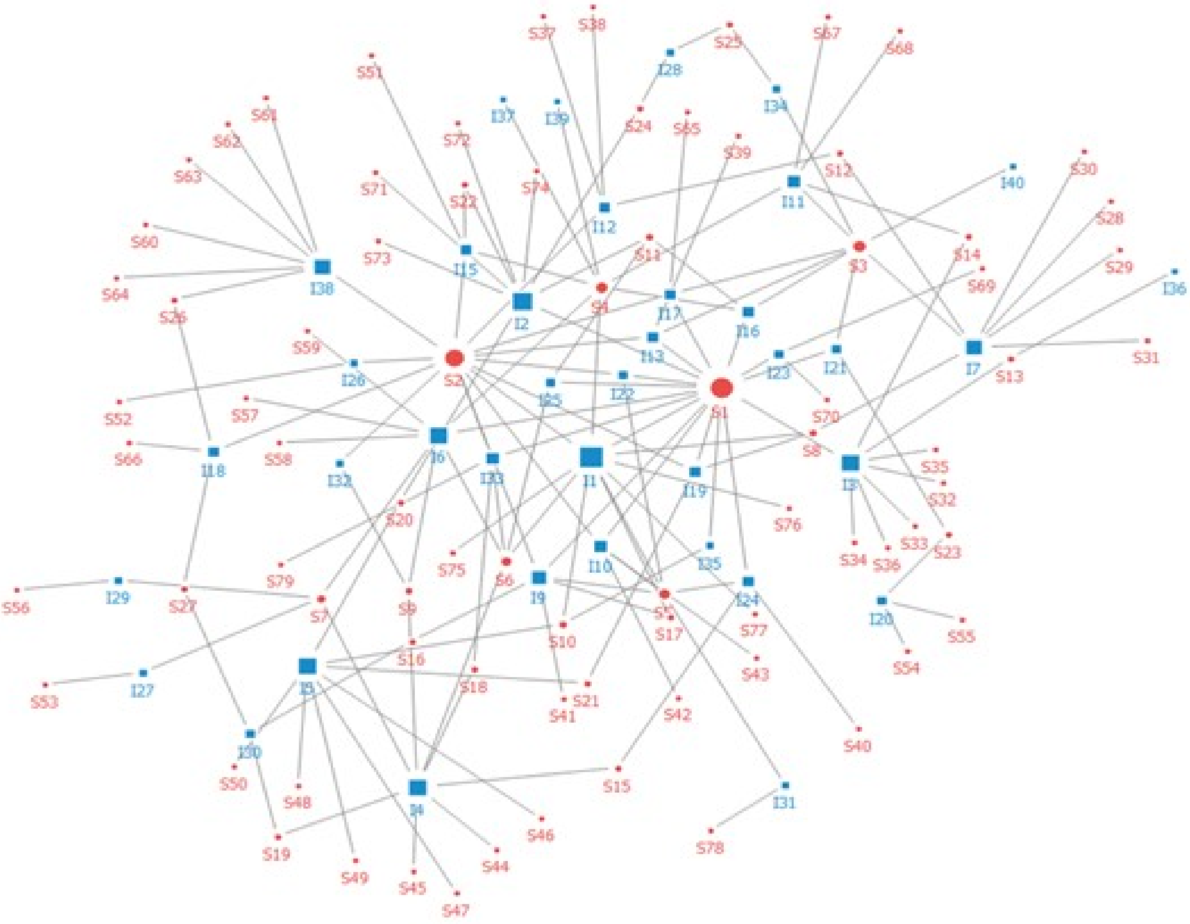
Themes and topics	Keywords
Government	
Topic 2	Digital, Health, Korea, Field, Support, Provision, Representative, Company, Communication
Topic 3	Health, Data, Market, Utilization, Investment, Cooperation, Nurturing, Research, Driving, Change
Topic 4	Government, Regulation, COVID-19, Bio, World, Platform, Plan, Individual, Data
Medical institution	
Topic 1	Government, Patient, Economy, Health, Technology, Medical Treatment, Construction, Cure, Region, Policy
Topic 6	Medical Care, Remote, Management, Smart, Seoul, Possibility, Special Zone, Country, Strategy, Apple
Company	
Topic 5	Service, Innovation, Device, Growth, Future, Startup, Center, Promotion, Diagnosis, Global
Topic 7	Industry, Company, Business, Hospital, Insurance, AI[a], Institution, Doctor, Expansion, Development

[a] AI: artificial intelligence.

Network Analysis

Network analysis was performed to analyze the ecosystem of the digital health care industry, and the relationship between nodes was identified according to the degree centrality (Figure 2). A total of 79 stakeholders and 40 key issues constituted the industrial ecosystem network. The major stakeholders included (1) government and regulatory agencies, represented by the government, Ministry of Health and Welfare, Ministry of Science and ICT, and Regulatory Reform Committee, and (2) medical institutions such as Seoul National University Hospital, Ajou University Hospital, and Gachon University Gil Medical Hospital. Stakeholders also included industrial and maintenance organizations such as the Korea Digital Health Industry Association, Asan Nanum Foundation, Samsung, LG Electronics, and KB Insurance.

Figure 2. Network analysis. Red indicates stakeholders (circles, S) and blue indicates the main issues (squares, I).



The stakeholder showing the greatest centrality was the Ministry of Health and Welfare. In other words, the Ministry of Health and Welfare raises core issues with respect to the network of the digital health care industry ecosystem, representing the actor connected with major issues and other actors in the industry. The government had the second highest centrality value, followed by the Ministry of Food and Drug Safety, Ministry of Science and ICT, and Korean Medical Association in descending order. Thus, the majority of government and regulatory agencies showed higher centrality, except for the three actors ranking in the top 10 with high centrality: the Korean Medical Association, Seoul National University Hospital, and SK Telecom. This seems to be because the digital health care industry ecosystem network became structured, led by the government, on the key issues and actors as the government-centered discussion on the activation of digital health care progressed (Table 3).

Next, we examined the major issues regarding stakeholders. The key issue in the digital health care industry ecosystem network was telemedicine introduction review, followed by passing of the 3 Data Acts, accuracy of AI technology in health care, deregulation of the health care sector, and smart health business (Table 4). Although South Korea has a strong personal information act, the development possibility of telemedicine has increased since revision of the 3 Data Acts was passed in the National Assembly. Particularly, telemedicine services have largely expanded in recent years since the state of public health emergency was declared according to the spread of COVID-19. Our findings further confirmed that deregulation in the health care sector, platform construction, and personnel training in related fields are emerging as major issues due to the expansion of private participation in digital health care. Moreover, other issues have been discussed, such as smart care policy, job creation, and AI real-time prediction of disease diagnosis and infection route (Table 4).

Table 3. Centrality of stakeholders.

Rank	Stakeholder	Number of nodes	Centrality
1	Ministry of Health and Welfare	17	0.425
2	Government	14	0.350
3	Ministry of Food and Drug Safety	8	0.200
4	Ministry of Science and ICT[a]	7	0.175
5	Korean Medical Association	6	0.150
6	Ministry of Strategy and Finance	5	0.125
7	Ministry of Trade, Industry and Energy	4	0.100
8	Seoul National University Hospital	3	0.075
9	Prime Minister	3	0.075
10	SK Telecom	3	0.075
11	Health Insurance Review & Assessment Service	3	0.075
12	Samsung	2	0.075
13	Watson Division	2	0.050
14	Vuno Inc	2	0.050
15	Democratic Party of Korea	2	0.050
16	Korea Health Industry Development Institute	2	0.050
17	National Assembly	2	0.050
18	People’s Solidarity for Participatory Democracy	2	0.050
19	Korea Economic Research Institute	2	0.050
20	Korea Insurance Research Institute	2	0.050
21	Kangbuk Samsung Hospital	2	0.050
22	Korea Pharmaceutical and Bio-Pharma Manufacturers Association	2	0.050
23	Osong Medical Innovation Foundation	2	0.050
24	National Health Insurance Corporation	2	0.050
25	Korea Internet & Security Agency	2	0.050
26	Gangwon-do	2	0.050
27	Ministry of SMEs[b] and Startups	2	0.050
28	LG Electronics	1	0.025
29	Medical Graduate School	1	0.025
30	KT	1	0.025
31	Green Cross Corporation	1	0.025
32	GE Healthcare	1	0.025
33	Ajou University Hospital		0.025
34	AI[c] Precision Medical Promotion Group	1	0.025
35	Gachon University Gil Medical Center	1	0.025
36	National Human Rights Commission of Korea	1	0.025
37	Apple	1	0.025
38	Regulatory Reform Committee	1	0.025
39	Seoul National University Bundang Hospital	1	0.025
40	Korean Pharmaceutical Association	1	0.025
41	Research Institute for Healthcare Policy under the KMA[d]	1	0.025

Rank	Stakeholder	Number of nodes	Centrality
42	U Health Industry Headquarters	1	0.025
43	Asan Medical Center	1	0.025
44	The People’s Party	1	0.025
45	Innovation Growth Headquarters	1	0.025
46	Samsung Fire Insurance	1	0.025
47	Kyobo Life Insurance	1	0.025
48	KB Insurance	1	0.025
49	Financial Supervisory Service	1	0.025
50	Financial Services Commission	1	0.025
51	Korea Digital Health Industry Association	1	0.025
52	Public Health Center	1	0.025
53	Korea Institute for Advancement of Technology	1	0.025
54	Siemens	1	0.025
55	KEPCO KDN	1	0.025
56	Korea International Trade Association	1	0.025
57	Electronics and Telecommunications Research Institute	1	0.025
58	Ministry of Public Administration and Security	1	0.025
59	Korea Venture Business Association	1	0.025
60	Minister Park Young-Sun	1	0.025
61	Deliberative Committee of Special Cases on Regulation	1	0.025
62	Korea Venture Investment Corporation	1	0.025
63	Chungcheongbuk-do	1	0.025
64	Busan	1	0.025
65	National Bio-bank of Korea	1	0.025
66	Ulsan	1	0.025
67	National Institute of Food and Drug Safety Evaluation	1	0.025
68	FDA[e]	1	0.025
69	Sejong-si	1	0.025
70	OECD[f]	1	0.025
71	National Cancer Center	1	0.025
72	Biotechnology Industry Organization	1	0.025
73	Korea Centers for Disease Control and Prevention	1	0.025
74	Korea Life Insurance Association	1	0.025
75	Regulatory Reform Committee	1	0.025
76	Seoul St. Mary's Hospital	1	0.025
77	Asan Nanum Foundation	1	0.025
78	Social Welfare Committee	1	0.025

Rank	Stakeholder	Number of nodes	Centrality
79	Korea Health Promotion Institute	1	0.025

[a] ICT: information and communications technology.

[b] SME: small and medium-sized enterprise.

[c] AI: artificial intelligence.

[d] KMA: Korea Meteorological Administration.

[e] FDA: Food and Drug Association.

[f] OECD: Organisation for Economic Co-operation and Development.

Table 4. Centrality of chief issues.

Rank	Chief issues	Number of nodes	Centrality
1	Review of telemedicine introduction	11	0.139
2	3 Data Acts passed	9	0.113
3	Accuracy of health care AI[a] technology	8	0.101
4	Deregulation in the health care sector	8	0.101
5	Health care insurance products	8	0.101
6	Fostering 5G front-back industries	8	0.101
7	Smart health business	7	0.088
8	Promotion of regulation-free special zones for fostering new industries	7	0.088
9	Concerns over village doctors going bankrupt	6	0.075
10	Establishment of precision medicine–integrated platforms	5	0.063
11	Personnel expansion for permission review	5	0.063
12	Health care platform market	4	0.050
13	Medical information big data	4	0.050
14	Conditional introduction of untact treatment	4	0.050
15	Regulatory sandbox	4	0.050
16	Resolving the personnel shortage of the Ministry of Food and Drug Safety	4	0.050
17	Construction of bio big data	4	0.050
18	Implementation of remote multidisciplinary diagnosis of primary medical institutions	4	0.050
19	Accelerating medical commercialization	4	0.050
20	Silver health care services	3	0.037
21	New drug development support	3	0.037
22	Medical Act amendment	3	0.037
23	Individual-led medical data activation	3	0.037
24	Activation of untact treatment	3	0.037
25	Health insurance fees for telemedicine	3	0.037
26	Judgment of whether the medical treatment is prompt	2	0.025
27	Blueprint of National Innovation Clusters	2	0.025
28	Probability prediction of diabetes and cardiovascular disease	2	0.025
29	Introduction of general digital norms	2	0.025
30	Silver robot expert	2	0.025
31	Need to expand public health services	2	0.025
32	Creation of funds dedicated to the regulation-free special zone	2	0.025
33	Expansion of untact health care services	2	0.025
34	Review for medical device cyber security permission	2	0.025
35	Smart care policy	2	0.025
36	Side effects of AI medical devices	1	0.012
37	Creation of jobs	1	0.012
38	Fostering small- and medium-sized ventures in new industries	1	0.012
39	AI real-time disease diagnosis and infection route prediction	1	0.012
40	Digital medicine	1	0.012

[a] AI: artificial intelligence.

Discussion

Principal Findings

As a result of network analysis on the digital health care industry, the government and the Ministry of Health and Welfare of Korea showed the highest centrality and were thus found to be major stakeholders; the major issues they had in common were review of telemedicine introduction, concerns over village doctors going bankrupt, and establishment of precision medicine–integrated platforms. Currently, the government is temporarily allowing telemedicine to respond to the spread of COVID-19, and intends to apply digital health care for remote patient monitoring and disease treatment to cope with insufficient medical demand. In addition, the issue of telemedicine is related to not only government agencies, including the National Assembly and Ministry of Science and ICT; but also medical institutions such as Seoul National University Hospital; and medical industry groups such as SK Telecom, Asan Nanum Foundation, and the Korean Medical Association.

In South Korea, telemedicine has been continuously discussed since the doctor/health care provider pilot project in 2002, but was not institutionalized due to the possibility of misdiagnosis, concerns over village doctors going bankrupt owing to the concentration of patients in large hospitals, and other related issues. As the digital health care industry has developed owing to the development of AI diagnosis technology, popularization of smart devices, and expansion of the utilization of medical data, the possibility of introducing telemedicine has also increased, and citizens' perception of telemedicine is changing owing to COVID-19. Therefore, the government, industrial sector, and medical care–related organizations should create institutional measures that could support digital health care through continuous discussions on the introduction of telemedicine that can contribute to improving public health. In particular, to prevent the concentration of large hospitals, which is a concern of medical institutions, it seems that patient convenience will improve, which could in turn improve the quality of medical care if primary medical institutions treat chronic diseases. Represcription of patients with mild conditions and patient monitoring for disease prevention could be achieved through the multidisciplinary treatment of primary medical institutions and large hospitals by utilizing digital health care products such as wearable devices. In this regard, an emergency medical system that can provide optimal treatment by collecting vital signs and images of patients in a severe emergency condition using digital health care in real time is being applied to the medical field.

Currently, there are only a few regional trauma and emergency medical centers in South Korea located in county areas, with the majority of such centers located in large cities such as Seoul and Gyeonggi. For emergency and trauma patients, the time required for treatment is important; thus, a balanced arrangement between regions is required. Therefore, digital health care can help to reduce the gap in medical infrastructure between rural and urban areas to better treat and manage patients with severe diseases. Expanding the functions of regional medical institutions by providing appropriate treatment for each emergency patient is expected to lay the foundation for expanding the utility of digital health care.

The major stakeholders of medical institutions and companies were found to be Seoul National University Hospital, Kangbuk Samsung Hospital, Ajou University Hospital, Samsung, and Vuno Inc, and their main issues were accuracy of health care AI technology and smart health business. Safety issues with respect to the utilization of digital health care have been constantly raised. Poor technical accuracy may lead to medical accidents; therefore, institutional strategies must be established in preparation of the possibility of misdiagnosis caused by product defects. It is necessary to improve the performance of digital health care products used for diagnosis and prescription, and to train professional personnel to develop and use the products appropriately. Furthermore, since data measurement must be accurate so as to increase the accuracy of diagnosis, it is judged thereby that periodic education for patients and medical staff in using the product will be necessary. Meanwhile, digital health care is an industry requiring medical information and continuous clinical data; with the increase in the use of digital health care, medical institutions are changing their position from a consumer to a supplier. Accordingly, if digital health care product development is accomplished through industry-university collaboration, it will be possible to develop safe digital health care products through continuously provided data, and in turn establish a smart health care business that can vitalize the industry.

Conclusion

This analysis confirmed that the major stakeholders in the digital health care industry of South Korea are largely composed of the government, medical institutions, and industrial companies, and that the issues related to digital health care largely consist of telemedicine, data, and health care business. We considered that all government, medical institutions, and industrial companies need to apply digital health care to the medical system through telemedicine and health care business establishment, and that cooperation is necessary among the government, medical institutions, corporations, research institutes, and related stakeholders. For practical cooperation, efficient use of data between institutions is required. Currently, medical data are stored in different ways between institutions, and there is a limit to the use of these data as there is no integrated management. Therefore, it is necessary to enable the use of integrated medical data for the commercialization of digital health care through standardized data linkage. This will not only revitalize the digital health care industry but will also lay the foundation for providing patient-tailored medical services, enabling the realization of precision medicine.

Meanwhile, for digital health care products to be effectively incorporated into the medical system, deregulation and the preparation for health insurance fees are necessary; however, there seems to be insufficient discussion on this aspect. The Ministry of Health and Welfare, a key stakeholder in the digital health care industry, is trying to vitalize the digital health care industry by preparing telemedicine health insurance rates; however, this analysis showed no connection of the Ministry with deregulation in the health care sector, the designation of regulation-free special zones, and other related issues. To vitalize the digital health care industry, it is necessary to expand the usability of patients through deregulation in related fields and application of health insurance fees; however, it seems that there has been no substantial institutional improvement on these aspects, as the discussion has mainly involved only government agencies, civic groups, and interest groups belonging to the related fields. Therefore, to promote the development of digital health care as a national innovative growth engine and its institutionalization, the development of a digital health care fee model that can improve the regulatory system and enhance the cost-effectiveness of patient treatment is essential, which will need to be centered on the Ministry of Health and Welfare as a key stakeholder. To revitalize the digital health care industry, as a national strategic project, the regulatory paradigm should be rationally established and centered on the market economy. This is expected to have a positive impact on the development of related industries by expanding into digital health care–related software, the medical service industry, and the insurance industry in the future, and ultimately enable a preventive-centered medical format through innovative medical service provision.

There are some limitations of this study that should be mentioned. Although newspaper articles have the advantage of providing a large amount of information, including opinions from various stakeholders on a specific topic, they also have a limitation in presenting data reflecting the popular experience on digital health care. Accordingly, future work exploring the issues regarding the digital health care industry by subdividing and analyzing data from SNS such as Twitter, blogs, and Facebook could better reflect the direct experience of digital health care and the views of the government, medical institutions, companies, and general consumers. In addition, we did not perform an analysis of the ecosystem according to the change in digital health care technology. In future research, it will be necessary to prepare a plan to provide digital health care for each patient type by analyzing the ecosystem according to technological change over time, which can help to identify issues and the structure of technology.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021R1I1A4A01057428), Bio-convergence Technology Education Program through the Korea Institute for Advancement Technology (KIAT) funded by the Ministry of Trade, Industry and Energy (No. P0017805), and new faculty research fund of Ajou University School of Medicine.

Conflicts of Interest

None declared.

References

1. Arendt F, Markiewitz A, Mestas M, Scherr S. COVID-19 pandemic, government responses, and public mental health: Investigating consequences through crisis hotline calls in two countries. Soc Sci Med 2020 Nov;265:113532 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113532] [Medline: 33223385]
2. 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]
3. Yoo B, Jeon H. A study on subjective perceptions of COVID-19 situation of senior living facilities worker. Korea Acad Care Manag 2021 Jun 30;39:31-54. [doi: 10.22589/kaocm.2021.39.31]
4. Lim Y. COVID-19 blues: a big data analysis. J Korean Soc Phys Ther 2021 May 31;33(2):829-852. [doi: 10.23844/kjcp.2021.05.33.2.829]
5. Santos-Moreno P, Rodríguez-Vargas GS, Casanova R, Rubio-Rubio J, Chávez-Chávez J, Rivera-Triana D, et al. Evaluation of a non-face-to-face multidisciplinary health care model in a population with rheumatoid arthritis vulnerable to COVID-19 in a health emergency situation. Healthcare 2021 Dec 17;9(12):1744 [FREE Full text] [doi: 10.3390/healthcare9121744] [Medline: 34946471]
6. Jin S. A study of factors affecting use intention of untact medical diagnosis and consultation services. J Korea Contents Assoc 2020;20(12):180-197. [doi: 10.5392/JKCA.2020.20.12.180]
7. Lee JS. A study on the activation of telemedicine in COVID-19 pandemic. J Knowl Inf Technol Syst 2020;15(5):815-829. [doi: 10.34163/jkits.2020.15.5.024]
8. Frank SR. Digital health care--the convergence of health care and the internet. J Ambul Care Manage 2000 Apr;23(2):8-17. [doi: 10.1097/00004479-200004000-00003] [Medline: 10848396]
9. Kim KB. A study of the digital healthcare industry in the Fourth Industrial Revolution. J Converg Inf Technol 2020;10(3):7-15. [doi: 10.22156/CS4SMB.2020.10.03.007]
10. Tortorella GL, Fogliatto FS, Saurin TA, Tonetto LM, McFarlane D. Contributions of Healthcare 4.0 digital applications to the resilience of healthcare organizations during the COVID-19 outbreak. Technovation 2022 Mar;111:102379. [doi: 10.1016/j.technovation.2021.102379]

11. Nikou S, Agahari W, Keijzer-Broers W, de Reuver M. Digital healthcare technology adoption by elderly people: a capability approach model. Telemat Informat 2020 Oct;53:101315. [doi: 10.1016/j.tele.2019.101315]
12. Lee WH. Evaluation and management of dysphagia based on digital health technologies. J Kor Dysphagia Soc 2021 Jul 30;11(2):105-110. [doi: 10.34160/jkds.2021.11.2.004]
13. Oh JH. A study on the perception of Data 3 Act through big data analysis. J Converg Secur 2021;21(2):19-28. [doi: 10.33778/kcsa.2021.21.2.019]
14. Becker D. Acceptance of mobile mental health treatment applications. Procedia Comput Sci 2016;98:220-227. [doi: 10.1016/j.procs.2016.09.036]
15. Kim YW, Han S, Kim KS. Determinants of intention to use digital healthcare service of middle and older users. Inf Soc Media 2018 Dec 31;19(3):1-23. [doi: 10.52558/ism.2018.12.19.3.1]
16. Lee J. A research on the necessity for telemedicine in the digital healthcare era. Dong A Law Rev 2020 Aug 31;88:245-272. [doi: 10.31839/dalr.2020.8.88.245]
17. Lee M, Park S, Lee K. Relationship between morbidity and health behavior in chronic diseases. J Clin Med 2020 Jan 02;9(1):121 [FREE Full text] [doi: 10.3390/jcm9010121] [Medline: 31906521]
18. Mobbs RJ, Ho D, Choy WJ, Betteridge C, Lin H. COVID-19 is shifting the adoption of wearable monitoring and telemedicine (WearTel) in the delivery of healthcare: opinion piece. Ann Transl Med 2020 Oct;8(20):1285-1285. [doi: 10.21037/atm-20-3678] [Medline: 33209865]
19. Yoon K, Park S, Choi S, Lee M. A proposal for public health information system-based health promotion services. Processes 2020 Mar 15;8(3):338. [doi: 10.3390/pr8030338]
20. BIG Kinds. URL: https://www.bigkinds.or.kr/ [accessed 2022-08-08]
21. Chen K, Zhang Z, Long J, Zhang H. Turning from TF-IDF to TF-IGM for term weighting in text classification. Exp Syst Appl 2016 Dec;66:245-260. [doi: 10.1016/j.eswa.2016.09.009]
22. Kim D, Seo D, Cho S, Kang P. Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. Inf Sci 2019 Mar;477:15-29. [doi: 10.1016/j.ins.2018.10.006]
23. Wang Z, Hahn K, Kim Y, Song S, Seo J. A news-topic recommender system based on keywords extraction. Multimed Tools Appl 2017 Dec 19;77(4):4339-4353. [doi: 10.1007/s11042-017-5513-0]
24. Ahadh A, Binish GV, Srinivasan R. Text mining of accident reports using semi-supervised keyword extraction and topic modeling. Process Safety Env Protect 2021 Nov;155:455-465. [doi: 10.1016/j.psep.2021.09.022]
25. Yun H, Park J, Yoon J. Introduction of topic modeling for extracting potential information from unstructured text data: issue analysis on news article of dementia-related physical activity. Korean J Sport Sci 2019 Sep;30(3):501-512. [doi: 10.24985/kjss.2019.30.3.501]
26. Fu Q, Zhuang Y, Gu J, Zhu Y, Guo X. Agreeing to disagree: choosing among eight topic-modeling methods. Big Data Res 2021 Feb;23:100173. [doi: 10.1016/j.bdr.2020.100173]
27. Kim SM, Kim Y. Research trend analysis on living lab using text mining. J Digit Converg 2020;18(8):37-48.
28. Leng Z, Sun H, Cheng J, Wang H, Yao Z. China's rare earth industry technological innovation structure and driving factors: a social network analysis based on patents. Resour Policy 2021 Oct;73:102233. [doi: 10.1016/j.resourpol.2021.102233]
29. Shukla M, Wu AF, Lavi I, Riddleston L, Hutchinson T, Lau JY. A network analysis of adolescent mental well-being during the coronavirus pandemic: evidence for cross-cultural differences in central features. Pers Individ Dif 2022 Feb;186:111316 [FREE Full text] [doi: 10.1016/j.paid.2021.111316] [Medline: 34629577]
30. Taylor S, Landry CA, Paluszek MM, Rachor GS, Asmundson GJ. Worry, avoidance, and coping during the COVID-19 pandemic: a comprehensive network analysis. J Anxiety Disord 2020 Dec;76:102327 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102327] [Medline: 33137601]
31. Lee M, Yoon K. Ecosystem of the medical device industry in South Korea: a network analysis approach. Health Policy Technol 2018 Dec;7(4):397-408. [doi: 10.1016/j.hlpt.2018.07.008]
32. Yuan CT, Nembhard IM, Kane GC. The influence of peer beliefs on nurses' use of new health information technology: a social network analysis. Soc Sci Med 2020 Jun;255:113002. [doi: 10.1016/j.socscimed.2020.113002] [Medline: 32353652]
33. Alotaibi N, Rhouma D. A review on community structures detection in time evolving social networks. J King Saud Univ Comput Inf Sci 2021 Aug:online ahead of print. [doi: 10.1016/j.jksuci.2021.08.016]
34. Lee M, Yoon K, Lee K. Social network analysis in the legislative process in the Korean medical device industry. Inquiry 2018 Aug 22;55:46958018791858 [FREE Full text] [doi: 10.1177/0046958018791858] [Medline: 30132349]
35. Behzadifar M, Gorji HA, Rezapour A, Rezvanian A, Bragazzi NL, Vatankhah S. Hepatitis C virus-related policy-making in Iran: a stakeholder and social network analysis. Health Res Policy Syst 2019 Apr 16;17(1):42 [FREE Full text] [doi: 10.1186/s12961-019-0442-1] [Medline: 30992014]
36. Li S, Garces E, Daim T. Technology forecasting by analogy-based on social network analysis: The case of autonomous vehicles. Technol Forecast Soc Change 2019 Nov;148:119731. [doi: 10.1016/j.techfore.2019.119731]
37. Nagarajan K, Muniyandi M, Palani B, Sellappan S. Social network analysis methods for exploring SARS-CoV-2 contact tracing data. BMC Med Res Methodol 2020 Sep 17;20(1):233 [FREE Full text] [doi: 10.1186/s12874-020-01119-3] [Medline: 32942988]

38. Das K, Samanta S, Pal M. Study on centrality measures in social networks: a survey. Soc Netw Anal Min 2018 Feb 28;8(1):13. [doi: 10.1007/s13278-018-0493-2]
39. Hwang JG, Ko YS, Lee C, Hwang JS. Analysis results in technical Ttends of 2018 Farnborough International Airshow via centrality analysis. J Korea Acad-Industr Coop Soc 2019;20(8):164-173. [doi: 10.5762/KAIS.2019.20.8.164]
40. Pomare C, Long JC, Churruca K, Ellis LA, Braithwaite J. Social network research in health care settings: design and data collection. Soc Netw 2022 May;69:14-21. [doi: 10.1016/j.socnet.2019.11.004]

Abbreviations

AI: artificial intelligence
ICT: information and communications technology
IoT: Internet of Things
mHealth: mobile health
SNS: social network site
TF-IDF: term frequency-inverse document frequency

Edited by R Kukafka; submitted 02.03.22; peer-reviewed by M Choi, HK Kim; comments to author 26.04.22; revised version received 10.05.22; accepted 25.07.22; published 17.08.22.

Please cite as:
Park Y, Park S, Lee M
Digital Health Care Industry Ecosystem: Network Analysis
J Med Internet Res 2022;24(8):e37622
URL: https://www.jmir.org/2022/8/e37622
doi:10.2196/37622
PMID:35976690

Viewpoint

Patient Design: The Importance of Including Patients in Designing Health Care

Bertalan Meskó[1,2], MD, PhD; Dave deBronkart[3,4], SB

[1]The Medical Futurist Institute, Budapest, Hungary
[2]Department of Behavioural Sciences, Semmelweis University, Budapest, Hungary
[3]e-Patient Dave, LLC, Boston, MA, United States
[4]Society for Participatory Medicine, Nutting Lake, MA, United States

Corresponding Author:
Bertalan Meskó, MD, PhD
The Medical Futurist Institute
Povl Bang-Jensen u 2/B1 4/1
Budapest, 1118
Hungary
Phone: 36 36703807260
Email: berci@medicalfuturist.com

Abstract

A paradigm shift is underway in the patient-clinician relationship, driven by irreversible changes in information access, yet the model under which clinicians are trained, care is conducted, and care delivery is designed has not changed significantly even though we call it "patient centered." Humanity endured centuries in which even doctors had little idea what the patient's problem really was. Science slowly solved that, and for a century, only doctors could know what was worth knowing. Today, the rise of the internet and digital health has led to the end of that era. We are already witnessing early signs of the era of participatory health: genuinely empowered people living their lives and managing their health according to their own priorities, in partnership and consultation with physicians as needed. This may feel like a threat to the physician's sacred role, but it is no more so than when physicians adopted informed consent and then shared decision-making. In the 2010s, many pharmaceutical, medical, and health care companies started to use patient centricity as a mantra. We argue that to drive this paradigm change fully into existence, we need to shift "patient centricity" from a relatively passive process, driven by industry needs, into a far more active, collaborative process driven by both parties' needs and preferences. To build this new world of practice and workflow, we simply must engage with patients as true partners. To achieve medicine's new potential, it must be optimized around the wants and priorities of the ultimate stakeholder—the party that has the most at stake in how it all plays out: the patient. Patient design is the approach that can make it happen.

(J Med Internet Res 2022;24(8):e39178) doi:10.2196/39178

KEYWORDS

patient; patient design; user design; patient centric; patient focus; digital health; future; empowerment; involvement; participatory; engagement; participation; patient centred; patient centered

The Short History of Patient Empowerment

Health care has been going through a paradigm shift in the 21st century, as per Thomas Kuhn's 1962 classic, *The Structure of Scientific Revolutions* [1]. Kuhn was an American philosopher who was influential in both academic and popular circles, introducing the term "paradigm shift." His "blockbuster" book claimed that sometimes a scientific field discovers it was wrong about something important.

Kuhn [1] wrote that "perhaps science does not develop by the accumulation of individual discoveries and inventions," but that "discovery commences with the awareness of anomaly, i.e., with the recognition that nature has somehow violated the paradigm-induced expectations that govern normal science."

We are at the point of detecting such anomalies in health care. For centuries, the dominant paradigm has been that patients do not and cannot contribute to their care, especially medical decisions concerning their case. However, as the ivory tower of medicine started breaking down in the early 21st century, empowered patients started bringing real value to their own cases, violating the paradigm-induced expectations that dictate the culture of medicine [1].

Kuhn [1] wrote that when too many anomalies accumulate, a field goes into crisis mode until a new paradigm is developed and accepted. We assert that the current paradigm cannot explain nor cope with the cluster of anomalies in which patients are genuinely creating value in health care, and to ignore this is to suboptimize health care in a way we can no longer afford.

Thus, this crisis stage has arrived, but as often happens, the causes of the anomalies are poorly understood, which leads to confusion. No new paradigm can arise, letting the field advance, while confusion reigns. Here, we present those causes—the factors that did not exist a generation ago, and now do:

- Consumerism: the cultural willingness of consumers to pursue their own priorities
- Information liquidity brought by the internet, which eradicated the belief that only people from the “priesthood” could know certain facts
- Advanced consumer health technologies putting unprecedented knowledge in the hands of consumers who had previously been uninformed
- Global supply chains making it possible for new products and technologies to reach patients worldwide
- The rise of social media enabling peer-to-peer communication among patients about needs and solutions

In truth, patient empowerment has been evolving for decades, but information liquidity and access to technology made it explode in this century—and become visible to the naked eye. At the time of his death in 2006, “Doc Tom” Ferguson, MD, was working on a white paper funded by the Robert Wood Johnson Foundation, “e-Patients: How They Can Help Heal Healthcare,” documenting what empowered and engaged patients had been doing as far back as the 1980s. His colleagues published the paper in 2007, and, in 2009, founded the Society for Participatory Medicine [2].

The cultural transformation we call digital health represents this paradigm shift and is a continuation of that vision [3].

The Practical Reality of Patient Empowerment

Today, except for the commercial obstacle inserted by paywalls, patients can have access to the same online health care resources, studies, and data as medical professionals [4]. Empowered patients want to get engaged in their health or disease management. There are many examples of how patients take their lives into their own hands. From joining patient communities online to using a range of digital health sensors, they bring new value to the table. In doing so, they violate the paradigm’s cultural expectation that only doctors know anything useful [5].

Empowered patients also put pressure on regulators. The #WeAreNotWaiting movement is a community of patients with diabetes taking disease management into their own hands by organizing themselves and developing applications, platforms, and other solutions to help each other beat their disease. They even created the “DIY pancreas” software, which automatically provides patients with the right doses of insulin based on their blood glucose level [6]. The software was created entirely by the patient community with no contribution from medical professionals.

It should be noted that this OpenAPS (Open Artificial Pancreas System) software is the second most-forked item on all of GitHub (Microsoft Corp) because almost all patients tweak the code to suit their own biological response. In other words, the app is designed from the ground up to be fully configurable to suit individual needs. This is the most advanced example of patient design we have seen.

Patient scholars have published in prestigious medical journals [7-9]. The #PatientsIncluded movement has led to involving patients in medical events either as speakers or cohosts. Governments such as that of New Zealand have started developing digital health policies featuring empowered patients.

These examples further underscore that a more patient-inclusive design approach is already emerging and will inevitably be the norm. The only thing holding it back is cultural resistance, which is why we say digital health is a cultural transformation.

The Rise of Patient Centricity and Patient Design

In the 2010s, myriad pharmaceutical, medical, and health care companies started to use patient centricity as a mantra. Each claimed that their company is patient centric and thus ahead of the others. Pharmaceutical company executives started making “putting patients first” part of their slogans and internal documents. A 2020 survey revealed that 85% of companies were raising their investment in patient-centric capabilities over the next 18 months [10].

Patients want more reliable and relatable health information from the companies that make their medication, so this was an obvious step forward for the industry. A 2019 survey indicated that 76% of patients expect pharmaceutical organizations to provide them with tools and support services [11].

At the same time, policy makers started adopting this theme too. The US Food and Drug Administration (FDA) launched the Patient Engagement Advisory Committee in 2017. The committee provides advice to the FDA commissioner or designee on complex issues relating to medical devices, the regulation of devices, and their use by patients.

The Need for a New Level of Patient Centricity: Patient Design

To drive this paradigm change fully into existence, we call for changing patient centricity from a relatively passive process, driven by industry needs, into a far more active, collaborative process driven by both parties’ needs and preferences. In short, it is no longer viable for patient centricity to mean, “We were *thinking about you* while *we* made *our* decisions.”

From the patient’s perspective, patient centricity has been a passive process since the inclusion of their opinion in the final design depends solely on those who invite the patient's opinion. That approach may sound patient centric, but in this scenario, patients' voices literally have no power since all decisions are still made by the project organizers. Sociologist Sherry Arnstein bluntly called this tokenism [12].

In contrast to this, patient design means patients are involved in the highest level of decision-making in the organization, essentially having patients advise the chief executive officer of a company or the head of a health care organization.

Patient design is a so-called "co-design" approach. Co-design is defined as "a creative practice that can be used to improve customer experience and enhance value" [13]. The approach involves a wide range of people who are the experts of their experience and therefore make creative contributions that come from the perspective of the person who has the need. This is only possible by admitting to ourselves that the cared-for person just might know what they want and be well-informed! On what basis would someone assert otherwise?

The short-term benefits of such a co-design approach could include:

- More original ideas arising from more diverse perspectives and priorities
- Better achievement of consumer value (by incorporating the voice of the person for whom the project exists)
- Improved knowledge of patient needs
- Immediate validation of ideas
- More efficient decision-making
- Reduced development time
- Generally better cooperation between patients and companies or organizations.

The long-term benefits of such a co-design approach could include higher degrees of patient satisfaction, increased levels of support and enthusiasm for innovation, and a better relationship between patients and companies (Table 1).

Table 1. Comparison of patients' role and power in token patient centricity versus patients as empowered design partners.

	Token "patient centricity"	Patients as design partners
Involvement of patients	Passive; when asked by the power holders	Active at all times
Patients' decision power	None; their opinions are sought but need not be heeded. The system is free to continue not responding	Shared
Type of input provided by patients	Share their opinions when invited	Actively influence design decisions, including what gets worked on
Mode of involvement	Through surveys, questionnaires, and focus groups, all organized by the power holders	By sitting on project committees and advisory boards that set agendas
What level of decision-making patients influence	Any level within the organization	The highest level of decision-making

Real-life Practical Examples of How Patient Designs Work

This social movement has already progressed to where examples exist to illustrate the shift in thinking—the paradigm change.

Physical Products

For physical products, the Patient Innovation website [14] shares innovations developed by patients. Some focus on a disease, some are just for a symptom, while others enable a particular activity. What they have in common is that they all feature patient-centered thinking: they are expressions of what patients *want* to improve.

Research

In research, patient voices are calling for researchers to change priorities to match patients' urgent needs. The father of a son with suicide ideation told Dr Thomas Insel after a speech [15], "Our house is on fire, and you're telling us what you learned about the chemistry of the paint." The scientific literature may contain volumes about "the paint," but Insel realized "this gap between our scientific progress and our public health failure." He left academia to pursue product development to solve real-world problems.

The urgency articulated so powerfully by Insel's audience member precisely echoes the urgency of AIDS activists in the last century who demanded that science respond more to patients' immediate needs, not just long-term science.

When the husband of Bettina Ryll, MD, PhD, was dying of melanoma, she switched hats and observed the clinical trials process as a family stakeholder and was incensed to discover that researchers chose their work priorities without consulting the people who were dying. Today, she advocates in *Nature* for researchers to do just that [16].

Clinical Design

In clinical design, the Oral and Maxillofacial Surgery Department of Radboud University Medical Centre Nijmegen in the Netherlands redesigned the whole department's rooms based on advice from patients. This involved changes that enabled a more balanced patient-physician relationship and a comfortable atmosphere. They prioritized round tables over square ones for more friendly conversations and suggested brightly lit rooms with warm colors.

These examples merely illustrate how differently things can play out when, in fact, the patient truly has agency in influencing the nature of care.

While it may seem that we portray patient design in an almost utopian manner, we must note that such transitions—like all cultural changes—are time-consuming and complicated. An analogy is Title IX, a law introduced in the United States in 1972 that required schools to provide sports for girls, not just boys. It took a generation before the US women's soccer team won the World Cup in 1991. A major consequence of an incorrect paradigm is that if we do not let women onto the pitch (or patients into the executive world), we have no chance of discovering their actual suppressed potential.

Conclusion

We are at the end of the only period in history where physicians knew important scientific facts and medical insights that patients could not. For health care to achieve its potential in this new era, our methods, along with our paradigm, must change.

Before 1900, medical practice mostly lacked any scientific basis. Doctors were not even exposed to patients in medical school (patients were merely called "clinical material" in the Flexner Report!). Yet by the end of the 20th century, the internet let knowledge flow.

Now in the 21st century, a plethora of personal health devices gives patients access to *more* information than their physicians have. The possibility of true participatory medicine is on the horizon—patients with increasing autonomy living their lives according to their own priorities, in partnership and consultation with physicians *as needed*.

To build this new world of practice and workflow, we simply must engage with patients as true partners. To achieve medicine's new potential, it must be optimized around the wants and priorities of the ultimate stakeholder—the party that has the most at stake in how it all plays out: the patient. Patient design is the approach that can make it happen.

Conflicts of Interest

None declared.

References

1. Kuhn T. The Structure of Scientific Revolutions. Chicago, IL: University of Chicago Press; 1962.
2. Ferguson T, The e-Patient Working Group. E-patients: how they can help us heal healthcare. 2007. URL: https://participatorymedicine.org/e-Patient_White_Paper_with_Afterword.pdf [accessed 2022-08-22]
3. 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]
4. deBronkart D. Open access as a revolution: knowledge alters power. J Med Internet Res 2019 Dec 11;21(12):e16368 [FREE Full text] [doi: 10.2196/16368] [Medline: 31825321]
5. deBronkart D, Eysenbach G. Gimme my damn data (and let patients help!): the #GimmeMyDamnData manifesto. J Med Internet Res 2019 Nov 22;21(11):e17045 [FREE Full text] [doi: 10.2196/17045] [Medline: 31755873]
6. Lewis D, Leibrand S, #OpenAPS Community. Real-world use of open source artificial pancreas systems. J Diabetes Sci Technol 2016 Nov 20;10(6):1411-1411 [FREE Full text] [doi: 10.1177/1932296816665635] [Medline: 27510442]
7. deBronkart D. How the e-patient community helped save my life: an essay by Dave deBronkart. BMJ 2013 Apr 02;346(apr02 1):f1990-f1990. [doi: 10.1136/bmj.f1990] [Medline: 23550048]
8. Anampa-Guzmán A, Freeman-Daily J, Fisch M, Lou E, Pennell NA, Painter CA, Collaboration for Outcomes using Social Media in Oncology. The rise of the expert patient in cancer: from backseat passenger to co-navigator. JCO Oncol Pract 2022 Aug;18(8):578-583. [doi: 10.1200/OP.21.00763] [Medline: 35344398]
9. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
10. Brückner A. Investing in patient-centric capabilities requires a clear strategy. PharmExec. 2016 May 31. URL: https://www.pharmexec.com/view/investing-patient-centric-capabilities-requires-clear-strategy [accessed 2022-08-22]
11. Patients expect pharmaceutical companies to provide services that help them manage their health, Accenture survey finds. Accenture. 2014 Mar 20. URL: https://tinyurl.com/y93xsvun [accessed 2022-08-22]
12. Arnstein SR. A ladder of citizen participation. J Am Plann Assoc 1969 Jul;35(4):216-224. [doi: 10.1080/01944366908977225]
13. King M. Co-design and patient-centred design. Colleaga. URL: https://www.colleaga.org/article/co-design-and-patient-centred-design [accessed 2022-08-22]
14. Patient Innovation. URL: https://patient-innovation.com/ [accessed 2022-08-22]
15. Chatterjee R, Dalrymple L, Lu T. Dr. Thomas Insel on why the U.S mental health system has failed and what can be done. NPR. 2022 Mar 07. URL: https://www.npr.org/2022/03/04/1084567343/dr-thomas-insel-on-why-the-u-s-mental-health-system-has-failed-and-what-can-be-d [accessed 2022-08-22]
16. Ryll B. From good to great: what patients can do for your medical research. Nat Med 2020 Oct 07;26(10):1508-1508. [doi: 10.1038/s41591-020-1097-8] [Medline: 33029016]

Abbreviations

OpenAPS: Open Artificial Pancreas System

FDA: US Food and Drug Administration

Edited by G Eysenbach, T Leung; submitted 02.05.22; peer-reviewed by J Kaswija, E Ray Chaudhuri, L Evans, R Marshall; comments to author 10.06.22; revised version received 17.08.22; accepted 19.08.22; published 31.08.22.

Please cite as:
Meskó B, deBronkart D
Patient Design: The Importance of Including Patients in Designing Health Care
J Med Internet Res 2022;24(8):e39178
URL: https://www.jmir.org/2022/8/e39178
doi:10.2196/39178
PMID:36044250

Original Paper

An Identity-Affirming Web Application to Help Sexual and Gender Minority Youth Cope With Minority Stress: Pilot Randomized Controlled Trial

Jose Bauermeister[1], MPH, PhD; Seul Ki Choi[1], PhD; Emma Bruehlman-Senecal[2], PhD; Jesse Golinkoff[1], MPH; Arianna Taboada[2], MSW, MSPH; Joshua Lavra[2], BS; Lionel Ramazzini[2], BFA; Fred Dillon[2], BA; Jana Haritatos[2], PhD

[1]Department of Family and Community Health, University of Pennsylvania, Philadelphia, PA, United States
[2]Hopelab, San Francisco, CA, United States

Corresponding Author:
Jose Bauermeister, MPH, PhD
Department of Family and Community Health
University of Pennsylvania
418 Curie Blvd
Suite 222L
Philadelphia, PA, 19104
United States
Phone: 1 215 898 9993
Email: bjose@upenn.edu

Abstract

Background: Efficacious mental health interventions for sexual and gender minority youth have had limited reach, given their delivery as time-intensive, in-person sessions. Internet-based interventions may facilitate reach to sexual and gender minority youth; however, there is little research examining their efficacy.

Objective: This study aims to describe the results of a pilot randomized controlled trial of *imi*, a web application designed to improve mental health by supporting lesbian, gay, bisexual, transgender, queer, and other sexual and gender minority identity affirmation, coping self-efficacy, and coping skill practice.

Methods: Sexual and gender minority youth (N=270) aged 13 to 19 (mean 16.5, SD 1.5) years and living in the United States were recruited through Instagram advertisements. Approximately 78% (210/270) of the sample identified as racial or ethnic minorities. Participants were randomized in a 1:1 fashion to the full *imi* intervention web application (treatment; 135/270, 50%) or a resource page–only version of the *imi* site (control; 135/270, 50%). The *imi* application covered four topical areas: gender identity; lesbian, gay, bisexual, transgender, queer, and other sexual and gender minority identity; stress and coping; and internalized homophobia and transphobia. Participants explored these areas by engaging with informational resources, exercises, and peer stories at a self-guided pace. Both arms were assessed via web-based surveys at baseline and 4-week follow-up for intervention satisfaction, stress appraisals (ie, challenge, threat, and resource), coping skills (ie, instrumental support, positive reframing, and planning), and mental health symptoms among other outcomes. Main *intent-to-treat* analyses compared the arms at week 4, controlling for baseline values on each outcome.

Results: Survey retention was 90.4% (244/270) at week 4. Participants in the treatment arm reported greater satisfaction with the intervention than participants in the control arm (t_{241}=–2.98; *P*=.003). The treatment arm showed significantly greater improvement in challenge appraisals (ie, belief in one's coping abilities) than the control (Cohen *d*=0.26; *P*=.008). There were no differences between the arms for threat (*d*=0.10; *P*=.37) or resource (*d*=0.15; *P*=.14) appraisals. The treatment arm showed greater increases in coping skills than the control arm (instrumental support: *d*=0.24, *P*=.005; positive reframing: *d*=0.27, *P*=.02; planning: *d*=0.26, *P*=.02). Mental health symptoms improved across both the treatment and control arms; however, there were no differences between arms. Within the treatment arm, higher engagement with *imi* (≥5 sessions, >10 minutes, or >10 pages) predicted greater improvement in stress appraisals (all *P* values <.05).

Conclusions: The results provide initial evidence that asynchronous psychosocial interventions delivered via a web application to sexual and gender minority youth can support their ability to cope with minority stress. Further research is needed to examine the long-term effects of the *imi* application.

Trial Registration: ClinicalTrials.gov NCT05061966; https://clinicaltrials.gov/ct2/show/NCT05061966

(J Med Internet Res 2022;24(8):e39094) doi:10.2196/39094

KEYWORDS

lesbian, gay, bisexual, transgender, queer, and other sexual and gender minority; LGBTQ+; youth; adolescence; discrimination; minority stress; mental health; resilience; sexual and gender minority; SGM; intersectionality

Introduction

Background

Compared with their cisgender, heterosexual peers, sexual and gender minority (SGM) youth are at increased risk of experiencing a wide variety of negative mental health outcomes [1]. In a recent national surveillance survey conducted by the Centers for Disease Control and Prevention, 55% of gay, lesbian, or bisexual youth reported poor mental health during the prior 30 days. SGM youth were also twice as likely as their non–SGM youth counterparts to report feeling sad or hopeless and nearly 3 times as likely to have considered attempting suicide [2]. These disparities may vary further by race and ethnicity. A 2021 national survey of SGM youth (aged 13-24 years) in the United States sponsored by the Trevor Project found that racial and ethnic minority SGM youth were more likely than White non-Hispanic SGM youth to have seriously considered suicide in the prior year [3]. Taken together, these data underscore the need to address the well-being of diverse SGM youth populations through innovative mental health interventions [4].

The minority stress model [5,6] provides a framework for understanding the higher prevalence of psychological distress and negative mental health outcomes for SGM youth, as well as for identifying interventions to improve SGM individuals' mental health [7]. The minority stress theory proposes that SGM health disparities can be explained in large part by discrimination from a hostile homophobic and transphobic culture, which creates stressors unique to minority identity [8]. These stressors include harassment, victimization, internalized homophobia and transphobia, and expectations of rejection. These disparities may be further compounded if individuals experience multiple minority stressors because of having >1 minority identity (eg, discrimination because of sexuality, gender, racial, and ethnic identity) [9-12]. For instance, racial and ethnic minority SGM youth may experience sexual or gender minority stress within their racial and ethnic communities while also experiencing racial and ethnic minority stress within lesbian, gay, bisexual, transgender, queer, and other SGM (LGBTQ+) communities.

The transactional model of stress and coping [13] notes that individuals' ability to respond to stress and reduce its impact on their well-being begins with an assessment of the stressor (ie, primary appraisal) and their confidence and ability to respond to the stressor (ie, secondary appraisal). Interventions designed to target and transform appraisals of stress from that of a threat to more of a challenge through cognitive and behavioral coping strategies have been shown to support the mental health and well-being of adolescents [14-16]. Efficacious mental health interventions for SGM youth have focused on providing resources that scaffold the ability of SGM youth to perceive minority stressors as a challenge to be faced and overcome rather than as a threat, including strengthening the coping skills of SGM youth, affirming SGM identities, and strengthening supportive social connections [17-19].

Although prior research suggests that face-to-face interventions that include these components may improve the mental health of SGM youth [20], the reach and scalability of these programs have been challenging, given their time intensity and need for synchronous interactions, which have become increasingly difficult to coordinate amidst the COVID-19 pandemic [21,22]. At the same time, the need for scalable mental health resources has become particularly acute in recent years. For example, data from a large US survey of teenagers conducted by Common Sense Media [23] indicated that the amount of time SGM youth spent searching for mental health information on the web substantially increased during the pandemic.

In recent years, technology-assisted interventions have been posited to help decrease implementation challenges by serving as supplemental strategies to face-to-face psychotherapy. For example, Lucassen et al [24] found that their modular computer-delivered cognitive behavioral therapy program was feasible, acceptable, and effective in their pilot study with 21 sexual minority adolescents, aged 13 and 19 years, in New Zealand. Other programs have sought to use web-based interventions to circumvent barriers to accessing affirming in-person services. For example, Craig et al [18] found that SGM youth in Canada (N=46; age 14-29 years) who participated in their 8-session, manualized, and synchronous pilot telehealth group intervention found the program to be acceptable. Although their design did not allow for randomization, preliminary efficacy analyses noted improvements in stress appraisals, cognitive and behavioral coping skills, and depressive symptomatology in the web-based group program when compared with youth in the wait-list control group. Taken together, these findings are promising and highlight technology's potential as a modality to deliver mental health interventions for SGM youth.

To date, there is limited research examining the efficacy of web-based platforms for helping SGM youth cope with minority stress asynchronously. For instance, in a pilot randomized trial, Schwinn et al [25] found that their tailored, 3-session web-based intervention resulted in decreases in perceived stress and increases in coping and problem-solving skills when compared with youth in the control arm at a 3-month follow-up. Egan et al [26] designed an innovative, web-accessible role-playing game intervention. Although they found high acceptability for their program among SGM youth participating in their randomized trial, they did not observe any improvements at the 1- or 2-month follow-up across coping skills, depression and anxiety symptoms, or knowledge and use of web resources.

Study Objectives

Although these findings suggest that SGM youth perceive web-based asynchronous interventions to be acceptable, there is a need to increase the empirical evidence base for the efficacy of these interventions [27]. Moreover, the generalizability of the aforementioned findings to racial and ethnic minority populations has been constrained, given the limited representation of these groups in prior research. Given the current state of the science, this study sought to test the acceptability and preliminary efficacy of an asynchronous web application—*imi*—among a predominantly racial and ethnic minority sample of SGM youth aged between 13 and 19 years living in the United States. The *imi* application was designed to facilitate SGM identity affirmation, promote a sense of connectedness to the LGBTQ+ community, and encourage cognitive and behavioral coping skill practice. In partnership with a racially and ethnically diverse group of SGM youth, we co-designed the *imi* application to be directly responsive to their needs by leveraging the visual, aesthetic, and interactive capacities of a web-based interface to deliver identity-affirming experiences that could support the intervention's engagement and efficacy.

Our study had 4 main objectives. First, we examined the acceptability of the *imi* application in a diverse sample of 270 SGM youth. Given our use of human-centered design principles and the involvement of SGM youth in the design of the *imi* application, we expected that participants randomized to receive the *imi* application would report greater acceptability and satisfaction than participants assigned to a resource-only version of the *imi* application (the control arm), which did not contain any of the newly created interactive coping and identity-affirming content designed with LGBTQ+ youth. Second, we examined the preliminary efficacy of the *imi* application as a digital tool for increasing adaptive stress appraisals among SGM youth (primary outcome). Given the *imi* application's focus on teaching cognitive and behavioral coping skills, we hypothesized that participants assigned to receive the *imi* application would be more likely to appraise stress as a surmountable challenge and less likely to appraise stress as threatening by the 4-week follow-up relative to the control arm. Third, we examined the preliminary efficacy of the *imi* application across five secondary outcomes related to the mental health of SGM youth: cognitive and behavioral coping skills, identity affirmation and connectedness to the LGBTQ+ community, internalization of blame for minority stress, sense of belonging, and anxiety and depression symptoms. We predicted that the *imi* application would be more likely to improve SGM youth's outcomes across these domains relative to the control arm. Finally, as exploratory analyses, we examined participants' engagement with the *imi* application relative to the control arm. We also explored whether participant engagement with the *imi* application (ie, counts of user sessions, time spent, and the number of pages visited) predicted improvement in primary and secondary outcomes.

Methods

Study Design

This pilot randomized controlled trial evaluated the acceptability and initial efficacy of the *imi* application at the end of the 4-week active study period. Participants were randomly assigned in a 1:1 fashion to receive either the *imi* application (treatment arm) or a resource page–only version of the *imi* site called "asterix," which linked out to a series of LGBTQ+-specific external mental health resources (resource-only control arm). We collected survey data via web-based self-completed Qualtrics surveys administered at baseline and at a 4-week follow-up.

Ethics Approval

The University of Pennsylvania Institutional Review Board approved all study procedures (protocol 849509), and the study was registered on ClinicalTrials.gov (NCT05061966). A waiver of parental consent was granted to ensure that youth who might not yet be out to their parents or have less parental support, and thus could benefit from an identity-affirming tool, could participate in the study.

Participants and Recruitment

Participants were SGM youth recruited through Instagram between October and November 2021. To be eligible for this study, the youth had to (1) be aged 13 to 19 years (inclusive); (2) identify as a sexual or gender minority; (3) reside within the United States; (4) have English literacy; (5) have access to a device with internet access, a web browser, and SMS text messaging capabilities; and (6) be willing to participate in study activities for 4 weeks.

A target sample size of 250 participants was selected to allow for the detection of arm differences in week 4 outcomes, which were medium to small in size or larger (Cohen $d \geq 0.35$) after accounting for the potential loss of participants because of attrition or noncompliance. These effect sizes are consistent with those observed in previous research on digital mental health tools for SGM youth [18,25].

Study Procedures

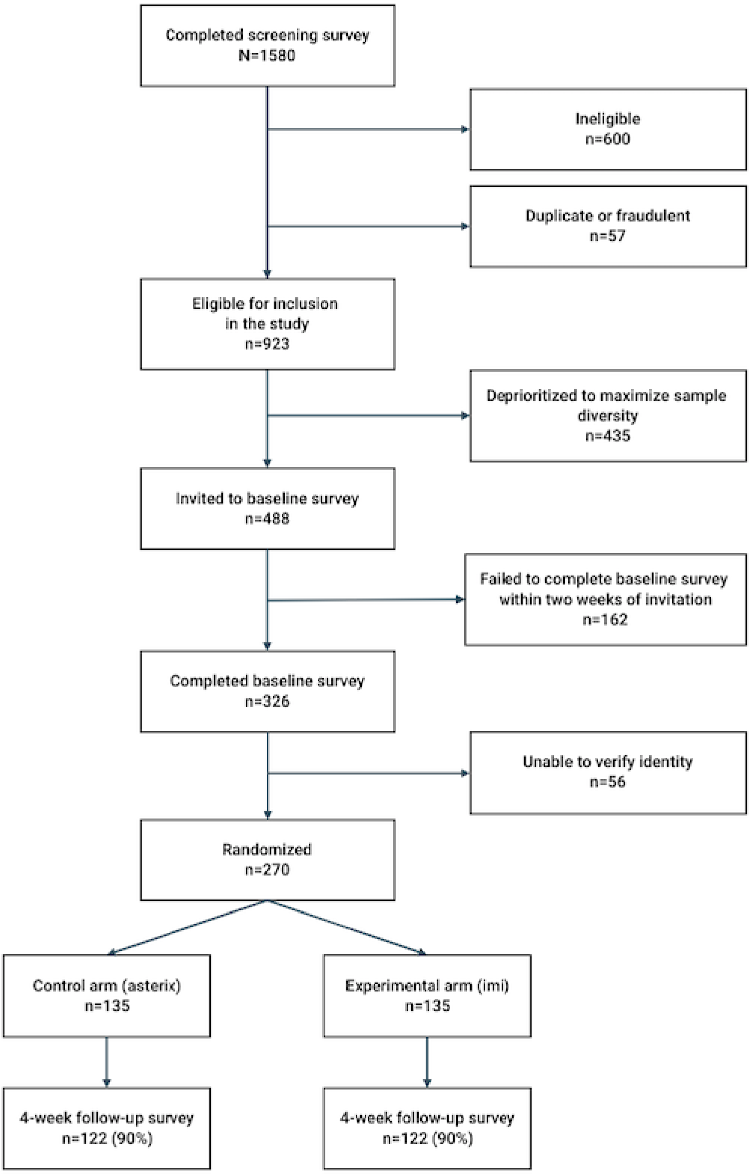
All study activities were conducted remotely, and web-based screening and survey assessments were delivered through the Qualtrics software. Prospective participants clicked on a paid advertisement and completed a brief screening survey. Individuals were emailed a link to the baseline survey, which contained the informed consent form (Multimedia Appendix 1), and participants were given 2 weeks to complete the survey. The baseline survey contained 8 of the same or similar questions as were asked in the screener. Following the established best practices for participant verification [28-30], the staff compared each applicant's screener and baseline data for these 8 questions, in addition to the metadata (eg, IP addresses registered in the United States, review of the time taken to complete the survey, and answers in hidden "honey pot" fields). The study staff also manually checked all screeners that met basic eligibility criteria to eliminate duplicate and fraudulent entries. If any significant inconsistencies were identified, applicants were emailed and asked to respond via email or phone to resolve the issue.

Once participants completed the baseline survey and passed the verification procedure, they were considered to be enrolled in the study. Enrolled participants were randomized in a 1:1 ratio into the *imi* (treatment; received full intervention content) or *asterix* (control; received pared-down resources-only version of the *imi* application) arm. Within 1 business day of completing the baseline survey, each participant was emailed a unique link to the *imi* or *asterix* web application. Participants were compensated with a US $30 Amazon e-gift card once they registered for an account on *imi* or *asterix*. The informed consent form encouraged participants to log into their digital resource at least twice a week and informed them that the study team would be able to track where they went within the web application, what features they used, what content they saw, and the duration of time they spent on the web application. Depending on their communication preferences, participants were emailed or texted every 7 days after enrollment to remind them to use the web application. On day 28, participants were sent a link to the follow-up survey, which they had 14 days to complete. Once completed, they received a US $40 Amazon e-gift card. Authors were not blind to participants' conditions during data collection or analysis; however, as all intervention activities were self-guided and all outcome measures were self-assessed by participants, there was no interaction between study staff and participants that could have led to response biases on the part of participants because of demand characteristics. No adverse events were reported during the trial.

Intervention Development Study Procedures

The *imi* application was built by Hopelab, a nonprofit social innovation laboratory, in collaboration with CenterLink, an international nonprofit organization and member-based association of LGBTQ+ centers serving their local and regional communities. Before launching the pilot trial, Hopelab conducted formative work through interviews, focus groups, co-design sessions, and surveys of SGM youth. The *imi* web application content and visual elements were tailored based on youth feedback and contributions. Screenshots containing example content are presented in Figure 1. The core intervention modules or "guides" in the *imi* application were designed based on a review of the existing efficacious minority stress interventions in the literature and honed through feedback with scientific advisors during a 6-month iterative design phase. This included extracting evidence-based exercises from the literature to support key psychosocial targets (eg, cognitive and behavioral coping skills and identity affirmation) that could be adapted to a digital platform and prototyping digital "mini-interventions" that would be further developed in the final *imi* application content.

Figure 1. CONSORT (Consolidated Standards for Reporting Trials) diagram.



Intervention Description

Experimental Arm

imi is a web application designed to facilitate LGBTQ+ identity affirmation, promote a feeling of connectedness to the LGBTQ+ community, and encourage cognitive and behavioral coping skill practice (Multimedia Appendix 2). The name *imi* (pronounced *eye-me*) is a nod to the idea that no matter who you are, you are you (ie, “I’m me”). The logo, designed as an ambigram that can be read in many orientations, represents the belief that even as one changes and evolves, they are exactly as they are. The *imi* application delivers fully automated information and skill practice across guides covering four content areas: (1) gender identity exploration (the *gender* guide), (2) sexual orientation and broader LGBTQ+ identity exploration (the *queerness* guide), (3) stress and coping (the *stress* guide), and (4) internalized homophobia and transphobia (the *stigma* guide). Examples of each guide’s goals and sample activities are as follows:

1. The *gender* guide allows youth to explore and affirm their gender identity and expression. For example, one activity uses a chat interface that allows users to experiment with different names and pronouns that fit their gender identity.
2. The *queerness* guide encourages youth to examine their LGBTQ+ identity through an intersectional lens and reflect upon what *queerness* means to them. For instance, one activity provides examples of the ways other LGBTQ+ youth define their queerness and guides the user through the creation of their own definition.
3. The *stress* guide provides psychoeducation on LGBTQ+ sources of minority stress (eg, discrimination, prejudice, and microaggressions) and teaches cognitive and behavioral coping skills through activities such as guided breathing, positive reframing exercises, and social support resources. Users are encouraged to select coping skills that work for them most effectively.
4. The *stigma* guide explains how negative and stereotypical messages become internalized and encourages users to examine and challenge their own internalized homophobia

and transphobia. For example, one activity scaffolds users in developing a personalized affirmation to help them combat negative self-talk.

Each of the four guides comprises four types of content: (1) learning segments, (2) activities, (3) community content, and (4) external resources. Learning segments provide information about LGBTQ+ relevant vocabulary (eg, pronouns and commonly used terms for sexual and gender identity), queer history, and psychoeducation on minority stress and internalized stigma. Activities include interactive exercises, such as chat interfaces, drawing activities, and guided relaxation and mindfulness practices. Community content includes video, audio, and written stories and images of LGBTQ+ youth. External resources connect youth to externally linked content designed for LGBTQ+ youth, such as the Trevor Project and the Gay, Lesbian, and Straight Education Network's *coming out* guides.

Both Arms

Participants in both arms received access to resource webpages that linked to freely accessible, preexisting crisis and noncrisis resources. Crisis resources included the National Suicide Prevention Lifeline, as well as resources specific to LGBTQ+ youth, such as TrevorChat. The noncrisis resources included moderated social networking and web-based chat spaces (eg, TrevorSpace and Q Chat Space), local LGBTQ+ centers (eg, CenterLink's center-finder tool), self-guided web resources (eg, The "It Gets Better Project"), databases of LGBTQ+-affirming therapists (eg, Gaylista), and a guide to free digital mental health tools (One Mind PsyberGuide). The resource section also contains *safer browsing tips*, which provide web-based privacy and safety advice specific to LGBTQ+ youth.

The control arm only received access to the resource webpages described previously and did not have access to any of the other content in the *imi* application. This pared-down, resources-only version of the *imi* site that was provided to the control group was named *asterix* This control allowed for a test of whether the learning, interactive, and community content of *imi* guides had benefits above and beyond any benefit that might be derived from simply having a curated, unified access point for existing, freely available web-based resources for SGM youth.

Participants in both arms were instructed to try to visit their respective web applications at least twice a week during the 4-week active trial period but could engage with the content available to them however they wished, in any order, at their discretion. On the basis of their preferences, participants received either weekly texts or emails reminding them to log into their web application.

Measures

Primary Outcomes: Stress Appraisals

The Stress Appraisal Measure for Adolescents [15] captures stress appraisals across 3 dimensions (challenge, threat, and resources). The 3-item Challenge subscale assesses perceptions of stress as a surmountable challenge (Cronbach α=.67). The 7-item Threat subscale measures perceptions of stress as having lasting, negative repercussions (α=.83). The 3-item Resources subscale assesses the belief that one has the necessary internal and external resources to cope with stress (α=.81). Responses to all items are recorded on a 5-point scale (1=*strongly disagree* to 5=*strongly agree*). A mean score was computed for each subscale, with higher values indicating greater endorsement of each respective stress appraisal.

Secondary Outcomes

Cognitive and Behavioral Coping Skills

Participants' use of specific cognitive and behavioral coping skills were measured with items adapted from the brief 2-item Coping Orientation to Problems Experienced (COPE) inventory by Carver [31], specifically the self-distraction (α=.46), active coping (α=.70), emotional support (α=.79), instrumental support (α=.78), venting (α=.66), positive reframing (α=.61), planning (α=.71), acceptance (α=.65), self-blame (α=.79), substance use (α=.96), and behavioral disengagement (α=.75) subscales. Instructions were modified such that participants indicated how they had been coping with stress in their lives over the past 2 weeks on a 4-point scale (1="I haven't been doing this at all" to 4="I've been doing this a lot"). A mean score was computed for each subscale, with higher values indicating greater use of that respective strategy.

Positive LGBTQ+ Identity

Two 5-item subscales from the Lesbian, Gay, and Bisexual (LGB) Positive Identity Measure [32] were used to measure positive LGBTQ+ identity factors. The first subscale, the *authenticity subscale* (α=.87), measured comfort with one's own LGBTQ+ identity (eg, "I am honest with myself about my LGBTQ+ identity"). The second subscale, the community subscale (α=.87), measured a sense of connection to the broader LGBTQ+ community (eg, "I feel supported by the LGBTQ+ community"). Items were modified from LGB to LGBTQ+ to be inclusive of a range of SGM identities. Participants responded on a 7-point scale (1=*strongly disagree* to 7=*strongly agree*), and a mean score was computed for each subscale, with higher scores indicating greater authenticity and connection, respectively.

Internalization of Blame for Minority Stress

The 5-item Coping with Discrimination Scale–Internalization subscale [33] assesses the tendency to blame oneself for instances of SGM-related discrimination (eg, "I tend to wonder if I did something to offend the others involved"). Participants responded on a 6-point scale (1=*never* to 6=*always*), and positively worded items were reverse coded. We computed a mean internalization score, where higher scores indicate greater internalization of blame for minority stress (α=.84).

Sense of Belonging

A 5-item version of the Thwarted Belongingness subscale of the Interpersonal Needs Questionnaire [34] was used to measure perceived belonging. Participants responded to the items (eg, "These days, I feel disconnected from other people.") on a 7-point scale (1=*not at all true*" to 7=*very true for me*). Positively worded items were reverse coded. We computed a sum score ranging from 5 to 35, where higher scores indicate a lack of sense of belonging (α=.81).

Anxiety and Depression Symptoms

The 7-item General Anxiety Disorder Scale [35] was used to examine symptoms of anxiety (eg, “Feeling nervous, anxious, or on edge”), and the 8-item Patient Health Questionnaire [36] was used to assess depressive symptoms (eg, “Feeling tired or having little energy”). These brief clinical measures have been used to screen for generalized anxiety and depression across a wide range of populations, including adolescents [37-39]. For both measures, participants rated the frequency of their symptoms over the past 2 weeks using a 4-point scale (0=*not at all* to 3=*nearly every day*). Items were summed to compute total scores for each construct, with higher scores indicating greater symptoms of anxiety and depression, respectively (General Anxiety Disorder-7 α=.88; Patient Health Questionnaire-8 α=.83).

Intervention Acceptability and Satisfaction

Participants rated the acceptability and their satisfaction with the *imi* and *asterix* applications at the 4-week follow-up. A modified version of the LGBTQ Appropriateness Scale [40] was used to assess the perceived relevancy and appropriateness of the web resources to SGM youth (eg, “This product appears to be relevant to people who identify as LGBTQ+”) using a 7-point scale (1=*strongly disagree* to 7=*strongly agree*). These questions were tailored for each web application’s features; we asked 12 items for the treatment arm (α=.91) and 9 items for the control arm (α=.86). Items were averaged, with a higher score indicating greater perceived appropriateness.

A measure of intervention satisfaction and suggestions for improvement created for this study were also included. The measure comprises a multiple-choice question (eg, “How would you rate your overall experience of this product?”; 1=*very bad* to 7=*excellent*) that was analyzed as a continuous variable and free-text responses (eg, “How could this product be more helpful to you?” [text box]).

A net promoter score (NPS; eg, “How likely would you be to recommend [*imi/asterix*] to a LGBTQ+ friend?”) was administered to further assess the perceived value of the interventions. Respondents answered on an 11-point scale (0=*not at all likely* to 10=*extremely likely*). Following established industry conventions for NPS, responses were recoded such that respondents who selected 9 or 10 were categorized as “Promoters,” those selecting 7 or 8 as “Passives,” and all others as “Detractors.”

Web Application Engagement

Participants’ actions in the *imi* and *asterix* applications were collected as paradata over the trial period. Each participant was provided with a unique link to their respective intervention. This link embedded participants’ study IDs as metadata within their individual accounts, allowing us to track how much time each participant spent in their web application, which pages they viewed, and which links they clicked. These paradata were transformed to characterize the amount, frequency, duration, and depth of engagement with the web-based intervention [41].

In this study, we derived four paradata metrics: (1) counts of user sessions, (2) time spent on each intervention, (3) the number of pages visited, and (4) external links clicked. Sessions were counted whenever there was a period of activity within the app, with a participant having the same session ID until they had a period of ≥15 minutes or more of inactivity with the application. We derived the variable *number of sessions* by counting the number of discrete sessions in which the participant engaged during the 4-week active trial period.

To measure time spent on the intervention, we tracked the number of minutes and seconds participants spent logged into their respective web resources during the active trial period. Each time a participant visited a page within the app, there was a record of their activity. We derived the variable *number of unique pages viewed* by summing the number of distinct pages a participant visited during the active trial period. Finally, each time a participant clicked on a link, a unique record of the click was generated. We summed the number of unique external links a participant clicked on during the active trial period to derive this value.

Analytic Strategy

Overview

Descriptive analyses were conducted to summarize demographic characteristics among the study arms at baseline. Preliminary analyses tested baseline equivalence between the study arms on demographics, assessment of attrition, and differential attrition by study arm. We used SAS (version 9.4; SAS Institute) to conduct all analyses.

Intervention Acceptability

To test the interventions’ acceptability, we compared participants’ ratings of the 2 web resources (eg, intervention satisfaction and NPSs) using chi-square tests for categorical variables and Student *t* tests for continuous variables. Within the treatment arm, open-ended feedback was coded by 2 coders using rapid qualitative analysis methods [42]. Core questions guiding the coding included “What did you find most helpful about *imi*?” and “If you could change anything about *imi* or the guides in it, what would you change and why?” Quotes were chosen to illustrate salient themes.

Preliminary Efficacy

Analyses of all primary and secondary outcome variables were performed using an intention-to-treat approach, which included all available data from participants randomly assigned to the 2 arms, regardless of whether participants created an account within their respective web resources. Our primary efficacy analysis sought to examine whether there were differences in our primary and secondary outcomes between the 2 arms. We used linear regression to test the main effect of arm (treatment=1 vs control=0) on week 4 outcomes, adjusting for the baseline value of each respective outcome as a covariate. Recognizing that we were testing 2 versions of a web application (the resource-only version of *imi* called *asterix* and the full interactive *imi* intervention), we also tested for changes over time within each web application. For these within-arm analyses, we examined the mean changes from baseline to follow-up using paired *t* tests.

Engagement

Given the absence of standardized and generalizable threshold indicators to suggest adequate engagement across digital health interventions, we adopted an exploratory approach to the analysis of these data and created thresholds to define participants' engagement with the intervention. After examining the distribution of the engagement data, we selected the following to define thresholds of use: ≥4 sessions, ≥10 minutes, ≥10 unique pages viewed, and ≥1 external link clicked. Given the high bivariate correlations between these indicators of engagement (Spearman $\rho>0.65$), as well as the exploratory nature of these analyses, we compared each engagement indicator separately by study arm using chi-square tests. Of note, treating the nonnormal data as continuous (with and without transformation of these data) yielded similar results.

Finally, we explored whether reaching these thresholds of use within the *imi* arm predicted differential gains across primary and secondary outcomes. We focused these analyses on three indicators capturing participants' engagement within *imi*: time spent on site, number of sessions completed, and number of pages viewed, which captured the depth of participant engagement with that content which was unique to the treatment arm. We ran separate regression models because of the high multicollinearity among engagement metrics and the importance of examining the different scopes of paradata (eg, the amount, frequency, duration, and depth of engagement). All models accounted for the baseline value of each respective outcome as a covariate.

Results

Screening and Enrollment

Of the 1580 individuals who completed the screening survey, 923 (58.4%) met all inclusion criteria and passed the duplicate and fraudulent entry checks (Multimedia Appendix 3). From this pool of 923 eligible participants, racial, ethnic, and gender minority youth were selectively invited to access the baseline survey to achieve a diverse participant pool.

We invited 488 participants in total to complete the baseline survey to ensure that the target enrollment of 250 participants would be reached, with the expectation that not all participants who expressed interest in participating would respond to further outreach or pass identity verification checks.

Of the 488 participants who were invited to complete the baseline survey, 162 (33.2%) failed to do so within the 2-week window, and another 56 (11.5%) participants did not pass the participant verification procedure (ie, there were discrepancies between information entered on screener and baseline surveys); thus, in total, 270 (55.3%) participants completed the baseline survey and were enrolled in the study.

Sample Characteristics

Participants had a mean age of 16.49 (SD 1.49) years. Most of the participants resided in a metropolitan area, with the majority residing in the Southern or Western regions of the United States. A large proportion of participants identified with multiple races, gender identities, and sexual orientations. To represent the diversity and heterogeneity of the sample, we report these variables in a nonmutually exclusive fashion in Table 1. The sample was racially and ethnically diverse, with 77.8% (210/270) of participants identifying as racial or ethnic minorities. Similarly, 41.9% (113/270) and 39.6% (107/270) of the sample identified with multiple gender identities and sexual orientations, respectively. Nonbinary (94/270, 34.8%) and bisexual (96/270, 35.6%) were the response options selected most frequently. Participants expressed various levels of outness about their sexual orientation, with 26.3% (71/270) of participants noting that they were completely or mostly in the closet, whereas 39.3% (106/270) were mostly or fully out.

Table 1. Sociodemographic characteristics by study arm (N=270).

Characteristics	All (N=270)	Control (n=135)	Intervention (n=135)
Age (years), mean (SD)	16.49 (1.49)	16.42 (1.51)	16.56 (1.46)
Geographic region, n (%)			
Metropolitan	249 (92.2)	126 (93.3)	123 (91.1)
Micropolitan	14 (5.2)	7 (5.2)	7 (5.2)
Small town	4 (1.5)	1 (0.7)	3 (2.2)
Rural areas	3 (1.1)	1 (0.7)	2 (1.5)
Census region, n (%)			
Northeast	37 (13.7)	16 (11.9)	21 (15.6)
Midwest	53 (19.6)	29 (21.5)	24 (17.8)
South	93 (34.4)	46 (34.1)	47 (34.8)
West	87 (32.2)	44 (32.6)	43 (31.9)
Education[a], n (%)			
Kindergarten to 5th grade	3 (1.1)	2 (1.5)	1 (0.7)
6th to 8th grade	98 (36.3)	48 (35.6)	50 (37)
9th to 11th grade	100 (37)	51 (37.8)	49 (36.3)
High school diploma or equivalent	52 (19.3)	26 (19.3)	26 (19.3)
Some postsecondary education	17 (6.3)	8 (5.9)	9 (6.7)
Subjective SES[b], n (%)			
Wealthy	1 (0.4)	1 (0.7)	0 (0)
Upper-middle class	43 (15.9)	18 (13.3)	25 (18.5)
Middle class	119 (44.1)	57 (42.2)	62 (45.9)
Working class	61 (22.6)	33 (24.4)	28 (20.7)
Low income or poor	31 (11.5)	19 (14.1)	12 (8.9)
I prefer not to respond	15 (5.6)	7 (5.2)	8 (5.9)
Sex at birth, n (%)			
Male	61 (22.6)	32 (23.7)	29 (21.5)
Female	209 (77.4)	103 (76.3)	106 (78.5)
Living status, n (%)			
Living with parent, parents, guardian, or guardians	222 (82.2)	109 (80.7)	113 (83.7)
Other	48 (17.8)	26 (19.3)	22 (16.3)
Race and ethnicity[c] (total count), n (%)			
American Indian or Alaska Native	14 (5.2)	10 (7.4)	4 (3)
Asian	68 (25.2)	33 (24.4)	35 (25.9)
Black or African American	64 (23.7)	29 (21.5)	35 (25.9)
Hispanic or Latinx	73 (27)	42 (31.1)	31 (23)
Middle Eastern or North African	9 (3.3)	7 (5.2)	2 (1.5)
Native Hawaiian or other Pacific Islander	2 (0.7)	1 (0.7)	1 (0.7)
White or Caucasian	130 (48.2)	64 (47.4)	66 (48.9)
Other	9 (3.3)	4 (3)	5 (3.7)
Racial or ethnic minority, n (%)			
Exclusive identifying as non-Hispanic White	60 (22.2)	28 (20.7)	32 (23.7)

Characteristics	All (N=270)	Control (n=135)	Intervention (n=135)
Identifying as racial or ethnic minority	210 (77.8)	107 (79.3)	103 (76.3)
Gender identity[c] (total count), n (%)			
Agender	6 (2.2)	2 (1.5)	4 (3)
Cisgender man	28 (10.4)	15 (11.1)	13 (9.6)
Cisgender woman	43 (15.9)	25 (18.5)	18 (13.3)
Genderqueer	44 (16.3)	22 (16.3)	22 (16.3)
Man	41 (15.2)	13 (9.6)	28 (20.7)
Woman	41 (15.2)	21 (15.6)	20 (14.8)
Nonbinary	94 (34.8)	51 (38.8)	43 (31.9)
Transgender man	50 (18.5)	21 (15.7)	29 (21.5)
Transgender woman	8 (3.0)	5 (3.7)	3 (2.2)
Other	30 (11.1)	13 (9.6)	17 (12.6)
Gender identity, n (%)			
Not questioning	218 (80.7)	107 (79.3)	111 (82.2)
Questioning	52 (19.3)	28 (20.7)	24 (17.8)
Multiple gender identities, n (%)			
Multiple identities	113 (41.9)	56 (41.5)	57 (42.2)
Single identity	157 (58.1)	79 (58.5)	78 (57.8)
Sexual orientation[c] (total count), n (%)			
Asexual	42 (15.6)	16 (11.9)	26 (19.3)
Bisexual	96 (35.6)	48 (35.6)	48 (35.6)
Gay	47 (17.4)	27 (20)	20 (14.8)
Lesbian	47 (17.4)	23 (17)	24 (17.8)
Pansexual	40 (14.8)	21 (15.6)	19 (14.1)
Queer	73 (27)	40 (29.6)	33 (24.4)
Straight or heterosexual	4 (1.5)	1 (0.7)	3 (2.2)
Other	21 (7.8)	8 (5.9)	13 (9.6)
Sexual orientation, n (%)			
Not questioning	225 (83.3)	110 (81.5)	115 (85.2)
Questioning	45 (16.7)	25 (18.5)	20 (14.8)
Multiple sexual orientations, n (%)			
Multiple identities	107 (39.6)	54 (40)	53 (39.3)
Single identity	163 (60.4)	81 (60)	82 (60.7)
Outness, n (%)			
Definitely in the closet	24 (8.9)	12 (8.9)	12 (8.9)
In the closet most of the time	47 (17.4)	19 (14.1)	28 (20.7)
Half in the closet, half out of the closet	93 (34.4)	51 (37.8)	42 (31.1)
Out of the closet most of the time	76 (28.2)	33 (24.4)	43 (31.9)
Completely out of the closet	30 (11.1)	20 (14.8)	10 (7.4)

[a] The highest level of education completed.

[b] SES: socioeconomic status.

[c] Nonmutually exclusive categories; participants were allowed to select all that apply.

Baseline Equivalence, Attrition, and Differential Attrition

Randomization resulted in baseline equivalence between the treatment and control arms on all demographics, primary and secondary outcomes. Our survey retention rate at the 4-week follow-up was 90.4% (244/270). In attrition analyses, comparing those who completed the follow-up survey (244/270, 90.4%) with those who did not (26/270, 9.6%), we found no significant condition differences in attrition linked to demographic characteristics or baseline scores on primary or secondary outcomes. Collapsing across the arms, participants who did not complete the follow-up survey were more likely to be younger (mean 15.42, SD 1.53 years vs mean 16.60, SD 1.44 years; 2-tailed $t_{268}=–3.95$; $P<.001$) and reported fewer cognitive and behavioral coping skills at baseline (Multimedia Appendix 4; all *P* values <.01).

Intervention Acceptability

Participants perceived the content of the intervention as appropriate to LGBTQ+ individuals, both in the treatment (mean 6.01, SD 0.85) and control arm (mean 5.85, SD 0.83; $t_{241}=–1.45$; $P=.15$). Participants in the treatment arm rated their overall experience with the *imi* application (mean 6.01, SD 1.06) more positively than participants in the control arm (mean 5.59, SD 1.14; $t_{241}=–2.98$; $P=.003$) and were also more likely to report that they would recommend the *imi* application to a friend ("Detractors" treatment: 26/121, 21.5%; "Detractors" control: 45/121, 37.2%; $\chi^2_2=8.9$; $P=.01$).

When asked what they liked and found most useful about the *imi* application, participants remarked on the benefits of hearing the stories and seeing the images of other LGBTQ+ youth:

> *I liked being able to read other queer folks' stories—no matter how many friends I have, it's always nice to hear about other people with experiences or identities similar to my own and learn from what they've done.*

Additional themes included being taught concrete strategies for managing stress and engaging with the activities (eg, the interactive chats and questionnaires) that encouraged identity exploration:

> *It gave lots of suggestions for stress relief, so I could focus on tackling stress myself. I don't like turning to other people for help that much, so the self-aspect of it was helpful.*
>
> *I really liked the activities—especially the one where I got to test out a new name. It made me feel seen.*

When asked what they would change or improve about the *imi* application, participants expressed a desire for more content and content on additional topics:

> *I'd add more content, the content is slightly lacking for now.*
>
> *I would also add a relationship section. Being in a relationship as a person in the LGBT community, there is a great need for knowledge on having thriving romantic, sexual, and even platonic relationships*

They also expressed interest in the addition of social networking and other interactive features:

> *I'd like to be able to interact with more people, not just the automatic responses...*
>
> *It would be cool if imi could track your location and find support groups in my area or a group chat to join with fellow lgbt people my age.*

Another common critique was that the *imi* application felt tailored to youth beginning to explore their identities and less suitable for those who are already more affirmed:

> *I believe this resource is excellent for someone part of the LGBT community who is questioning themselves and actively needs help or would benefit from it. If you're already out of the closet and comfortable with who you are, it may not be very beneficial.*

Preliminary Efficacy of the Intervention

Primary Outcome: Stress Appraisals

Stress appraisal improved over time across both arms. The treatment arm experienced significant improvements in challenge ($t_{121}=4.51$; $P<.001$), threat ($t_{121}=–2.73$; $P=.007$) and resource appraisals ($t_{121}=4.24$; $P<.001$) from baseline to the end of the follow-up. The control arm also showed improvements in challenge ($t_{121}=1.96$; $P=.052$) and resource appraisals ($t_{121}=2.83$; $P=.005$; Table 2).

Controlling for baseline scores (Table 3), the treatment arm reported significantly higher challenge appraisals at follow-up than the control arm (Cohen $d=0.25$; coefficient for treatment arm $b=0.26$; $P=.008$). We did not observe a difference between the study arms for threat appraisals ($d=0.10$; $b=–0.06$; $P=.37$) or resource appraisals ($d=0.15$; $b=0.14$; $P=.14$).

Table 2. Within-arm changes in primary and secondary outcomes from baseline to the 4-week follow-up (N=244).

Outcomes	Control				Intervention			
	Baseline (n=135), mean (SD)	Follow-up (n=122), mean (SD)	*t* test (*df*)	*P* value[a]	Baseline (n=135), mean (SD)	Follow-up, (n=122), mean (SD)	*t* test (*df*)	*P* value[a]
Primary outcomes								
Stress appraisals								
Challenge	3.15 (0.88)	3.32 (0.93)	1.96 (121)	.052	3.30 (0.77)	3.64 (0.85)	4.51 (121)	<.001
Threat	4.03 (0.72)	3.92 (0.69)	–1.79 (121)	.08	3.99 (0.73)	3.85 (0.77)	–2.73 (121)	.007
Resource	3.42 (1.00)	3.67 (0.90)	2.83 (121)	.005	3.46 (0.98)	3.83 (0.92)	4.24 (121)	<.001
Secondary outcomes								
Cognitive and behavioral coping skills								
Self-distraction	3.20 (0.72)	3.23 (0.65)	0.25 (121)	.80	3.29 (0.70)	3.23 (0.65)	–1.33 (121)	.19
Active coping	2.46 (0.82)	2.47 (0.72)	0.22 (121)	.83	2.46 (0.78)	2.65 (0.84)	1.80 (121)	.06
Emotional support	2.31 (0.86)	2.43 (0.94)	1.88 (121)	.06	2.41 (0.89)	2.59 (0.94)	2.30 (121)	.02
Instrumental support	2.27 (0.86)	2.30 (0.87)	0.65 (121)	.52	2.32 (0.95)	2.64 (0.90)	3.15 (121)	.002
Venting	2.43 (0.83)	2.42 (0.82)	0.53 (121)	.57	2.43 (0.83)	2.48 (0.86)	0.91 (121)	.36
Positive reframing	2.22 (0.88)	2.25 (0.78)	0.17 (121)	.87	2.23 (0.82)	2.45 (0.86)	2.98 (121)	.003
Planning	2.57 (0.85)	2.56 (0.86)	–0.24 (121)	.81	2.59 (0.96)	2.79 (0.83)	2.80 (121)	.006
Acceptance	2.79 (0.82)	2.73 (0.77)	–1.13 (121)	.26	2.78 (0.79)	2.87 (0.77)	0.70 (121)	.49
Self-blame	3.04 (0.88)	2.76 (0.88)	–3.73 (121)	.003	3.07 (0.92)	2.91 (0.87)	–1.71 (121)	.09
Substance use	1.42 (0.81)	1.38 (0.78)	–1.16 (121)	.25	1.27 (0.69)	1.33 (0.74)	0.63 (121)	.53
Behavioral disengagement	2.16 (0.84)	1.99 (0.82)	–1.67 (121)	.10	2.13 (0.90)	2.01 (0.80)	–1.26 (121)	.21
Positive LGBTQ+[b] identity								
Authenticity	5.06 (1.38)	5.08 (1.30)	0.20 (121)	.84	5.04 (1.38)	5.10 (1.33)	1.17 (121)	.24
LGBTQ+ community	4.94 (1.35)	4.97 (1.18)	0.47 (121)	.64	4.69 (1.36)	4.82 (1.23)	1.44 (121)	.15
Internationalization of minority stress								
Internalization	3.21 (1.24)	2.94 (1.23)	–1.97 (121)	.051	3.29 (1.25)	3.19 (1.27)	–0.79 (121)	.43
Sense of belonging								
Thwarted belongingness	19.58 (6.12)	18.21 (6.24)	–2.53 (121)	.01	18.83 (6.29)	17.15 (6.99)	–3.53 (121)	<.001
Anxiety and depression symptoms								
Anxiety[c]	11.53 (5.31)	10.30 (5.78)	–2.55 (121)	.01	11.74 (5.44)	9.92 (5.56)	–4.42 (121)	<.001
Depression[c]	12.40 (5.97)	11.45 (5.75)	–2.59 (121)	.01	13.00 (5.31)	11.61 (5.95)	–3.33 (121)	.001

[a] Paired *t* test.

[b] LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other sexual and gender minority.

[c] Considered as continuous variables in the analyses.

Table 3. Between-arm differences in week 4 primary and secondary outcomes (N=244).

Outcomes	Cohen *d*	Modeling differences by arm[a]	
		Coefficients	*P* value
Primary outcomes			
Stress appraisals			
Challenge	0.25	0.26	.008
Threat	0.1	–0.06	.37
Resource	0.15	0.15	.14
Secondary outcomes			
Cognitive and behavioral coping skills			
Self-distraction	0.15	–0.04	.62
Active coping	0.16	0.17	.07
Emotional support	0.04	0.09	.37
Instrumental support	0.24	0.29	.005
Venting	0.03	0.05	.63
Positive reframing	0.27	0.22	.02
Planning	0.26	0.23	.02
Acceptance	0.17	0.14	.13
Self-blame	0.13	0.13	.18
Substance use	0.16	0.05	.55
Behavioral disengagement	0.02	0.02	.86
Positive LGBTQ+[b] identity			
Authenticity	0.09	0.06	.57
LGBTQ+ community	0.09	–0.002	.99
Internationalization of minority stress			
Internalization	0.07	0.15	.29
Sense of belonging			
Thwarted belongingness	0.08	–0.64	.34
Anxiety and depression symptoms			
Anxiety[c]	0.14	–0.55	.31
Depression[c]	0.07	–0.17	.74

[a] The effect of study arm (*imi* vs *asterix*) on the outcome at follow-up controlling for the outcome at baseline.

[b] LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other sexual and gender minority.

[c] Considered as continuous variables in the analyses.

Secondary Outcomes

Cognitive and Behavioral Coping Skills

At the 4-week follow-up, the treatment arm showed significant improvements in emotional support (t_{121}=2.30; *P*=.02), instrumental support (t_{121}=3.15; *P*=.002), positive reframing (t_{121}=2.98; *P*=.003), and planning to cope (t_{121}=2.80; *P*=.006). We observed no other changes over time in the other COPE subscales within the treatment arm. The control arm had significant reductions in self-blame (t_{121}=–3.73; *P*=.003). We observed no other changes over time in the other COPE subscales within the control arm.

The effect of the intervention on cognitive and behavioral coping skills was greater among the treatment arm than the control arm (instrumental support: *d*=0.24, *b*=0.29, *P*=.005; positive reframing: *d*=0.27, *b*=0.22, *P*=.02; planning: *d*=0.26, *b*=0.23, *P*=.02). However, we observed no differences in emotional support (*d*=0.04; *b*=0.09; *P*=.37) or self-blame (*d*=0.13; *b*=0.13; *P*=.18) between the arms. We observed no other differences between the arms in the COPE subscales.

Positive LGBTQ+ Identity

We did not observe any significant changes over time in the authenticity or the community subscales of the LGB Positive Identity Measure in either arm, nor did we observe differences in improvements between the arms.

Internalization of Blame for Minority Stress

We did not observe reductions in internalization of blame for minority stress over time in either arm or differences between the two arms.

Sense of Belonging

At the 4-week follow-up, both the treatment arm (t_{121}=–3.53; *P*<.001) and the control arm (t_{121}=–2.53; *P*=.01) showed significant reductions in thwarted belongingness. We did not observe differences in reductions of thwarted belongingness between the arms (*d*=–0.08; *b*=–0.64; *P*=.34).

Anxiety and Depression Symptoms

We also observed reductions in anxiety and depression symptoms for both arms. Among the treatment arm, we found significant reductions in anxiety (t_{121}=–4.42; *P*<.001) and depression (t_{121}=–3.33; *P*=.001) from baseline to week 4. We found similar results among control arm participants (reductions in anxiety: t_{121}=–2.55, *P*=.01; reductions in depression: t_{121}=–2.35, *P*=.01). However, the treatment arm did not report significantly lower anxiety (*d*=0.14; *b*=–0.55; *P*=.31) or depression (*d*=0.07; *b*=–0.17; *P*=.74) at the follow-up than the control arm.

Intervention Engagement

Approximately 98.5% (133/135) of the participants in the treatment arm and 97.8% (132/135) of the participants in the control arm created an account in their respective web resource within 4 weeks of being invited to access it. Participants in the treatment arm did not log significantly more sessions than participants in the control arm (t_{268}=–1.84; *P*=.07); however, they spent significantly more time in the product (t_{268}=–7.08; *P*<.001) and viewed more pages (t_{268}=–10.30; *P*<.001). Similarly, although there were no significant differences between the treatment and control arms in thresholds of use for the number of sessions logged (≥5 sessions; χ^2_1=1.0; *P*=.39), significantly more participants in the treatment arm spent ≥10 minutes in the product (χ^2_1=49.2; *P*<.001) and viewed >10 unique pages (χ^2_1=101.9; *P*<.001) than participants in the control arm (Table 4). Although the treatment arm showed higher engagement than the control overall, the control arm participants clicked on more unique external links than the treatment arm (t_{268}=4.51; *P*<.001) and were more likely to click on at least one external link (χ^2_1=13.1; *P*<.001).

Table 4. Engagement metrics by arm (N=270).

Engagement metrics	Control (n=135)	Intervention (n=135)	*P* value[a]
Sessions			
Total sessions completed, median (range)	3 (0-19)	4 (0-18)	.07
Low (0-4 sessions), n (%)	84 (62.2)	76 (56.3)	.39[b]
High (≥5 sessions), n (%)	51 (37.8)	59 (43.7)	N/A[c]
Time			
Total time spent (minutes), median (range)	3.08 (0.03-37.70)	12.14 (0.48-152.35)	<.001
Low (0-10 minutes), n (%)	118 (87.4)	64 (47.4)	<.001[b]
High (>10 minutes), n (%)	17 (12.6)	71 (52.6)	N/A
Unique pages[d]			
Unique pages viewed, median (range)	5 (1-8)	13 (1-50)	<.001
Low (0-10 pages), n (%)	135 (100)	61 (45.2)	<.001[b]
High (>10 pages), n (%)	0 (0)	74 (54.8)	N/A
External links			
Number of links clicked, median (range)	1 (0-14)	0 (0-6)	<.001
None (0), n (%)	68 (50.4)	97 (71.9)	<.001[b]
Any (>0), n (%)	67 (49.6)	38 (28.2)	N/A

[a] Student *t* test for continuous variables and chi-square tests for categorical variables.

[b] Compares dichotomized engagement (high vs low) and study arm (control vs intervention).

[c] N/A: not applicable. Refer to the *P* value for low engagement for a statistical comparison of high and low engagement by study arm.

[d] The maximum number of unique pages in the control web application (*asterix*) was 8 and the maximum for the *imi* application was 73.

Engagement and Outcome Improvements Within the *imi* Application Arm

Exploratory analyses examining outcomes within the treatment arm revealed a positive relationship between the 3 engagement indicators and several of the primary and secondary outcomes (Table 5).

Table 5. Linear regressions examining differences in primary and secondary outcomes by engagement indicators among participants assigned to the *imi* application (n=122).

Outcomes	The number of sessions (high [≥5] vs low[a])		Total minutes (high [>10] vs low[a])		The number of unique pages viewed (high [>10] vs low[a])	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Primary outcomes						
Stress appraisals						
Challenge	0.16	.25	0.42	.003	0.38	.008
Threat	0.03	.76	–0.13	.21	–0.13	.22
Resource	0.33	.03	0.50	.001	0.51	.001
Secondary outcomes						
Cognitive and behavioral coping skills						
Self-distraction	0.17	.13	0.18	.11	0.07	.51
Active coping	0.15	.31	0.28	.06	0.18	.23
Emotional support	0.01	.95	0.32	.03	0.24	.10
Instrumental support	–0.03	.87	0.28	.06	0.17	.26
Venting	0.13	.36	0.24	.09	0.22	.13
Positive reframing	0.14	.34	0.08	.60	0.05	.76
Planning	–0.01	.94	0.25	.046	–0.01	.94
Acceptance	0.08	.52	0.10	.41	–0.01	.93
Self-blame	–0.02	.91	–0.32	.02	–0.22	.13
Substance use	–0.10	.41	–0.05	.67	–0.05	.66
Behavioral disengagement	–0.11	.42	–0.11	.39	–0.08	.54
Positive LGBTQ+[b] identity						
Authenticity	–0.03	.84	0.26	.12	–0.13	.44
LGBTQ+ community	0.11	.51	0.27	.09	0.03	.88
Internationalization of minority stress						
Internalization	0.005	.98	–0.03	.90	0.20	.41
Sense of belonging						
Thwarted belongingness	–1.51	.11	–1.30	.17	–0.21	.83
Anxiety and depression symptoms						
Anxiety[c]	0.17	.82	–1.12	.13	–0.45	.55
Depression[c]	0.10	.90	–0.55	.48	0.23	.77

[a] The effect of engagement (reached thresholds of use vs not reached) on the outcome at follow-up, controlling for outcome at baseline.

[b] LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other sexual and gender minority.

[c] Considered as continuous variables in the analyses.

Sessions

Controlling for baseline scores, participants in the treatment group who engaged in ≥5 sessions during the intervention period had greater improvements in resource appraisals (*b*=0.33; *P*=.03) at follow-up than those who engaged in <5 sessions.

Total Minutes

Controlling for respective baseline scores, participants who spent >10 minutes in the *imi* application during the intervention period had significantly greater improvements in challenge appraisals (*b*=0.42, *P*=.003), resource appraisals (*b*=0.50;

P=.001), and emotional support coping skills (*b*=0.32; *P*=.03) at follow-up than those who spent ≤10 minutes.

Unique Pages

Participants who viewed >10 unique pages within the *imi* application during the intervention period had significantly greater improvements in challenge appraisals (*b*=0.38, *P*=.008) and resource appraisals (*b*=0.51; *P*=.001) at follow-up than those who viewed ≤10 pages, controlling for baseline scores.

Discussion

Principal Findings

Digital interventions show promise in supporting the mental health of SGM youth, given the appeal of the modality, their suitability to deliver engaging asynchronous content, and their unique capacity to reach large numbers of SGM youth, including youth who may be unable to access in-person services because of transportation concerns, availability of local services, or concerns about privacy and stigma. In this study, we found high acceptability for both the full *imi* intervention web application and the resource page–only control, with participants reporting greater satisfaction and engagement with the *imi* web application. SGM youth assigned to the *imi* web application also had greater improvements in stress appraisals and coping skills. Given the potential for digital interventions such as the *imi* application to support the mental health needs of SGM youth, we discuss our findings in detail in the following sections and note opportunities to advance work in this area.

Participants indicated that the *imi* application's content was acceptable and well-suited for SGM youth populations. Consistent with our hypothesis, participants randomly assigned to receive the *imi* application were more likely than those assigned to the resource-only control arm to report a positive experience with the intervention and to recommend it to SGM youth friends. These sentiments were echoed in the qualitative feedback, in which participants highlighted the benefits of viewing the stories of other SGM youth and learning approaches for coping with stress. Participants also offered several ways of improving the *imi* application, including developing greater breadth (eg, more overall content) and depth (eg, expansion to include a section focused on LGBTQ+ relationships) of content, and the addition of new features (eg, the ability to interact with other SGM youth through the web application). Collectively, these findings align with the growing body of evidence suggesting that SGM youth consider digital interventions to be acceptable modalities through which to receive mental health support [18,19,24]. They also point to concrete directions for intervention improvement.

For individuals to respond effectively against a stressor, they must be able to feel confident in their ability to address it [43]. Consistent with our hypothesis, SGM youth in the *imi* application arm were more likely than their peers in the control arm to appraise stressful situations as a surmountable challenge at the 4-week follow-up. The *imi* application arm was also less likely to report appraising stressful events as having lasting, negative repercussions. Participants in both study arms reported gains in having the internal and external resources to deal with stressful situations. The absence of a significant difference in resource appraisals between the 2 arms may be explained by the fact that both sites contained the same *imi*-based curated list of resources tailored to SGM youth. Taken together, these findings are promising and suggest that the *imi* application may help SGM youth situate stressors in their lives as transient and addressable, which, in turn, may reduce the acuity of these experiences on their mental health.

We also observed important initial changes in SGM youth's coping skills after the 4-week follow-up period. Compared with the control arm, participants in the *imi* application arm reported greater gains in their ability to seek out instrumental support, positively reframe challenging situations, and engage in planning as coping skills. Interestingly, we observed improvements in measured areas of mental health across both arms (eg, anxiety and depression symptoms), suggesting that web applications of varying scope tailored to SGM youth are promising sources of mental health support. Unfortunately, the absence of a no-treatment control group prevents us from inferring whether both interventions are efficacious relative to the absence of any program. Although comparing the full *imi* intervention with a resource-only subset of the *imi* website (ie, *asterix*) makes it harder to detect differences between the arms, we felt that withholding referrals to care for SGM youth would be unethical. We also selected this control as we wished to examine whether the full, interactive *imi* intervention provides benefits to SGM youth that go above and beyond what might be gained by being given access to a unified, vetted list of freely available existing resources. Previous studies suggest that SGM youth are increasingly and disproportionately searching for support on the web; however, this opens them to increased experiences of encountering homophobic and transphobic content and other hate speech [23]. It may be the case that gathering vetted resources in a single safe site designed specifically for them (ie, the resources section of the *imi* application) supports the well-being of SGM youth in and of itself. Although our design lacked a no-treatment control and, therefore, was unable to test this hypothesis, it does highlight the additive benefit of interactive identity affirmation and coping content contained in the full *imi* intervention; that is, although participants who were assigned both versions of the web application (the resource-only version of *imi* called *asterix* and the full interactive *imi* intervention) experienced decreased feelings of anxiety and depression after 4 weeks, only those given the full intervention experienced the additional benefit of greater stress appraisal and cognitive and behavioral coping skill improvements.

Future Directions

The lack of observed differences between the arms across other domains (eg, identity affirmation and connectedness to the LGBTQ+ community, internalization of minority stress, and sense of belonging) may be related to several additional considerations. First, it is possible that the 4-week study period may be too short to allow for meaningful changes to emerge in certain types of outcomes. For example, it may be that identity-related shifts require more time to manifest. Second, consistent with participants' qualitative feedback, it is possible that SGM youth require a greater amount of content and activities in a web application to, for example, feel more connected to the broader LGBTQ+ community. Third, some participants noted in their qualitative recommendations that the *imi* application may be most useful for SGM youth who are earlier in the exploration and affirmation of their identities. However, it remains unclear whether the intervention's effects may be greater for SGM youth earlier in their coming out process or those questioning their identities, a question which this study was not statistically powered to address. Future research is needed to examine whether the intervention's effect on the outcomes requires greater content to be developed within the web application, whether the intervention effects are moderated by how comfortable users are with their identities, or whether a longer follow-up period may allow for changes in distal outcomes to be observed.

Participants' engagement with the web application also offers insights into its acceptability and preliminary efficacy. Most participants across both arms created an account on their assigned web application, with participants reporting a comparable number of log-ins over the 4-week follow-up period. As expected, participants in the *imi* application arm viewed more unique pages than participants in the control arm. In and of itself, this finding does not indicate greater engagement with the *imi* application, as the control web application contained only a small subset of pages from the *imi* application. Consistent with our hypothesis, participants in the *imi* application arm also spent more time logged into the intervention than their peers in the control arm. However, *imi* participants were less likely than their control counterparts to click on external links. When taken together, these findings may be explained by two alternate (but complementary) possibilities: *imi* participants may have been more likely to spend time on the web application and less likely to require clicking on external links as the created content and activities were engaging and sufficiently helpful. Alternatively, given the limited content available to SGM youth in the control arm, participants assigned to *asterix* may have spent their time reviewing the content offered through the links to curated LGBTQ+-affirming resources, a possibility we could not assess in this study as we could not track participants' behavior on external websites. Future research examining participants' experiences within and outside of the web application may be warranted.

As an exploratory analysis, we examined whether SGM youth's engagement with the *imi* application could affect the strength of the intervention's effect. Consistent with our hypothesis, and in alignment with prior research examining the role of engagement on a digital intervention's efficacy [41,44,45], higher engagement with the *imi* application (eg, ≥10 minutes spent on the tool; >10 unique pages viewed) was linked to greater improvement on our primary outcome variables. This finding supports the interpretation that engaging with the *imi* application's content improves coping abilities. On the basis of our engagement data, it is promising to see that more than half of the participants assigned to the full *imi* intervention achieved these thresholds. However, it is worthwhile noting that SGM youth were incentivized for creating an account within the tool and sent reminders by the study team to log back in over time. It remains crucial to understand whether these thresholds and their associated benefits persist outside of the clinical trial setting. Efforts to explore engagement in a naturalistic study may be worthwhile to examine and affirm external validity.

Strengths and Limitations

Several limitations are worth noting. First, although the intervention effects moved in favor of the *imi* application arm in primary and secondary outcomes, our ability to detect these effects with statistical precision was limited by our small sample size and short follow-up period. Future scaled-up versions of the *imi* intervention with larger sample sizes, a longer duration, and a greater number of follow-up periods may be warranted to examine efficacy and effectiveness with greater precision. Second, some of the indicators used to measure our outcomes (eg, authenticity and LGBTQ+ community connectedness) were originally developed with adult populations. Given the unique needs of SGM youth, it is possible that the measures used in our study were not optimal for use with SGM youth populations. Future research examining the psychometric appropriateness of these measures with SGM youth populations may be warranted. Third, our ability to recruit youth from the lowest end of the age spectrum was somewhat limited. Only 11.1% (30/270) of the sample was aged 13 to 14 years. This is a common limitation in studies of SGM youth [46]. However, given the qualitative feedback that youth earlier in the journey of identity exploration may benefit more from the tool, future implementation research should explore pathways for making the tool more accessible to both younger youth and youth who may be in the earlier stages of identity exploration or questioning their identities. Similarly, although the diversity of race and ethnicities, as well as sexual and gender identities represented in our sample, is a strength, this initial pilot study was not powered to assess whether the intervention has differential efficacy based on these demographic characteristics. Finally, we weighed the advantages and disadvantages of enforcing a type I error (ie, false positive) correction relative to a type II error (ie, false negative). Given the exploratory nature of our trial and the small sample size of this pilot study, we did not include a family-wise error rate correction, which is justified in early exploration [47]. Future research examining the effects of both the *imi* and *asterix* applications with larger and population-representative samples may be warranted.

Our study also has several strengths worth emphasizing. First, it is one of the few clinical trials examining how to design and deliver a coping intervention that may reduce the negative effects of minority stress on SGM youth. Second, compared with face-to-face programs for SGM youth and synchronous digital interventions, our findings underscore the feasibility and acceptability of an asynchronous digital intervention that overcomes access and engagement barriers by being freely accessible on demand, scalable, confidential, and not requiring a significant time commitment. Our study demonstrated that a brief (as little as 10 minutes), self-guided intervention may have significant benefits for coping with stress. Finally, compared with prior trials, our study increases the likelihood that the findings are generalizable, given our commitment to recruiting and retaining a diverse group of SGM youth across races, ethnicities, sexual orientations and gender identities, geographies, and socioeconomic backgrounds.

Conclusions

This study demonstrated that a brief web-based intervention can provide self-guided, asynchronous, and confidential support that improves the ability of SGM youth to cope with minority stress. As a tool, the *imi* application may provide public health utility and value by expanding the reach and scalability of mental health programs, particularly for SGM youth who may be unable to participate in time-intensive, synchronous interactions. It may also serve as an ancillary tool for community-based agencies seeking to engage their SGM youth clients via the internet. Further research is needed to examine the long-term effects of the *imi* application and its potential for scalability and population health impact.

Acknowledgments

The authors are grateful to all their partners who participated in the co-design of the *imi* application and the study reported herein. In particular, the authors would like to thank Chris McCarthy, Denise Ho, Danielle Ramo, Janise Chan, Grace Greene, Cassandra Marzke, Primo Lagaso Goldberg, Griffin Somaratne, Starr Jiang, Rayan Madjidi, Adriana Allman, Sam Dinga, Will Coleman, Jett Allen, Amanda Keller, Lauren Jacobs, Katie Suppes, Meryl Franklin, Sam Aguilar, Kevin McCloskey, Forrest Marie Linsell, David Biele, Tess Allen, Audrey Ford, Brooks Banker, Deborah Levine, and Justin Tindall, who were instrumental in the design and development of the *imi* application. The authors would also like to thank their scientific advisors Annesa Flentje, Shelley Craig, Edward Alessi, Ashley Austin, Alison Cerezo, and Amaranta Ramirez for their thoughtful input and feedback.

Conflicts of Interest

imi is a progressive web application codeveloped by Hopelab, CenterLink, and It Gets Better Project. Although all 3 organizations were involved in the development of the product, CenterLink will ultimately be responsible for the operation and distribution of the *imi* application. As a free digital tool created and distributed by nonprofit organizations, none of the organizations involved stand to profit financially from the product. The research reported here, as well as the development of the *imi* application, were supported by the nonprofit Hopelab Foundation. Hopelab develops behavioral interventions to improve the health and well-being of young people. The design, conduct, analysis, and reporting of this study represent a scientific collaboration between Hopelab and the Program on Sexuality, Technology, and Action Research at the University of Pennsylvania School of Nursing. EBS, FD, AT, JL, LR, and JH are employed by Hopelab Foundation. The study sponsor was involved in the study design; collection, analysis, and interpretation of data; writing of the article; and decision to submit it for publication.

Multimedia Appendix 1
Informed consent form.
[DOCX File , 26 KB - jmir_v24i8e39094_app1.docx]

Multimedia Appendix 2
The *imi* application informational video.
[MP4 File (MP4 Video), 127192 KB - jmir_v24i8e39094_app2.mp4]

Multimedia Appendix 3
Screenshots from the *imi* (treatment) and *asterix* (control) web applications.
[PNG File , 1679 KB - jmir_v24i8e39094_app3.png]

Multimedia Appendix 4
Assessment of attrition and differential attrition by study arm.
[DOCX File , 32 KB - jmir_v24i8e39094_app4.docx]

Multimedia Appendix 5
CONSORT eHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 278 KB - jmir_v24i8e39094_app5.pdf]

References

1. National Academies of Sciences, Engineering, and Medicine. Understanding the Well-Being of LGBTQI+ Populations. Washington, DC, USA: National Academies Press (US); 2020.
2. Jones SE, Ethier KA, Hertz M, DeGue S, Le VD, Thornton J, et al. Mental health, suicidality, and connectedness among high school students during the COVID-19 pandemic - adolescent behaviors and experiences survey, United States, January-June 2021. MMWR Suppl 2022 Apr 01;71(3):16-21 [FREE Full text] [doi: 10.15585/mmwr.su7103a3] [Medline: 35358165]

3. National Survey on LGBTQ Youth Mental Health 2021. The Trevor Project. 2021. URL: https://www.thetrevorproject.org/survey-2021/?section=SuicideMentalHealth [accessed 2022-07-18]
4. National Academies of Sciences, Engineering, and Medicine. Reducing Inequalities Between Lesbian, Gay, Bisexual, Transgender, and Queer Adolescents and Cisgender, Heterosexual Adolescents: Proceedings of a Workshop. Washington, DC, USA: National Academies Press (US); 2022 Presented at: NASEM '22; August 25–27, 2021; Virtual. [doi: 10.17226/26383]
5. Meyer IH. Minority stress and mental health in gay men. J Health Soc Behav 1995 Mar;36(1):38-56. [Medline: 7738327]
6. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697 [FREE Full text] [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
7. Pachankis JE. The psychological implications of concealing a stigma: a cognitive-affective-behavioral model. Psychol Bull 2007 Mar;133(2):328-345. [doi: 10.1037/0033-2909.133.2.328] [Medline: 17338603]
8. Hatzenbuehler ML, Phelan JC, Link BG. Stigma as a fundamental cause of population health inequalities. Am J Public Health 2013 May;103(5):813-821 [FREE Full text] [doi: 10.2105/AJPH.2012.301069] [Medline: 23488505]
9. Ayala G, Bingham T, Kim J, Wheeler DP, Millett GA. Modeling the impact of social discrimination and financial hardship on the sexual risk of HIV among Latino and Black men who have sex with men. Am J Public Health 2012 May;102 Suppl 2:S242-S249. [doi: 10.2105/AJPH.2011.300641] [Medline: 22401516]
10. Corliss HL, Rosario M, Birkett MA, Newcomb ME, Buchting FO, Matthews AK. Sexual orientation disparities in adolescent cigarette smoking: intersections with race/ethnicity, gender, and age. Am J Public Health 2014 Jun;104(6):1137-1147. [doi: 10.2105/AJPH.2013.301819] [Medline: 24825218]
11. Shangani S, Gamarel KE, Ogunbajo A, Cai J, Operario D. Intersectional minority stress disparities among sexual minority adults in the USA: the role of race/ethnicity and socioeconomic status. Cult Health Sex 2020 Apr;22(4):398-412 [FREE Full text] [doi: 10.1080/13691058.2019.1604994] [Medline: 31144598]
12. Dyar C, Taggart TC, Rodriguez-Seijas C, Thompson Jr RG, Elliott JC, Hasin DS, et al. Physical health disparities across dimensions of sexual orientation, race/ethnicity, and sex: evidence for increased risk among bisexual adults. Arch Sex Behav 2019 Jan;48(1):225-242 [FREE Full text] [doi: 10.1007/s10508-018-1169-8] [Medline: 29633061]
13. Wenzel L, Glanz K, Lerman C. Stress, coping, and health behavior. In: Glanz A, Glanz K, Rimer BK, Lewis FM, editors. Health Behavior and Health Education: Theory, Research, and Practice. 3rd edition. San Francisco, CA, USA: Jossey-Bass; 2002:210-239.
14. Jamieson JP, Crum AJ, Goyer JP, Marotta ME, Akinola M. Optimizing stress responses with reappraisal and mindset interventions: an integrated model. Anxiety Stress Coping 2018 May;31(3):245-261. [doi: 10.1080/10615806.2018.1442615] [Medline: 29471669]
15. Rowley AA, Roesch SC, Jurica BJ, Vaughn AA. Developing and validating a stress appraisal measure for minority adolescents. J Adolesc 2005 Aug;28(4):547-557. [doi: 10.1016/j.adolescence.2004.10.010] [Medline: 16022889]
16. Yeager DS, Lee HY, Jamieson JP. How to improve adolescent stress responses: insights from integrating implicit theories of personality and biopsychosocial models. Psychol Sci 2016 Aug;27(8):1078-1091 [FREE Full text] [doi: 10.1177/0956797616649604] [Medline: 27324267]
17. Craig SL, Austin A, McInroy LB. School-based groups to support multiethnic sexual minority youth resiliency: preliminary effectiveness. Child Adolesc Soc Work J 2013 Aug 15;31(1):87-106. [doi: 10.1007/s10560-013-0311-7]
18. Craig SL, Leung VW, Pascoe R, Pang N, Iacono G, Austin A, et al. AFFIRM online: utilising an affirmative cognitive-behavioural digital intervention to improve mental health, access, and engagement among LGBTQA+ youth and young adults. Int J Environ Res Public Health 2021 Feb 05;18(4):1541 [FREE Full text] [doi: 10.3390/ijerph18041541] [Medline: 33562876]
19. Pachankis JE, Hatzenbuehler ML, Rendina HJ, Safren SA, Parsons JT. LGB-affirmative cognitive-behavioral therapy for young adult gay and bisexual men: a randomized controlled trial of a transdiagnostic minority stress approach. J Consult Clin Psychol 2015 Oct;83(5):875-889 [FREE Full text] [doi: 10.1037/ccp0000037] [Medline: 26147563]
20. Craig SL, Eaton AD, Leung VW, Iacono G, Pang N, Dillon F, et al. Efficacy of affirmative cognitive behavioural group therapy for sexual and gender minority adolescents and young adults in community settings in Ontario, Canada. BMC Psychol 2021 Jun 07;9(1):94 [FREE Full text] [doi: 10.1186/s40359-021-00595-6] [Medline: 34099063]
21. Silliman Cohen RI, Bosk EA. Vulnerable youth and the COVID-19 pandemic. Pediatrics 2020 Jul;146(1):e20201306. [doi: 10.1542/peds.2020-1306] [Medline: 32345686]
22. Fish JN, McInroy LB, Paceley MS, Williams ND, Henderson S, Levine DS, et al. "I'm kinda stuck at home with unsupportive parents right now": LGBTQ youths' experiences with COVID-19 and the importance of online support. J Adolesc Health 2020 Sep;67(3):450-452 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.06.002] [Medline: 32591304]
23. Rideout V, Fox S, Peebles A, Robb MB. Coping with COVID-19: How young people use digital media to manage their mental health. Common Sense and Hopelab. San Francisco, CA, USA: Common Sense and Hopelab; 2021. URL: https://www.commonsensemedia.org/sites/default/files/research/report/2021-coping-with-covid19-full-report.pdf [accessed 2022-07-18]

24. Lucassen MF, Merry SN, Hatcher S, Frampton CM. Rainbow SPARX: a novel approach to addressing depression in sexual minority youth. Cogn Behav Pract 2015 May;22(2):203-216. [doi: 10.1016/j.cbpra.2013.12.008]
25. Schwinn TM, Thom B, Schinke SP, Hopkins J. Preventing drug use among sexual-minority youths: findings from a tailored, Web-based intervention. J Adolesc Health 2015 May;56(5):571-573. [doi: 10.1016/j.jadohealth.2014.12.015] [Medline: 25744209]
26. Egan JE, Corey SL, Henderson ER, Abebe KZ, Louth-Marquez W, Espelage D, et al. Feasibility of a Web-accessible game-based intervention aimed at improving help seeking and coping among sexual and gender minority youth: results from a randomized controlled trial. J Adolesc Health 2021 Oct;69(4):604-614. [doi: 10.1016/j.jadohealth.2021.03.027] [Medline: 34140199]
27. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 2;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
28. Teitcher JE, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
29. Bauermeister J, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data quality in Web-based HIV/AIDS research: handling invalid and suspicious data. Field methods 2012 Aug 01;24(3):272-291 [FREE Full text] [doi: 10.1177/1525822X12443097] [Medline: 23180978]
30. Guest JL, Adam E, Lucas IL, Chandler CJ, Filipowicz R, Luisi N, et al. Methods for authenticating participants in fully Web-based mobile app trials from the iReach project: cross-sectional study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e28232 [FREE Full text] [doi: 10.2196/28232] [Medline: 34463631]
31. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
32. Riggle ED, Mohr JJ, Rostosky SS, Fingerhut AW, Balsam KF. A multifactor lesbian, gay, and bisexual positive identity measure (LGB-PIM). Psychol Sex Orientat Gend Divers 2014 Dec;1(4):398-411. [doi: 10.1037/sgd0000057]
33. Ngamake ST, Walch SE, Raveepatarakul J. Validation of the Coping with Discrimination Scale in sexual minorities. J Homosex 2014;61(7):1003-1024. [doi: 10.1080/00918369.2014.870849] [Medline: 24325286]
34. Van Orden KA, Cukrowicz KC, Witte TK, Joiner TE. Thwarted belongingness and perceived burdensomeness: construct validity and psychometric properties of the Interpersonal Needs Questionnaire. Psychol Assess 2012 Mar;24(1):197-215 [FREE Full text] [doi: 10.1037/a0025358] [Medline: 21928908]
35. 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]
36. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
37. Allgaier AK, Pietsch K, Frühe B, Sigl-Glöckner J, Schulte-Körne G. Screening for depression in adolescents: validity of the patient health questionnaire in pediatric care. Depress Anxiety 2012 Oct;29(10):906-913. [doi: 10.1002/da.21971] [Medline: 22753313]
38. Mossman SA, Luft MJ, Schroeder HK, Varney ST, Fleck DE, Barzman DH, et al. The Generalized Anxiety Disorder 7-item scale in adolescents with generalized anxiety disorder: signal detection and validation. Ann Clin Psychiatry 2017 Nov;29(4):227-34A [FREE Full text] [Medline: 29069107]
39. Moyer DN, Connelly KJ, Holley AL. Using the PHQ-9 and GAD-7 to screen for acute distress in transgender youth: findings from a pediatric endocrinology clinic. J Pediatr Endocrinol Metab 2019 Jan 28;32(1):71-74. [doi: 10.1515/jpem-2018-0408] [Medline: 30530884]
40. Lyons A, Rozbroj T, Pitts M, Mitchell A, Christensen H. Improving e-therapy for mood disorders among lesbians and gay men: a practical toolkit for developing tailored web and mobile phone-based depression and anxiety interventions. La Trobe University. Melbourne, Australia: The Australian Research Centre in Sex, Health and Society; 2015. URL: https://www.latrobe.edu.au/__data/assets/pdf_file/0005/632381/Improving-etherapy-for-mood-disorders-among-lesbians-and-gay-men,-A-practical-toolkit.pdf [accessed 2022-07-18]
41. Hightow-Weidman LB, Bauermeister JA. Engagement in mHealth behavioral interventions for HIV prevention and care: making sense of the metrics. Mhealth 2020 Jan 5;6:7 [FREE Full text] [doi: 10.21037/mhealth.2019.10.01] [Medline: 32190618]
42. Vindrola-Padros C, Johnson GA. Rapid techniques in qualitative research: a critical review of the literature. Qual Health Res 2020 Aug;30(10):1596-1604. [doi: 10.1177/1049732320921835] [Medline: 32667277]
43. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York, NY, USA: Springer; 1984.
44. 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 Mar;59:299-307 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]

45. Bauermeister JA, Muessig KE, Flores DD, LeGrand S, Choi S, Dong W, et al. Stigma diminishes the protective effect of social support on psychological distress among young black men who have sex with men. AIDS Educ Prev 2018 Oct;30(5):406-418 [FREE Full text] [doi: 10.1521/aeap.2018.30.5.406] [Medline: 30332312]
46. Mustanski B. Future directions in research on sexual minority adolescent mental, behavioral, and sexual health. J Clin Child Adolesc Psychol 2015;44(1):204-219 [FREE Full text] [doi: 10.1080/15374416.2014.982756] [Medline: 25575125]
47. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. J Res Educ Eff 2012 Apr;5(2):189-211. [doi: 10.1080/19345747.2011.618213]

Abbreviations

COPE: Coping Orientation to Problems Experienced
LGB: lesbian, gay, and bisexual
LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other sexual and gender minority
NPS: net promoter score
SGM: sexual and gender minority

Edited by G Eysenbach, T Leung; submitted 30.04.22; peer-reviewed by T Piehler; comments to author 07.06.22; revised version received 25.06.22; accepted 26.06.22; published 01.08.22.

Please cite as:
Bauermeister J, Choi SK, Bruehlman-Senecal E, Golinkoff J, Taboada A, Lavra J, Ramazzini L, Dillon F, Haritatos J
An Identity-Affirming Web Application to Help Sexual and Gender Minority Youth Cope With Minority Stress: Pilot Randomized Controlled Trial
J Med Internet Res 2022;24(8):e39094
URL: https://www.jmir.org/2022/8/e39094
doi:10.2196/39094
PMID:35916700

Original Paper

Effect of a Reminder System on Pre-exposure Prophylaxis Adherence in Men Who Have Sex With Men: Prospective Cohort Study Based on WeChat Intervention

Bing Lin[1,2], PhD; Jiaxiu Liu[3], MA; Wei He[1,2], MA; Haiying Pan[1,2], MA; Yingjie Ma[1,2], MA; Xiaoni Zhong[1,2], MA

[1]School of Public Health and Management, Chongqing Medical University, Chongqing, China
[2]Research Center for Medicine and Social Development, Chongqing, China
[3]School of Medical Informatics, Chongqing Medical University, Chongqing, China

Corresponding Author:
Xiaoni Zhong, MA
School of Public Health and Management
Chongqing Medical University
No.1 Medical College Road
Yuzhong District
Chongqing, 400042
China
Phone: 86 13308368059
Email: zhongxiaoni@cqmu.edu.cn

Abstract

Background: The efficacy of pre-exposure prophylaxis (PrEP) is highly dependent on adherence, and one of the main reasons for poor adherence is forgetfulness. Therefore, it is important to explore how to remind users to take their medicine on time.

Objective: This study aims to explore the effect of a reminder system on PrEP adherence in men who have sex with men (MSM) to improve adherence. The main function of the reminder system based on the WeChat social media app is to send daily messages to PrEP users reminding them to take their medicine.

Methods: An open-label, multicenter, prospective cohort study of PrEP in HIV-negative MSM was conducted from November 2019 to June 2021. Study participants who met the criteria were randomly divided into 2 groups: no-reminder group and reminder group. Both groups received daily oral PrEP with follow-up every 3 months. Adherence was measured on the basis of self-report and was defined as the percentage of medications taken on time. Participants in the reminder group scanned a WeChat QR code and received a reminder message every day. Participants in the no-reminder group took daily oral medicines without reminders. The longitudinal trajectories of adherence for both groups were displayed to compare the variability in adherence at each time point. The association between the changes in adherence (no change, improvement, decline) at each time point and the use of the reminder system was analyzed by multinomial logistic regression models to further explore the effectiveness of the system.

Results: A total of 716 MSM were included in the analysis, that is, 372 MSM in the no-reminder group and 344 MSM in the reminder group. Adherence in the no-reminder group fluctuated between 0.75 and 0.80 and that in the reminder group gradually increased over time from 0.76 to 0.88. Adherence at each time point was not statistically different between the 2 groups. Further analysis showed that an improvement in adherence in the early stage was associated with the use of the reminder system (odds ratio [OR] 1.65, 95% CI 1.01-2.70; *P*=.04). An improvement in average adherence compared to initial adherence was positively associated with the use of the reminder system (OR 1.82, 95% CI 1.10-3.01; *P*=.02).

Conclusions: The effect of the reminder system on PrEP adherence in MSM was more significant in the early stage, which is related to the increased motivation of users and the development of medicine-taking habits. The reminder system is potentially effective for early-stage medicine management, encouraging users to develop healthy medicine-taking habits and to increase their adherence.

Trial Registration: Chinese Clinical Trial ChiCTR190026414; http://www.chictr.org.cn/showproj.aspx?proj=35077

(J Med Internet Res 2022;24(8):e37936) doi:10.2196/37936

KEYWORDS

pre-exposure prophylaxis (PrEP); adherence; reminder system; men who have sex with men (MSM); WeChat; oral PrEP; HIV prevention; MSM; reminder; message

Introduction

Globally, men who have sex with men (MSM) bear a disproportionate burden of HIV, and they are a high-risk group for infection [1]. In China, more than a quarter of new HIV diagnoses can be attributed to MSM [2]. Despite various interventions, the risk of HIV infections among MSM in China is increasing, with the proportion of new HIV infections rising from 2.5% to 25.5% [3,4]. Given the existing circumstances, several studies have been exploring ways to prevent HIV infection in MSM populations, including biomedical interventions such as oral pre-exposure prophylaxis (PrEP) [5,6]. PrEP is a biological HIV prevention intervention that focuses on reducing the risk of HIV infection through daily (or event-based) oral antiretroviral medication [7]. In fact, several clinical trials and cohort studies have demonstrated the safety and efficacy of oral PrEP in reducing the risk of HIV infection in MSM [8]. However, the efficacy of PrEP for HIV prevention is highly dependent on adherence [9,10].

PrEP advocates and researchers agree that one of the important issues that needs to be addressed before PrEP can be scaled up is the concern about nonadherence [11,12]. As a result, researchers have focused their attention on PrEP adherence and the factors that influence it. Prior studies have shown that the potential barriers to daily oral PrEP include living arrangement, side effects, stigma, and forgetfulness [13,14]. In addition, another study found that forgetting to take medicines is the most common objective reason for nonadherence, accounting for 70.21% [15]. Taken together, forgetfulness is one of the main reasons for poor adherence, and reminding users to take medicines is an important way to improve adherence. Exploring how to obtain and maintain high adherence in MSM populations is crucial for promoting the effective implementation of PrEP intervention strategies and reducing new HIV infections.

Different methods can be considered to remind MSM to take their medicines on time. Many attempts have been made in this direction by researchers. Mobile health (mHealth) technologies are effective and cost-effective strategies to improve individual and public health [16,17]. Among them, SMS text messaging is often used to help remind users to take their medicines, which can improve adherence, owing to its ubiquity and ease of use in mobile devices. The feasibility and acceptability of SMS text messaging as a potential tool for primary HIV prevention has been demonstrated [18,19]. Meanwhile, other studies have also developed a novel mobile app to support PrEP adherence through artificial intelligence and an electronic sexual diary, which have received positive feedback from users, thereby providing a further basis for future effectiveness studies [20]. These studies of improving adherence through a “reminder function” have been tested in Africa, Thailand, Peru, and the United States, but implementation in China is uncommon [21,22].

Like Facebook and Twitter, WeChat is a popular social media app in China. According to the data analysis, WeChat is the most popular social media platform with over 1 billion registered users [23]. Approximately 93% of the residents in the major Chinese cities log on to WeChat every day [24]. The high ownership of mobile phones and the widespread popularity of the WeChat app suggest that this is a promising platform for providing low-cost interventions. Moreover, WeChat-based interventions have shown feasibility and acceptability in HIV prevention and control in China [25,26], which includes MSM [23]. However, studies using the WeChat app to remind MSM to take their medicine daily during PrEP use are limited in China.

Therefore, we conducted a prospective cohort study of PrEP adherence in the MSM population in Western China. Our reminder system is based on the WeChat app for mobile phones that scans individual QR codes and connects to the backend management system to send daily reminder messages. We compared the variability of adherence between the no-reminder group and the reminder group at each time point by plotting the longitudinal trajectory of adherence. At the same time, the relationship between changes in adherence (no change, improvement, decline) and the use of the reminder system was investigated to confirm the influence of the reminder system on adherence and to provide a theoretical foundation for the improvement of the reminder system in the future so as to improve adherence and increase the effectiveness of PrEP.

Methods

Ethics Approval

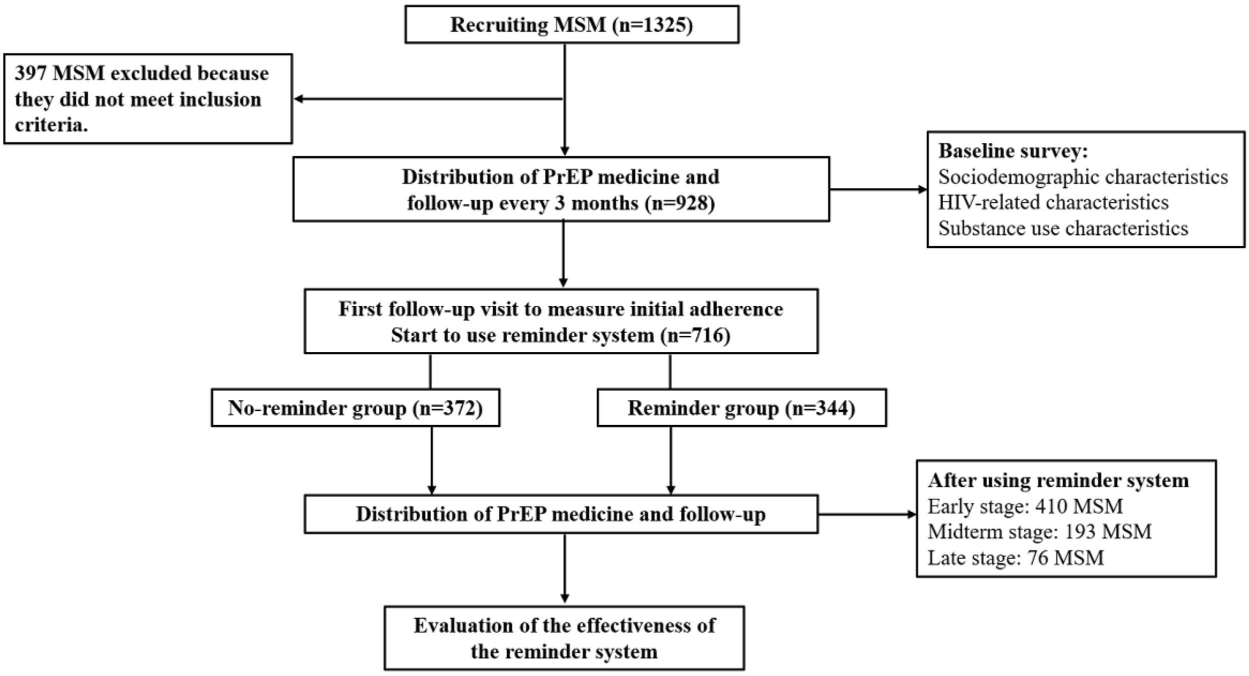
All procedures of this study were in accordance with the ethics approval granted by the ethics committee of Chongqing Medical University (2019001). The ethics committee of Chongqing Medical University has reviewed the proposed use of human participants in the abovementioned project.

Data Collection

This study was a PrEP open-label, multicenter, prospective cohort study conducted from November 2019 to June 2021 in 3 regions of Western China: Chongqing, Sichuan, and Xinjiang (Chinese clinical trial registration ChiCTR190026414). MSM who met the criteria were recruited through collaboration with local nongovernment organizations and peer recommendations. Inclusion criteria for the study population were (1) physiological male (assigned male sex at birth), (2) older than 18 years, (3) had engaged in sex with male partners in the past 6 months, (4) negative HIV antigen-antibody test, (5) willing to use medicines under guidance and subject to follow-up arrangements, and (6) signed informed consent form.

The MSM population was screened for inclusion criteria and then divided into 2 groups: no-reminder group and reminder group. After completing the baseline survey, participants were followed up every 3 months and given their medicines by our study researchers. They received daily oral dose of Lamivudine and Tenofovir Disoproxil Fumarate tablets (300 mg/tablet). The first 3 months after joining the cohort were considered as an observation period with no use of the reminder system. The initial adherence of the study participants was measured and the reminder system was started at the first follow-up visit. Participants in the reminder group scanned a WeChat QR code (unique identity QR code, scanned only once) to receive daily medicine reminder messages, while the no-reminder group took daily oral medicines without reminders. We designated the intervals corresponding to each follow-up time point as early, midterm, and late stages after the initial follow-up visit (beginning to employ the reminder system). A total of 716 MSM were included in the analysis: 372 MSM in the no-reminder group and 344 MSM in the reminder group. Figure 1 presents the flow chart of the recruitment, survey, and follow-up of the study participants.

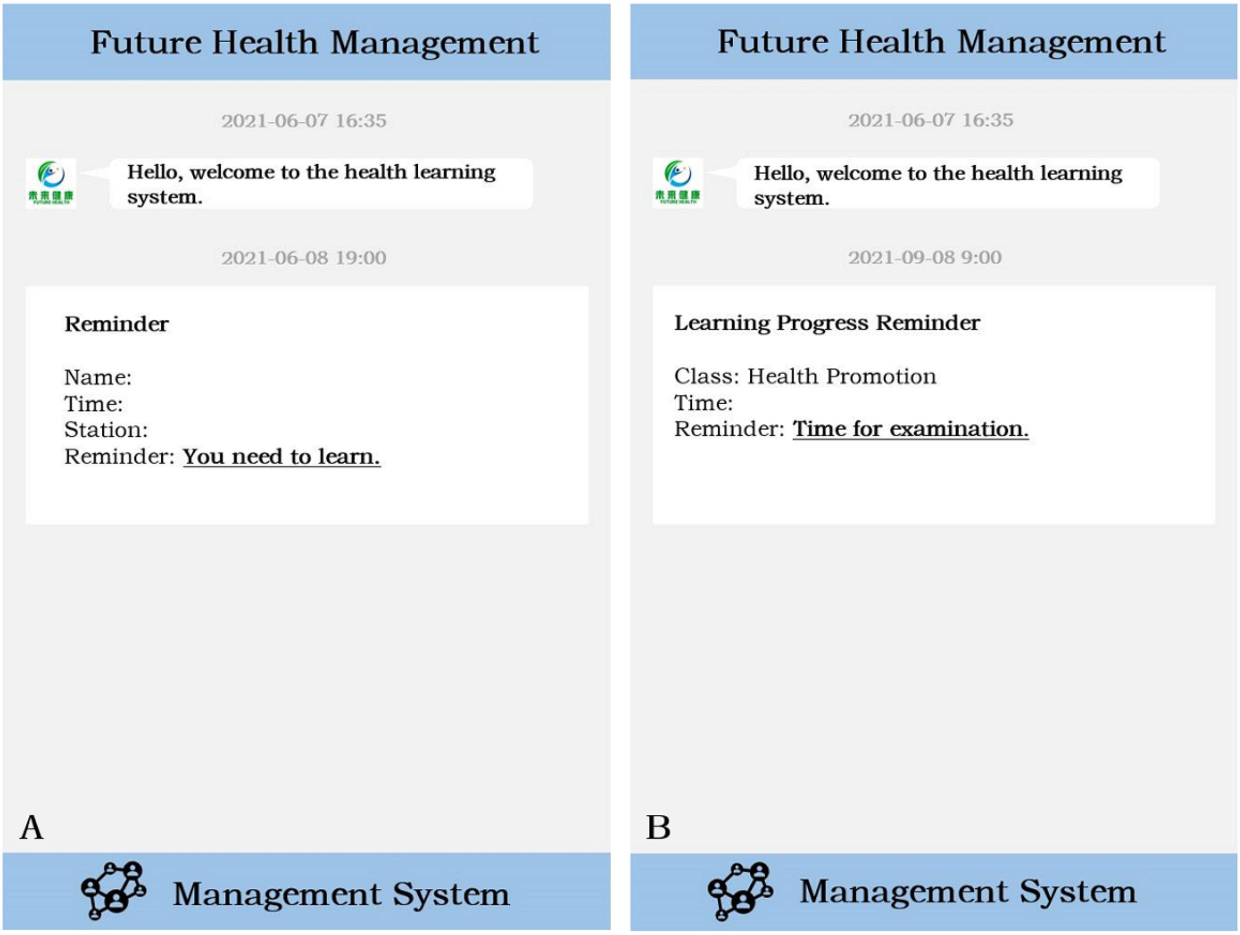
Figure 1. Flowchart of the recruitment, survey, and follow-up of the study participants. MSM: men who have sex with men. PrEP: pre-exposure prophylaxis.



Reminder System of PrEP

Our study developed a PrEP reminder system in conjunction with Chongqing Future Health Management Company Limited. This system relies on the internet and intelligent medicine information tracking management system using cloud computing, big data, intelligent hardware, and other new generation information technology products to provide effective medicine management for the study participants through the process of medicine-taking plan, medicine reminder, and health knowledge promotion to achieve the expected prevention effect. Our researchers typed in the basic information of the reminder group into the background management system and generated an independent personal QR code. Participants were bound to the backend system after scanning the QR code using their WeChat apps. Daily reminders were sent from the next day, and users could set the time according to their request. In order to protect the privacy of the users, the content of our messages was relatively obscure. For example, when users received a “You need to learn” message from Future Health Management, it is a reminder that they should take the medicine that day. When users received a “Time for examination” message from Future Health Management, it is a reminder that they should visit the study center for the 3-month follow-up (Figure 2).

Figure 2. A: reminds users that they should take their medicine that day and B: reminds users that they should go for a follow-up visit.



2021-06-07 16:35

Hello, welcome to the health learning system.

2021-06-08 19:00

Reminder

Name:
Time:
Station:
Reminder: You need to learn.

2021-06-07 16:35

Hello, welcome to the health learning system.

2021-09-08 9:00

Learning Progress Reminder

Class: Health Promotion
Time:
Reminder: Time for examination.

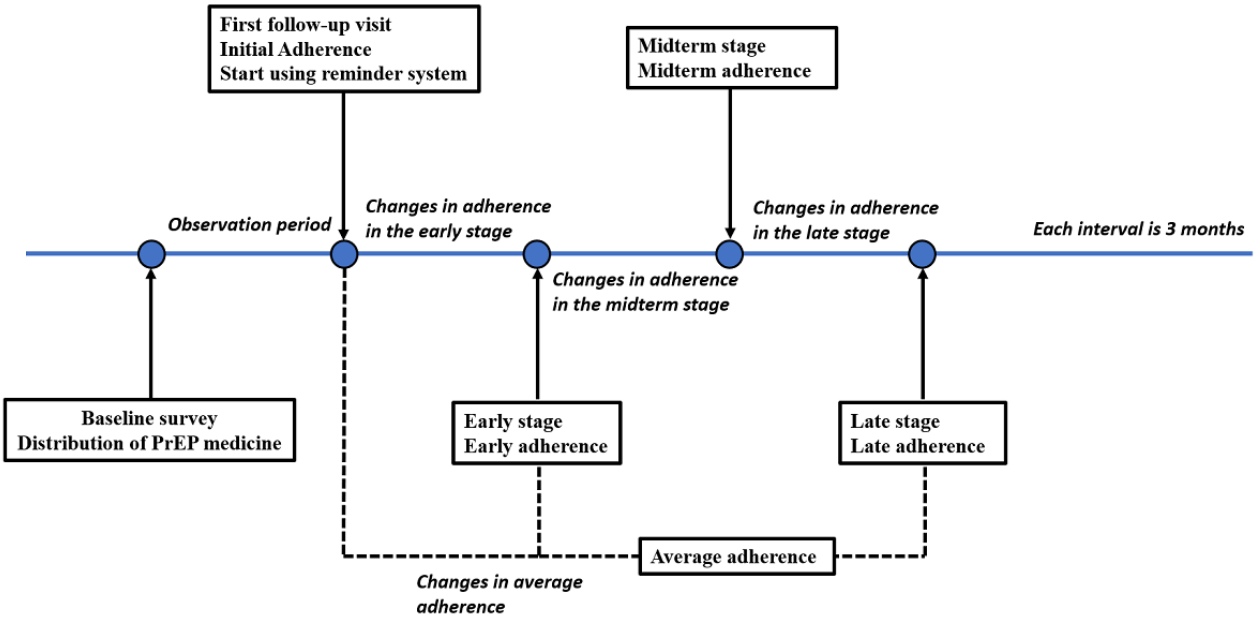
Baseline Variables

The baseline survey included demographic characteristics, HIV-related characteristics, and substance use characteristics. Among them, the demographic characteristics include age, household register location, education attainment, employment status, marital status, and monthly personal income. HIV-related characteristics mainly included HIV testing and counseling, number of sexual partners, condom use, and HIV risk perception. Participants were asked, “In the last month, how many male sexual partners (casual and regular) have you had? How many times did you have sex with a male sexual partner? How many of these times did you use a condom?” These questions were used to measure the number of male sexual partners and condom use during sex. We defined that condoms were used at each sexual intercourse and if the number of condom use was greater than or equal to the number of sexual intercourses. At the same time, participants were asked, “How likely do you think you are to get AIDS?” We used this question to measure the participants’ perceived HIV risk. Participants responded on a scale of 1-5, with a score of 1 indicating very small and 5 indicating very large. We defined a score ≥3 as the subgroup with higher perceived HIV risk. Substance use mainly included alcohol use. Participants were asked, “in the last month, how often did you drink alcohol?” Participants responded, “basically every day,” “at least 3 times a week,” “at least 1 time a week,” “less than 1 time a week,” or “never drink alcohol.” We divided the alcohol use into 2 categories: one for never drinking and one for ever drinking.

Adherence

At each follow-up visit, adherence was measured by self-report. Study participants were asked, “in the last two weeks, have you missed any doses? how many days did you miss?” Participants answered “yes” or “no” and the number of days missed (0-14 days). Adherence was equal to the percentage of days adhered. All answers were checked by our researchers. If illogical answers occurred, quality control and corrections were made in the study site. After completing the baseline survey, participants entered an observation period, taking their medicines without reminders. The participants arrived for the first follow-up visit 3 months later, at which point adherence was referred to as “initial adherence.” In the meantime, study participants in the reminder group began using the reminder system. After this, adherence during the follow-up period was defined as early stage adherence, midterm adherence, and late stage adherence. Changes in adherence at each follow-up time point (no change, improvement, decline) were compared to the adherence in the previous period. For example, the changes in adherence in the early stage is defined as early stage adherence minus initial adherence. Average adherence is the mean of adherence in the early, midterm, and late periods, and changes in average adherence is defined as average adherence minus initial adherence. The definition of adherence in each period is shown in Figure 3.

Figure 3. Explanation of adherence (initial adherence, early stage adherence, midterm adherence, late stage adherence, average adherence) and changes in adherence by period. PrEP: pre-exposure prophylaxis.



Statistical Analyses

We compared the variability of the baseline demographic characteristics, HIV-related characteristics, and substance use characteristics between the no-reminder group and reminder group of the MSM population. Trajectories of adherence were plotted for the 2 groups. The nonparametric test (Kruskal-Wallis test) was used to compare the variability of adherence between the 2 groups at each follow-up time point. A multinomial logistic regression model was developed using the changes in adherence (no change, improvement, decline) in each period of follow-up as the dependent variable, and the relationship between the reminder system and the changes in adherence was further assessed by odds ratio (OR) and 95% CI. Variables with $P \leq .15$ were first screened by one-way analysis of variance, which was included in the multinomial logistic regression model. The efficacy of the reminder system was assessed by exploring the factors influencing the improvement and decline of adherence, using no change in adherence as the reference. $P<.05$ indicated a statistical difference. All statistical analyses were performed with SAS software (SAS Institute).

Results

A total of 716 MSM were included in the analysis, with 372 in the no-reminder group and 344 in the reminder group. The baseline demographic characteristics, HIV-related characteristics, and substance use characteristics of the 2 groups were not statistically different in one-way analysis of variance. The results of the descriptive analyses and chi-square tests for each variable are shown in Table 1.

Adherence trajectories were plotted for the no-reminder group and reminder group (Figure 4). The median of initial adherence was 1.00 (IQR 0.64-1.00) in the no-reminder group and 1.00 (IQR 0.47-1.00) in the reminder group, with no statistical difference in initial adherence between the 2 groups (P=.48). After using the reminder system, adherence in the no-reminder group fluctuated between 0.75 and 0.80; adherence in the reminder group gradually increased over time from 0.76 to 0.88. There was no statistical difference in adherence between the 2 groups at the early (P=.69), midterm (P=.96), and late (P=.37) stages of follow-up. After using the reminder system, the mean adherence for the 3 periods in the no-reminder group was 0.79 and the mean adherence in the reminder group was 0.78, which was not statistically different (P=.82) (Table 2).

Table 1. Differences in the baseline demographic characteristics, HIV-related characteristics, and substance use characteristics between the no-reminder group and reminder group in the men who have sex with men population (N=716).

Variables	No-reminder group (n=372), n (%)	Reminder group (n=344), n (%)	*P* value
Age (years)			.93
18-30	165 (44.47)	149 (43.31)	
31-45	158 (42.59)	151 (43.90)	
>45	48 (12.94)	44 (12.79)	
Household register location[a]			.52
Urban	254 (69.21)	245 (71.43)	
Rural	113(30.79)	98 (28.57)	
Education attainment[a]			.87
Junior high school and below	27 (7.28)	27 (7.85)	
High school	84 (22.64)	70 (20.35)	
College	94 (25.34)	93 (27.03)	
Undergraduate training or higher	166 (44.74)	154 (44.77)	
Employment status[a]			.54
Employed	307 (82.75)	275 (79.94)	
Internal student	34 (9.16)	40 (11.63)	
Jobless	30 (8.09)	29 (8.43)	
Marital status[a]			.45
Married	50 (13.51)	53 (15.50)	
Single	320 (86.49)	289 (84.50)	
Monthly personal income[a] **(¥1=US $0.15)**			.41
≤¥1000	25 (6.79)	22 (6.40)	
¥1000-¥3000	59 (16.03)	71 (20.64)	
¥3000-¥5000	125 (33.97)	104 (30.23)	
>¥5000	159 (43.21)	147 (42.73)	
HIV testing[a]			.78
Yes	341 (92.41)	317 (92.96)	
No	28 (7.59)	24 (7.04)	
HIV counseling[a]			.37
Yes	267 (71.97)	257 (74.93)	
No	104 (28.03)	86 (25.07)	
Number of male sexual partners in the last month[a]			.50
0	38 (10.33)	27 (7.99)	
1	204 (55.43)	198 (58.58)	
≥2	126 (34.24)	113 (33.43)	
Condoms were used at each sexual intercourse[a]			.65
Yes	301 (83.84)	279 (82.54)	
No	58 (16.16)	59 (17.46)	
Finding sex partners through the internet			.67
Yes	250 (67.20)	226 (65.70)	
No	122 (32.80)	118 (34.30)	

Variables	No-reminder group (n=372), n (%)	Reminder group (n=344), n (%)	*P* value
HIV risk perception[a]			.79
High	156 (42.05)	148 (43.02)	
Low	215 (57.95)	196 (56.98)	
If your male sexual partners know you are using pre-exposure prophylaxis, their attitude is			.08
Positive	174 (46.77)	156 (45.35)	
Neutral	151 (40.59)	160 (46.51)	
Negative	47 (12.63)	28 (8.14)	
Have you ever been diagnosed with a sexually transmitted disease by a doctor[a]			.18
Yes	18 (4.89)	25 (3.73)	
No	350 (95.11)	317 (92.69)	
Commercial sex[a]			.83
Yes	15 (4.05)	15 (4.39)	
No	355 (95.95)	327 (95.61)	
Alcohol use in the last month			.45
Yes	217 (58.33)	191 (55.52)	
No	155 (41.67)	153 (44.48)	

[a] Indicates missing data.

Figure 4. Trajectory plot of adherence at each time point in the men who have sex with men population in the no-reminder group and reminder group.

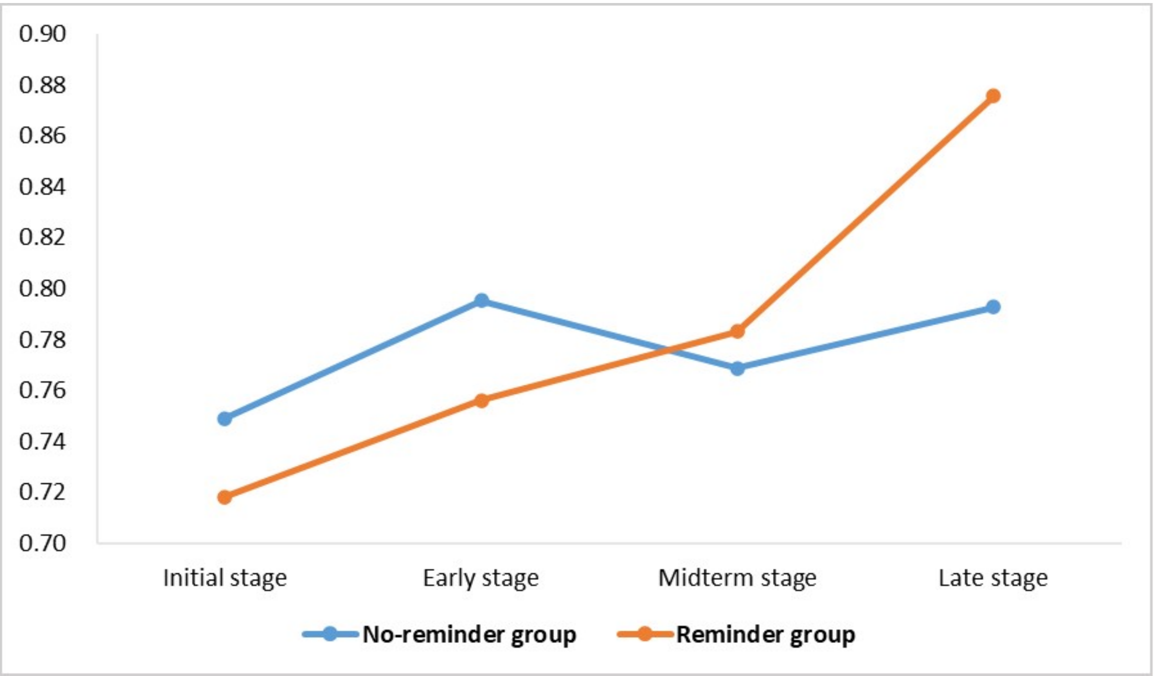


Table 2. Differences in adherence at each time point among the men who have sex with men population in the no-reminder group and reminder group.

	No-reminder group			Reminder group			*P* value[a]
	n	Mean (SD)	Median (IQR)	n	Mean (SD)	Median (IQR)	
Initial adherence (N=716)	372	0.75 (0.37)	1.00 (0.64-1.00)	344	0.72 (0.39)	1.00 (0.47-1.00)	.48
After using the reminder system							
Early stage adherence (n=410)	217	0.80 (0.33)	1.00 (0.79-1.00)	193	0.76 (0.37)	1.00 (0.64-1.00)	.69
Midterm adherence (n=193)	100	0.77 (0.37)	100 (0.68-1.00)	93	0.78 (0.35)	1.00 (0.79-1.00)	.96
Late stage adherence (n=76)	41	0.79 (0.31)	1.00 (0.64-1.00)	35	0.88 (0.23)	1.00 (0.79-1.00)	.37
Average adherence (n=410)	217	0.79 (0.31)	0.93 (0.72-1.00)	193	0.78 (0.32)	0.97 (0.64-1.00)	.82

[a] Kruskal-Wallis test was used.

To further explore the effect of the reminder system on adherence, the percentage of changes in adherence (no change, improvement, decline) at each time point in the MSM population is shown in Figure 5. A percentage bar graph was used to show the distribution of the changes in adherence. No change in adherence was used as a reference for the dependent variable, and the no-reminder group was used as a reference for the independent variable. After the one-way analysis of variance, variables with $P \leq .15$ were included in a multinomial logistic regression model for adjusting. The adjusted variables in the early stage included age, HIV counseling, and male sexual partners’ attitude. The adjusted variables in the midterm stage included age, household register location, education attainment, male sex partners, and their attitudes. The adjusted variables in the late stage included household register location and male sexual partners’ attitude. In the model for changes in the average adherence, the adjusted variables included age, HIV testing, HIV risk perception, alcohol use, and male sexual partners’ attitude. After adjusting for basic demographic characteristics, HIV-related characteristics, and substance use characteristics with no change in adherence as a reference, analysis based on multinomial logistic regression models was obtained (Table 3). An improvement in adherence in the early stage was positively associated with the use of the reminder system (OR 1.65, 95% CI 1.01-2.70; *P*=.04). The result showed that the reminder system was not able to prevent a decline in adherence in the early stage, and the difference was not statistically significant (*P*=.78). The changes in adherence in both the midterm and late stages were not statistically associated with the reminder system (*P*>.05). An improvement in the average adherence compared to the initial adherence was associated with the use of the reminder system (OR 1.82, 95% CI 1.10-3.01; *P*=.02); a decline in the average adherence was not associated with the reminder system (*P*=.67).

Figure 5. Distribution of changes in adherence.

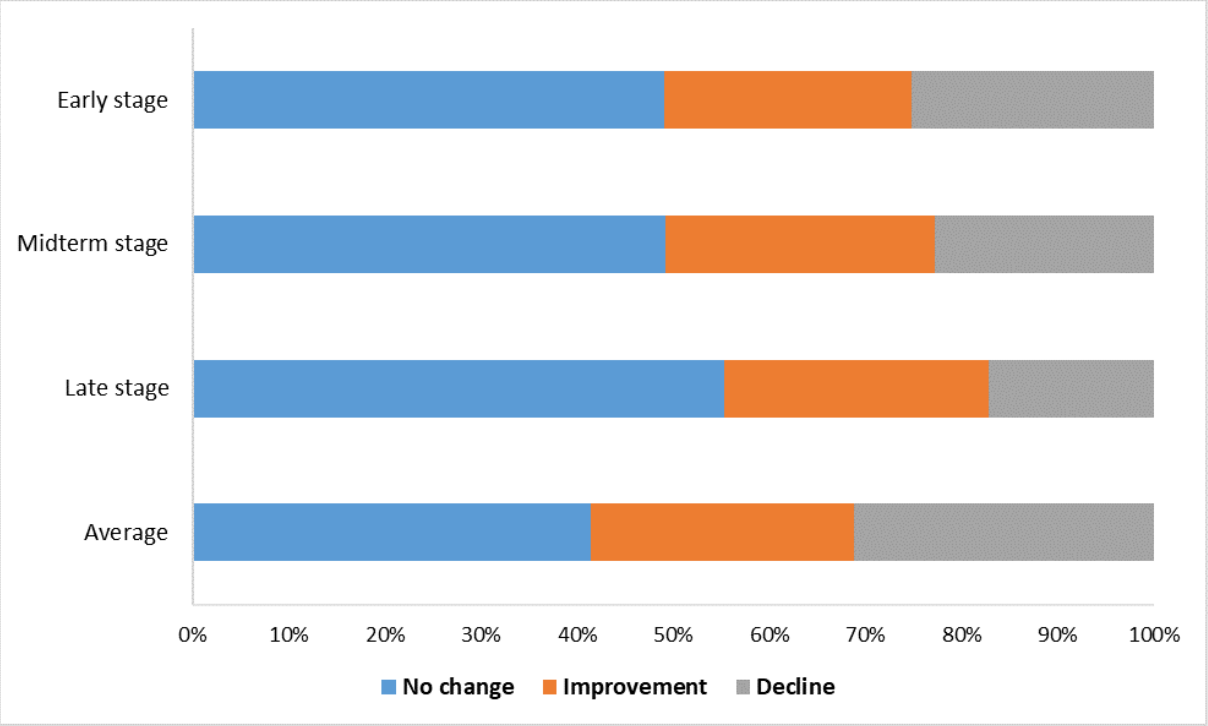


Table 3. Multinomial logistic regression analysis of the changes in adherence at each time point after using the reminder system.

Time point, variables	Improvement versus no change		Decline versus no change	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Early stage				
Reminder group	1.65 (1.01-2.70)	.04	0.93 (0.57-1.53)	.78
No-reminder group	Reference	Reference	Reference	Reference
Midterm stage				
Reminder group	1.43 (0.67-3.03)	.36	1.01 (0.46-2.24)	.98
No-reminder group	Reference	Reference	Reference	Reference
Late stage				
Reminder group	0.73 (0.23-2.26)	.58	0.16 (0.03-1.03)	.054
No-reminder group	Reference	Reference	Reference	Reference
Average				
Reminder group	1.82 (1.10-3.01)	.02	0.90 (0.55-1.48)	.67
No-reminder group	Reference	Reference	Reference	Reference

Discussion

Principal Findings

Our study explored the effect of a reminder system on PrEP adherence to promote adherence and support HIV prevention efforts in the MSM population. We found that the reminder system contributed to an improvement in adherence in the early stage. Previous studies have mentioned that text message reminders are one of the interventions that can improve adherence [27]. Two pilot studies and a PrEP demonstration project in Kenya and Brazil have also demonstrated that incorporating text message reminders into PrEP practices is an acceptable method [28-30]. These reminder methods have been found to be effective in improving adherence by increasing the user's knowledge of the medicine, motivation, and perceived support through text messages [31]. Meanwhile, according to Rogers' Protective Motivation Theory, users' motivation factors may be the driving force for individual behavior change [32]. A prior study has also confirmed that the Protective Motivation Theory can provide a valuable reference for the study of PrEP adherence in the MSM population [33]. The newly developed reminder system in our study was based on text message, and the presentation of this information can increase users' motivation to take their medicines and encourage them to change their behavior, especially in relation to PrEP adherence.

We believe that the use of the reminder system can encourage users to form medicine-taking habits. Habit is defined as behavior that is performed subconsciously without thought [34]. In the book of "The Power of Habit," Charles Duhigg defines habit as a 3-step loop of cue, routine, and reward [35]. Habits are behaviors that are induced by situational cues, and our reminder system serves as a cue. Previous findings have shown that habit formation interventions have the potential to improve adherence to antiretroviral therapy [36]. In the meanwhile, as mentioned in the previously published review, the idea of "habit" has been applied to the Medication Usage Skills for Effectiveness program [35]. In their program, users selected cues to help them remember to take their medicine, such as a specific time, a meal, or a reliable daily ritual. According to the findings of that study, the reminder system provided daily cues to users, motivating them to develop the habit of taking medicine and improving their adherence. However, the reminder system was ineffective in preventing adherence declines in the early stage as well as changes in adherence in the midterm and late stages of follow-up. We believe that there are several main reasons for this.

First, according to previous survey results, 70.21% of MSM did not take their medicines because they forgot, while 29.08% said they were too busy, 28.01% were worried about the side effects, and 18.44% thought it was troublesome to take medicine [15]. Forgetting to take medicine is the main factor affecting adherence, but there are still other factors that can potentially impact adherence. Reminder systems are primarily targeted at MSM who forget to take their medicine, and the effect of reminder systems on adherence may not be significant for those users with low willingness to take medicine. Therefore, reminder systems are ineffective in stopping the decline in adherence in MSM in the early stage.

Second, as we mentioned before, practice over many repetitions can help develop habits. The increased level of early stage adherence may be related to the reminder system that motivates users to develop the habit of taking medicine. However, repetitive activity eventually transforms the individual's cognitive control from a conscious to an automatic process [37]. This suggests that once a habit is formed, the act of taking medicines shifts from a conscious motivated behavior to an unconscious automatic one, which does not require external reminders. This may lead to a less effective reminder system in the midterm and late stages. Therefore, if the reminder system is used for medicine management in the early stage, it can not only promote the formation of medicine habits and improve adherence among users but also make the use of reminder systems more targeted and effective.

In addition, while most participants expressed positive attitude toward using the reminder system, a few expressed concerns about this approach. Fear of privacy disclosure was the main reason. For example, it is still possible for others to see the messages after receiving it. Owing to the special nature of MSM population, the difference in public awareness, and the complexity of the public opinion environment, privacy disclosure in this population may cause physical and psychological harm and negative impact on personal life. Once their privacy is leaked, they may face problems such as discrimination, stigma, breakdown of social relationships, and even experience anxiety, depression, and other psychological disorders [38-40]. After receiving a reminder message, users may be likely to quickly close the page to prevent it from being seen by others. There are also some users who may unbind with the backend system in order not to be discovered by their family and friends. However, this may reduce the reminder system's effectiveness, resulting in poor performances in the midterm and late stages. Thus, one of the most important aspects of system optimization is the privacy protection of the users.

Moreover, as a new developed reminder system for improving PrEP adherence, it does have some problems that were not considered. For instance, what are the attitudes of users toward this daily reminder? We do not have any surveys on this. Is the frequency of the daily reminders feasible? Would it be more effective to have a weekly reminder instead? It emphasizes the necessity for more personalization. Administrators can communicate with users face-to-face, design reminders to their individual needs, set reminder frequency, and provide feedback and corrections. It was also mentioned in an in-depth qualitative interview that most users showed a strong preference for feedback mechanisms [19]. The ability to respond and receive information may help increase users' motivation to participate in the program and improve health outcomes. For example, users can interact with the system, and the system will send encouraging words like "Good job" after taking the medicine. In addition, we can try to include health information and education in the WeChat reminder messages, for example, the risk of HIV infection if you do not use PrEP consistently. In this way, we can increase the feedback from the system to the users. The comments and suggestions provided by users provide an important basis for understanding messaging preferences and operating procedures, which further help managers optimize text messages and overall implementation methods.

The reminder system has several advantages. First, although the development and management of the reminder system is challenging, the issue of message costs is fully considered compared to SMS text messaging. Many literatures on SMS text messaging usage do not focus on reducing message costs [41,42], while the reminder system we use does not incur any cost as long as it is bound to the backend system for the first time. Second, according to previous studies, users prefer WeChat as a platform for receiving information and interacting with each other [25]. Compared to receiving text messages, app-based reminders are visually and formally more vivid and interesting, which provide a more convenient platform for our reminder system and are more likely to inspire interest and confidence in users. At the same time, preliminary experience with reminder system practice suggests that given the widespread use of the WeChat app compared to smartphones and other internet resources, this reminder method may be a relatively simple, convenient, and a quick tool to support HIV prevention efforts among high-risk populations [17,43,44]. MSM may benefit from a technology-based intervention that can also be integrated into routine HIV education for high-risk populations. In addition to PrEP, adherence also plays an important role in areas such as HIV antiretroviral therapy, clinical research in traditional Chinese medicine, and chronic diseases [45-47]. Therefore, there is no denying the importance and potential applicability of this reminder system for medicine adherence. The usage of a reminder system not only promotes healthy medicine-taking habits but also decreases the possibility of privacy breaches and costs associated with long-term use. This system is expected to be used on a broader scale after the improvements and upgrades, with a focus on evaluating social and economic benefits as well as targeted dissemination of health education.

Limitations

There were some limitations in this study. First, adherence was measured by self-report. Owing to recall bias, inaccurate reporting of adherence may occur [48,49]. However, it may also have a social desirability bias, which will cause adherence to be overestimated [50,51]. A previous study has also pointed out that self-reported data may be reliable [52]. Second, the initial design protocol for our study was a prospective cohort study based on a randomized controlled trial in which study participants were randomized to a no-reminder group and a reminder group, and they started using the reminder system once they entered the cohort. However, owing to the impact of COVID-19, the reminder system's development was delayed, and a portion of the MSM was already recruited. Because of the particularity and privacy of the MSM, finding eligible participants was very difficult. To avoid attrition of the study participants, we had to change the protocol to include the first 3 months of follow-up as the observation period without using a reminder system. When the first 3-month follow-up visit was conducted, the reminder system was initiated, which addressed the problem of the reminder system's development delay. Nevertheless, this could have a potential impact on our results. Since the study participants took the medicine 3 months in advance, the efficacy of the reminder system was underestimated as they may have developed good adherence during this time. In addition, for the reminder group, we did not ask more questions about the reminder messages. For example, did the participants read the reminder messages every day? Therefore, we were unable to perform a subgroup analysis to assess the effect of different groups. Future studies could focus on more detailed subgroups. Lastly, the specificity of MSM population and the impact of COVID-19 posed significant challenges in recruitment and the follow-up for our study; therefore, longitudinal cohort maintenance in this population was not optimistic.

Conclusions

Our study constructed a longitudinal cohort of PrEP adherence, explored the factors associated with changes in adherence, and further confirmed the effect of the WeChat-based reminder system on improving PrEP adherence in the MSM population in Western China. The effectiveness of the reminder system in improving adherence is currently significant in the early stage, which is considered to be related to the increased motivation of the users and the development of medicine-taking habits. Although it still can be improved and upgraded, there is no denying the importance and potential applicability of this reminder system for early-stage medicine management, which can help users build medicine-taking habits and increase adherence. Meanwhile, enhancing the privacy protection of reminder system and providing more personalized services and informative feedback are priorities for future studies, which will help improve adherence in the MSM population and promote PrEP implementation.

Acknowledgments

The authors thank all men who have sex with men participants in this study. In addition, the authors thank the following persons for their support to this study: Ailong Huang, Jianghong Dai, Yuwen Gao, Ruibin Deng, and Cuie Kong. Furthermore, the authors thank the members of the nongovernment organizations and the doctors working in the First Hospital of Chongqing Medical University.

Conflicts of Interest

None declared.

References

1. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, Partners PrEP Study Team. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med 2012 Aug 02;367(5):399-410 [FREE Full text] [doi: 10.1056/NEJMoa1108524] [Medline: 22784037]
2. Peng P, Su S, Fairley CK, Chu M, Jiang S, Zhuang X, et al. A global estimate of the acceptability of pre-exposure prophylaxis for HIV among men who have sex with men: a systematic review and meta-analysis. AIDS Behav 2018 Apr;22(4):1063-1074. [doi: 10.1007/s10461-017-1675-z] [Medline: 28176168]
3. Qin Q, Guo W, Tang W, Mahapatra T, Wang L, Zhang N, et al. Spatial analysis of the human immunodeficiency virus epidemic among men who have sex with men in China, 2006-2015. Clin Infect Dis 2017 Apr 01;64(7):956-963 [FREE Full text] [doi: 10.1093/cid/cix031] [Medline: 28362948]
4. Hu Y, Zhong X, Peng B, Zhang Y, Liang H, Dai J, et al. Comparison of depression and anxiety between HIV-negative men who have sex with men and women (MSMW) and men who have sex with men only (MSMO): a cross-sectional study in Western China. BMJ Open 2019 Jan 03;9(1):e023498 [FREE Full text] [doi: 10.1136/bmjopen-2018-023498] [Medline: 30610021]
5. Longino A, Montano MA, Sanchez H, Bayer A, Sanchez J, Tossas-Milligan K, et al. Increasing PrEP uptake and adherence among MSM and TW sex workers in Lima, Perú: what and whom do different patients trust? AIDS Care 2020 Feb;32(2):255-260 [FREE Full text] [doi: 10.1080/09540121.2019.1634787] [Medline: 31242753]
6. Chemnasiri T, Varangrat A, Amico KR, Chitwarakorn A, Dye BJ, Grant RM, et al. Facilitators and barriers affecting PrEP adherence among Thai men who have sex with men (MSM) in the HPTN 067/ADAPT Study. AIDS Care 2019 Jun 03;32(2):249-254. [doi: 10.1080/09540121.2019.1623374] [Medline: 31159584]
7. Davies O, Ustianowski A, Fox J. Pre-exposure prophylaxis for HIV prevention: why, what, who and how. Infect Dis Ther 2016 Dec;5(4):407-416 [FREE Full text] [doi: 10.1007/s40121-016-0128-8] [Medline: 27677264]
8. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, iPrEx study team. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829 [FREE Full text] [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
9. R Amico K, McMahan V, Goicochea P, Vargas L, Marcus JL, Grant RM, et al. Supporting study product use and accuracy in self-report in the iPrEx study: next step counseling and neutral assessment. AIDS Behav 2012 Jul;16(5):1243-1259. [doi: 10.1007/s10461-012-0182-5] [Medline: 22460228]
10. Closson EF, Mitty JA, Malone J, Mayer KH, Mimiaga MJ. Exploring strategies for PrEP adherence and dosing preferences in the context of sexualized recreational drug use among MSM: a qualitative study. AIDS Care 2018 Feb;30(2):191-198. [doi: 10.1080/09540121.2017.1360992] [Medline: 28830220]
11. Krakower D, Mayer KH. Engaging healthcare providers to implement HIV pre-exposure prophylaxis. Current Opinion in HIV and AIDS 2012;7(6):593-599. [doi: 10.1097/coh.0b013e3283590446] [Medline: 23032736]
12. Kirby T, Thornber-Dunwell M. Uptake of PrEP for HIV slow among MSM. The Lancet 2014 Feb;383(9915):399-400. [doi: 10.1016/S0140-6736(14)60137-9]
13. Mugo PM, Sanders EJ, Mutua G, van der Elst E, Anzala O, Barin B, et al. Understanding adherence to daily and intermittent regimens of oral HIV pre-exposure prophylaxis among men who have sex with men in Kenya. AIDS Behav 2015 May;19(5):794-801 [FREE Full text] [doi: 10.1007/s10461-014-0958-x] [Medline: 25432877]

14. Krakower DS, Mimiaga MJ, Rosenberger JG, Novak DS, Mitty JA, White JM, et al. Limited awareness and low immediate uptake of pre-exposure prophylaxis among men who have sex with men using an internet social networking site. PLoS One 2012;7(3):e33119 [FREE Full text] [doi: 10.1371/journal.pone.0033119] [Medline: 22470438]
15. Qu D, Zhong X, Xiao G, Dai J, Liang H, Huang A. Adherence to pre-exposure prophylaxis among men who have sex with men: A prospective cohort study. Int J Infect Dis 2018 Oct;75:52-59 [FREE Full text] [doi: 10.1016/j.ijid.2018.08.006] [Medline: 30125688]
16. Mbuagbaw L, van der Kop ML, Lester RT, Thirumurthy H, Pop-Eleches C, Ye C, et al. Mobile phone text messages for improving adherence to antiretroviral therapy (ART): an individual patient data meta-analysis of randomised trials. BMJ Open 2013 Dec 17;3(12):e003950 [FREE Full text] [doi: 10.1136/bmjopen-2013-003950] [Medline: 24345901]
17. Shrestha R, Karki P, Copenhaver M. Interest in use of mHealth technology in HIV prevention and associated factors among high-risk drug users enrolled in methadone maintenance program. AIDS Care 2017 Sep;29(9):1144-1148 [FREE Full text] [doi: 10.1080/09540121.2017.1325439] [Medline: 28478706]
18. Muwonge TR, Ngure K, Katabira E, Mugo N, Kimemia G, Burns BFO, Partners Mobile Adherence to PrEP (PMAP) Team. Short message service (SMS) surveys assessing pre-exposure prophylaxis (PrEP) adherence and sexual behavior are highly acceptable among HIV-uninfected members of serodiscordant couples in East Africa: a mixed methods study. AIDS Behav 2019 May;23(5):1267-1276 [FREE Full text] [doi: 10.1007/s10461-018-2326-8] [Medline: 30406335]
19. Shrestha R, Altice FL, DiDomizio E, Sibilio B, Ranjit YS, Copenhaver MM. Feasibility and acceptability of an mHealth-based approach as an HIV prevention strategy among people who use drugs on pre-wxposure prophylaxis. Patient Prefer Adherence 2020;14:107-118 [FREE Full text] [doi: 10.2147/PPA.S236794] [Medline: 32021122]
20. Liu AY, Laborde ND, Coleman K, Vittinghoff E, Gonzalez R, Wilde G, et al. DOT diary: developing a novel mobile app using artificial intelligence and an electronic sexual diary to measure and support PrEP adherence among young men who have sex with men. AIDS Behav 2021 Apr;25(4):1001-1012 [FREE Full text] [doi: 10.1007/s10461-020-03054-2] [Medline: 33044687]
21. Brown W, Giguere R, Sheinfil A, Ibitoye M, Balan I, Ho T, et al. Challenges and solutions implementing an SMS text message-based survey CASI and adherence reminders in an international biomedical HIV PrEP study (MTN 017). J Biomed Inform 2018 Apr;80:78-86 [FREE Full text] [doi: 10.1016/j.jbi.2018.02.018] [Medline: 29501908]
22. Haberer JE, Bukusi EA, Mugo NR, Pyra M, Kiptinness C, Oware K, et al. Effect of SMS reminders on PrEP adherence in young Kenyan women (MPYA study): a randomised controlled trial. The Lancet HIV 2021 Mar;8(3):e130-e137. [doi: 10.1016/s2352-3018(20)30307-6]
23. Li S, Zhang J, Mao X, Lu T, Gao Y, Zhang W, et al. Feasibility of indirect secondary distribution of HIV self-test kits via WeChat among men who have sex with men: national cross-sectional study in China. J Med Internet Res 2021 Oct 26;23(10):e28508 [FREE Full text] [doi: 10.2196/28508] [Medline: 34698651]
24. 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]
25. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and feasibility testing of an mHealth (text Message and WeChat) intervention to improve the medication adherence and quality of life of people living with HIV in China: pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e10274. [doi: 10.2196/10274]
26. Zha P, Mahat G, Qureshi R, Zhao L, Shen Q, Chao Y, et al. Utilising a WeChat intervention to improve HIV and AIDS education among college students in China. Health Education Journal 2021 Jul 09;80(8):1002-1013. [doi: 10.1177/00178969211023084]
27. Jennifer V, Bill K, LindaGail B, Connie C, Sybil H, Sinead D. Interventions to improve daily medication use among adolescents and young adults: what can we learn for youth PrEP services? AIDS (London, England). 2020;. FREE Full text Medline 2021 Mar 1;35(3):463-475. [doi: 10.1097/qad.0000000000002777] [Medline: 33252486]
28. Curran K, Mugo NR, Kurth A, Ngure K, Heffron R, Donnell D, et al. Daily short message service surveys to measure sexual behavior and pre-exposure prophylaxis use among Kenyan men and women. AIDS Behav 2013 Nov;17(9):2977-2985 [FREE Full text] [doi: 10.1007/s10461-013-0510-4] [Medline: 23695519]
29. Grinsztejn B, Hoagland B, Moreira RI, Kallas EG, Madruga JV, Goulart S, et al. Retention, engagement, and adherence to pre-exposure prophylaxis for men who have sex with men and transgender women in PrEP Brasil: 48 week results of a demonstration study. The Lancet HIV 2018 Mar;5(3):e136-e145. [doi: 10.1016/s2352-3018(18)30008-0]
30. Velloza J, Ngure K, Kiptinness C, Quame-Amaglo J, Thuo N, Dew K, et al. A clinic-based tablet application to support safer conception among HIV serodiscordant couples in Kenya: feasibility and acceptability study. Mhealth 2019;5:4 [FREE Full text] [doi: 10.21037/mhealth.2019.01.04] [Medline: 30976596]
31. Ware N, Pisarski E, Tam M, Wyatt M, Atukunda E, Musiimenta A. The Meanings in the messages: how SMS reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS. . FREE Full text Medline 2016 May 15;30(8):1287-1294. [doi: 10.1097/qad.0000000000001035] [Medline: 26807967]
32. Maddux J, Rogers R. Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. Journal of Experimental Social Psychology 1983 Sep 02;19(5):469-479. [doi: 10.1016/0022-1031(83)90023-9] [Medline: 28136248]

33. Liu J, Deng R, Lin B, Pan H, Gao Y, Dai J, et al. Risk management on pre-exposure prophylaxis adherence of men who have sex with multiple men: a multicenter prospective cohort study. RMHP 2021 Apr;Volume 14:1749-1761. [doi: 10.2147/rmhp.s295114]
34. Gardner B. A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. Health Psychol Rev 2015;9(3):277-295 [FREE Full text] [doi: 10.1080/17437199.2013.876238] [Medline: 25207647]
35. Vrijens B, Urquhart J, White D. Electronically monitored dosing histories can be used to develop a medication-taking habit and manage patient adherence. Expert Rev Clin Pharmacol 2014 Sep;7(5):633-644. [doi: 10.1586/17512433.2014.940896] [Medline: 25088004]
36. Jennings Mayo-Wilson L, Devoto B, Coleman J, Mukasa B, Shelton A, MacCarthy S, et al. Habit formation in support of antiretroviral medication adherence in clinic-enrolled HIV-infected adults: a qualitative assessment using free-listing and unstructured interviewing in Kampala, Uganda. AIDS Res Ther 2020 Jun 08;17(1):30 [FREE Full text] [doi: 10.1186/s12981-020-00283-2] [Medline: 32513192]
37. Aarts H, Paulussen T, Schaalma H. Physical exercise habit: on the conceptualization and formation of habitual health behaviours. Health Educ Res 1997 Sep;12(3):363-374. [doi: 10.1093/her/12.3.363] [Medline: 10174219]
38. Mao Y, Li X, Qiao S, Zhou Y, Zhao Q. Ethnicity, stigma and adherence to antiretroviral therapy (ART) among people living with HIV/AIDS in Guangxi, China. J AIDS Clin Res 2017;08(01):652. [doi: 10.4172/2155-6113.1000652] [Medline: 30140549]
39. Allam RR, Takamiya M, Pant R, Gandham S, Yeldandi VV, Thomas J, et al. Factors associated with non-adherence to antiretroviral therapy among female sex workers living with HIV in Hyderabad, India. Int J STD AIDS 2020 Jul;31(8):735-746. [doi: 10.1177/0956462420920145] [Medline: 32631214]
40. Zeng C, Li X, Qiao S, Yang X, Shen Z, Zhou Y. Anticipated stigma and medication adherence among people living with HIV: the mechanistic roles of medication support and ART self-efficacy. AIDS Care 2020 Aug;32(8):1014-1022 [FREE Full text] [doi: 10.1080/09540121.2020.1728213] [Medline: 32336130]
41. Bock B, Heron K, Jennings E, Morrow K, Cobb V, Magee J, et al. A text message delivered smoking cessation intervention: the initial trial of TXT-2-Quit: randomized controlled trial. JMIR Mhealth Uhealth 2013 Jul 30;1(2):e17 [FREE Full text] [doi: 10.2196/mhealth.2522] [Medline: 25098502]
42. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
43. Patel A, Kessler J, Braithwaite R, Nucifora K, Thirumurthy H, Zhou Q. Economic evaluation of mobile phone text message interventions to improve adherence to HIV therapy in Kenya. Medicine. . FREE Full text Medline 2017;96(7):e6078. [doi: 10.1097/md.0000000000006078] [Medline: 28207516]
44. Shrestha R, Huedo-Medina TB, Altice FL, Krishnan A, Copenhaver M. Examining the acceptability of mHealth technology in HIV prevention among high-risk drug users in treatment. AIDS Behav 2017 Nov;21(11):3100-3110 [FREE Full text] [doi: 10.1007/s10461-016-1637-x] [Medline: 28025735]
45. Zheng W, Chang B, Chen J. Improving participant adherence in clinical research of traditional chinese medicine. Evid Based Complement Alternat Med 2014;2014:376058 [FREE Full text] [doi: 10.1155/2014/376058] [Medline: 24527045]
46. Georgette N, Siedner MJ, Zanoni B, Sibaya T, Petty CR, Carpenter S, et al. The acceptability and perceived usefulness of a weekly clinical SMS program to promote HIV antiretroviral medication adherence in KwaZulu-Natal, South Africa. AIDS Behav 2016 Nov;20(11):2629-2638 [FREE Full text] [doi: 10.1007/s10461-016-1287-z] [Medline: 26781866]
47. Al-Noumani H, Al-Harrasi M, Jose J, Al-Naamani Z, Panchatcharam SM. Medication adherence and patients' characteristics in chronic diseases: a national multi-center study. Clin Nurs Res 2022 Mar;31(3):426-434. [doi: 10.1177/10547738211033754] [Medline: 34287084]
48. Buscher A, Hartman C, Kallen MA, Giordano TP. Validity of self-report measures in assessing antiretroviral adherence of newly diagnosed, HAART-naïve, HIV patients. HIV Clin Trials 2011;12(5):244-254 [FREE Full text] [doi: 10.1310/hct1205-244] [Medline: 22180522]
49. Musinguzi N, Muganzi C, Boum Y, Ronald A, Marzinke M, Hendrix C. Comparison of subjective and objective adherence measures for preexposure prophylaxis against HIV infection among serodiscordant couples in East Africa. AIDS. . Medline 2016;30(7):1121-1129. [doi: 10.1097/qad.0000000000001024] [Medline: 26785125]
50. Williams AB, Amico KR, Bova C, Womack JA. A proposal for quality standards for measuring medication adherence in research. AIDS Behav 2013 Jan;17(1):284-297 [FREE Full text] [doi: 10.1007/s10461-012-0172-7] [Medline: 22407465]
51. Baxi S, Liu A, Bacchetti P, Mutua G, Sanders E, Kibengo F. Comparing the novel method of assessing PrEP adherence/exposure using hair samples to other pharmacologic and traditional measures. JAIDS Journal of Acquired Immune Deficiency Syndromes. . FREE Full text Medline 2015;68(1):13-20. [doi: 10.1097/qai.0000000000000386] [Medline: 25296098]
52. Marins LMS, Torres TS, Leite IDC, Moreira RI, Luz PM, Hoagland B, et al. Performance of HIV pre-exposure prophylaxis indirect adherence measures among men who have sex with men and transgender women: Results from the PrEP Brasil Study. PLoS One 2019;14(8):e0221281 [FREE Full text] [doi: 10.1371/journal.pone.0221281] [Medline: 31430318]

Abbreviations

MSM: men who have sex with men
OR: odds ratio
PrEP: pre-exposure prophylaxis

Edited by G Eysenbach; submitted 12.03.22; peer-reviewed by B Rewari, X Xu; comments to author 12.06.22; revised version received 16.06.22; accepted 12.07.22; published 11.08.22.

Please cite as:
Lin B, Liu J, He W, Pan H, Ma Y, Zhong X
Effect of a Reminder System on Pre-exposure Prophylaxis Adherence in Men Who Have Sex With Men: Prospective Cohort Study Based on WeChat Intervention
J Med Internet Res 2022;24(8):e37936
URL: https://www.jmir.org/2022/8/e37936
doi:10.2196/37936
PMID:35969436

Original Paper

The Effects of Internet-Based Storytelling Programs (Amazing Adventure Against Stigma) in Reducing Mental Illness Stigma With Mediation by Interactivity and Stigma Content: Randomized Controlled Trial

Tiffany H C Fong[1*], BSSc; Winnie W S Mak[1*], PhD

Department of Psychology, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[*] all authors contributed equally

Corresponding Author:
Winnie W S Mak, PhD
Department of Psychology
The Chinese University of Hong Kong
Shatin, NT
Hong Kong
Hong Kong
Phone: 852 39436577
Fax: 852 26035019
Email: wwsmak@psy.cuhk.edu.hk

Abstract

Background: Mental illness stigma has been a global concern, owing to its adverse effects on the recovery of people with mental illness, and may delay help-seeking for mental health because of the concern of being stigmatized. With technological advancement, internet-based interventions for the reduction of mental illness stigma have been developed, and these effects have been promising.

Objective: This study aimed to examine the differential effects of internet-based storytelling programs, which varied in the levels of interactivity and stigma content, in reducing mental illness stigma.

Methods: Using an experimental design, this study compared the effects of 4 storytelling websites that varied in the levels of interactivity and stigma content. Specifically, the conditions included an interactive website with stigma-related content (*combo* condition), a noninteractive website with stigma-related content (*stigma* condition), an interactive website without stigma-related content (*interact* condition), and a noninteractive website without stigma-related content (*control* condition). Participants were recruited via mass emails to all students and staff of a public university and via social networking sites. Eligible participants were randomized into the following four conditions: *combo* (n=67), *stigma* (n=65), *interact* (n=64), or *control* (n=67). The participants of each group viewed the respective web pages at their own pace. Public stigma, microaggression, and social distance were measured on the web before the experiment, after the experiment, and at the 1-week follow-up. Perceived autonomy and immersiveness, as mediators, were assessed after the experiment.

Results: Both the *combo* (n=66) and *stigma* (n=65) conditions were effective in reducing public stigma and microaggression toward people with mental illness after the experiment and at the 1-week follow-up. However, none of the conditions had significant time×condition effects in reducing the social distance from people with mental illness. The *interact* condition (n=64) significantly reduced public stigma after the experiment (P=.02) but not at the 1-week follow-up (P=.22). The *control* condition (n=67) did not significantly reduce all outcomes associated with mental illness stigma. Perceived autonomy was found to mediate the effect of public stigma (P=.56), and immersiveness mediated the effect of microaggression (P=.99).

Conclusions: Internet-based storytelling programs with stigma-related content and interactivity elicited the largest effects in stigma reduction, including reductions in public stigma and microaggression, although only its difference with internet-based storytelling programs with stigma-related content was not statistically significant. In other words, although interactivity could strengthen the stigma reduction effect, stigma-related content was more critical than interactivity in reducing stigma. Future stigma reduction efforts should prioritize the production of effective stigma content on their web pages, followed by considering the value of incorporating interactivity in future internet-based storytelling programs.

Trial Registration: ClinicalTrials.gov NCT05333848; https://clinicaltrials.gov/ct2/show/NCT05333848

(J Med Internet Res 2022;24(8):e37973) doi:10.2196/37973

KEYWORDS

mental illness stigma; internet-based; interactivity; storytelling; social distance; microaggression

Introduction

Mental Illness Stigma

According to the social cognitive model of stigma, mental illness stigma is defined as having stereotypical thoughts, prejudicial feelings, and discriminatory actions against people with mental illness in situations with power differentials [1]. Mental illness stigma is a global concern owing to its detrimental effects imposed on people with mental illness across various life domains (eg, education, housing, employment, and health care) during their recovery and their willingness to seek help [1-3].

Given the possible negative impacts of mental illness stigma, various approaches have been proposed to reduce it, with education and contact being identified as the 2 most common and effective approaches in inducing positive attitudinal changes and reducing public stigma [4-7]. Knowledge enhancement through psychoeducation can improve mental health literacy, which corrects misunderstanding of mental health–related issues, challenges stigmatizing beliefs, and supports recovery from mental illness [7]. Contact is effective in confronting stigmatization through an equal, interpersonal exchange that fosters perspective taking and empathy [4]. The primary form of contact is an in vivo one, which requires the person with lived experience to share his or her experiences of mental illness and stigmatization with an audience live. However, it can be costly and taxing for speakers to share their experiences repeatedly [8]. Hence, researchers have investigated the possibilities of using parasocial contacts, which include video-based, filmed, and web-based contacts [4,8-11]. These have exhibited similar effects as were found with in vivo contacts [4,11].

Internet-Based Stigma Reduction Interventions

Recently, internet-based programs addressing mental illness stigma have been established worldwide owing to their low cost, accessibility, and scalability [12-14]. Studies have shown the effectiveness of internet-based stigma reduction programs in the form of social media platforms, digital games, webinars, filmed social contact, and websites [9,15-19]. Research has also shown that internet-based and face-to-face stigma reduction programs are equally effective [20,21]. Research has also found increased empathy and reduced intergroup anxiety to be significant mediators explaining the effects of interventions on the reduction of mental illness stigma [22,23]. However, the content and design of these internet-based stigma reduction programs vary greatly, and limited efforts have been made to investigate the common factors contributing to their effectiveness.

In addition to incorporating the critical determinants, namely education and contact, in stigma reduction, many internet-based interventions have made use of interactivity and storytelling in their designs and have demonstrated positive results in reducing mental illness stigma [24-28]. However, the types of interactivity are diverse, and it is unknown whether the addition of interactivity induces significant positive attitudinal changes that should be valued. Despite the popularity of incorporating elements of interactivity into websites to enhance the engagement of users, there is minimal empirical work investigating the effect of interactivity on stigma reduction. Thus, rather than focusing on psychological variables (ie, empathy and intergroup anxiety) as mediators (the effects of which have been established), this study focused on the mediators related to the form of interventions and aimed to examine the differential effects of internet-based storytelling programs with the presence or absence of interactivity and stigma content in reducing mental illness stigma.

Stigma-Related Content

As previously mentioned, educating about mental illness and fostering an understanding toward people with mental illnesses helps individuals to clear misunderstandings and empathize with their experiences [7]. Hence, it is pivotal to introduce stigma-related content into stigma reduction programs. However, research has shown significant positive attitudinal changes in *control* conditions where stigma-related content is not present [28,29]. Although the situation is rare, it is proposed that the effect might be attributed to the social desirability effect and priming effect in which people have been prompted to answer questions related to mental illness attitudes at preassessment [28,29]. In this study, stigma-related content was one of the independent variables that accounted for this possible effect. Despite some cases of stigma reduction in the absence of stigma content, we hypothesized that the reduction of stigma will only manifest in conditions with the presence of stigma content.

Storytelling

In previous research related to interventions aimed at reducing mental illness stigma, the use of storytelling has demonstrated significant reduction in stigma [24,28,30]. Corrigan and Kosyluk [31,32] have identified three crucial components that make storytelling effective in reducing mental illness stigma, namely people with lived experience as storytellers, in-person story delivery, and descriptions encompassing both ups and downs on the recovery journey of mental illness. Disclosure from storytellers enables people to understand the experience shared in a deeper way with the aid of contexts and connection with the storytellers [33]. Storytelling was found to be positively related to people’s reflection, motivation, and engagement, which not only allows people to have a better understanding of people with mental illnesses cognitively but also cultivates empathy in people [30].

Interactivity

Empirically, few studies have shown that communication using interactive media could exert synergistic effects with education in reducing stigmatizing attitudes toward people with mental illness [26,27,34]. However, past efforts investigating internet-based stigma reduction programs that have incorporated interactivity have mainly used interactivity in different delivery formats. For instance, a study allowed participants to choose the sequence of reading materials but not the content [26]. Another study allowed individuals to undergo simulated contact by using web-based interactions with avatars who shared their emotional distress [27]. There is still a dearth of empirical studies investigating the effects of interactivity in internet-based stigma reduction programs.

Scholars have proposed that the effect might be attributed to the inducement of positive affect during the processing of mental illness information on the internet [26,35]. According to the Systemic Thinking Model, in interactive environments, interactivity allows individuals to be the agent and effect physical environmental changes that best align with their thinking needs and flow [36]. Individuals actively manipulate information and receive contingent feedback, which facilitates the encoding of new information [37]. In turn, information processing and learning are facilitated [36]. Empirical research has found that interactivity induces positive affect, including increased satisfaction and pleasure [38]. Individuals tend to favor interactive information through attitude transfer mechanisms [39]. In addition, research shows that incidental positive affect can reduce complex affective judgments toward outgroup members and lead to more prosocial orientations [40]. These findings suggest that interactivity could lead to positive affect and reduced stigmatizing attitudes toward people with mental illness even without the presence of stigma content aiming to correct misconceptions.

Interactivity is a favorable element in learning. McMillan proposed four main types of interaction, namely user-to-user, user-to-content, user-to-medium, and medium-to-medium interactions [41]. Most internet-based programs have used user-to-content interaction, where individuals can interact with information on the internet [42]. Kim and Stout [26] examined a type of user-to-content interaction by allowing individuals to control the navigation sequence of web pages. In this study, individuals interacted with the contents of web pages by choosing their actions and responses, and the possible factors mediating the interactivity effect were analyzed.

Possible Mechanisms of Change

Research has found interactivity to have a significant role in improving information processing through enhanced motivation, which facilitates stigma reduction [26]. Perceived autonomy and immersiveness have been found to enhance motivation [43,44]. Thus, they might be possible mediators of the effect of interactivity although their relationships remain untapped.

Perceived autonomy refers to the perception of being the agent, which makes one consider the experience and behavior as concordant with one's integrated sense of the self [45]. When perceived autonomy is high, autonomy need satisfaction is achieved and cognitive changes are facilitated [46]. By allowing individuals to select their responses as an avatar on the website, the heightened perceived autonomy might facilitate the endorsement of responses and actions selected, which may foster cognitive changes. Therefore, perceived autonomy might be a possible mediator of cognitive changes that correct misbeliefs about mental illnesses.

Immersiveness indicates the subjective feelings of participating in a comprehensive and realistic experience [47]. Storytelling has been evident in inducing immersiveness [48]. With higher levels of immersiveness, individuals are driven to take in messages conveyed in the story [49], which in turn leads to cognitive and affective changes related to mental illness stigma [50-53]. Interactivity also allows people to engage with the content, which is also considered an immersive medium [47]. Thus, immersiveness was proposed to be another possible mediator leading to stigma reduction.

Aims and Hypotheses

This experimental study aimed to investigate the effect of internet-based storytelling programs on the manipulation of stigma-related content and interactivity. We hypothesized that an internet-based storytelling program with a combination of interactivity and stigma content would lead to the most significant reduction in public stigma, microaggression, and social distance from people with mental illnesses, followed by an internet-based storytelling program with stigma content only and interactivity only, compared with the control group. Second, we hypothesized that the effects observed in stigma reduction would be mediated by perceived autonomy and immersiveness owing to the presence of interactivity.

Methods

Study Design

This experimental study compared the following four internet-based storytelling programs: an interactive stigma content website (*combo* condition), a noninteractive stigma content website (*stigma* condition), an interactive nonstigma content website (*interact* condition), and a noninteractive nonstigma content website (*control* condition).

Ethics Approval

Ethics approval for behavioral research was obtained from the Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong, and the study was registered on ClinicalTrials. gov (NCT05333848). The findings of this study were reported in accordance with the recommendations of the CONSORT (Consolidated Standards of Reporting Trials) guidelines.

Participants

This study targeted people who were aged ≥18 years and able to read and understand Chinese. Recruitment was performed by sending mass emails to students and staff of a public university in Hong Kong and by posting advertisements on social media. Individuals who were interested in participating in the study visited the registration link, where they were screened through a web-based survey on basic contact information and age. The experimenter (THCF) then provided eligible individuals a Zoom (Zoom Video Communications Inc) appointment link, where individuals indicated their preferred day and time to participate in the experiment. A Zoom link was provided to individuals upon the completion of their booking. At the scheduled Zoom experimental session, participants were given detailed information about the study aims, length of the program, and participant involvement. Participants provided informed consent by checking the *I agree* button at the end of the study description page. Afterward, participants received another web-based questionnaire link to complete the pre-experiment questionnaire. The participants were randomly assigned to 1 of the 4 experimental conditions through block randomization. Participants completed the pre-experiment, postexperiment, and 1-week follow-up questionnaires on the web.

Storytelling Programs

The four internet-based storytelling programs were administered via the internet with 4 different web pages that were displayed in the Chinese language. The design of the web pages, including the story content and the use of graphics, was created to appeal widely to the adult population without catering specifically to a certain gender or age group. The Amazing Adventure Against Stigma website [54] was used in the experimental study for the *combo* and *stigma* conditions. Animation with fictional backgrounds (eg, mountains, canyons, and safari) and avatars of diverse body types, height, skin color, and gender-neutral hairstyle and clothing were used to minimize the effect of cultural and environmental effects that may be more prominent in using real persons and real settings on the participants and to maximize the possibility of adults with diverse backgrounds relating to the avatars. It is also easier and more economical to create control animations by using existing software than to produce videos with real people in real-life settings. Each web page took approximately 20 minutes to browse through. The presence or absence of interactivity and the presence or absence of stigma content were manipulated on the four web pages. All web pages involved a story. For the *combo* and *stigma* conditions, the story was identical, which was about the journey of a person experiencing mental illness stigma. For the *interact* and *control* conditions, the story was also identical and nonstigma related, illustrating the typical day of a person. Interactivity was manipulated by adding interactive elements to the *combo* and *interact* conditions, where participants could choose their actions and responses on web pages. In these 2 conditions, the participants had internet-based contact with the protagonist in the story, where they learned about the life experiences of the protagonist through the story portrayed on the web pages. Contact and interaction with the protagonist were in the moment as the story regarding the lived experience of the protagonist unfolded along the journey and the participants chose their responses to continue their interaction with the protagonist. The story content in the *combo* and *stigma* conditions was organized based on the disclosure of a person with a lived experience of mental illness. The person with a lived experience of mental illness accompanied participants to visualize their microaggressive encounters in various life domains (eg, work, family, and social circle) and the public's misunderstanding of mental illness with the aid of visual images on the web page. The story also incorporated messages about the interconnection between people with or without mental illness. The selection of the aforementioned story content was based on the previous identification of both education and contact as the most effective approaches for inducing stigma reduction [4-7].

The story content in the *interact* and *control* conditions formulated a typical day for a person, which began with the morning routine, followed by having breakfast, going to work, working encounters, and ending the day. The ordinary storyline was created to minimize affective arousal and, in turn, minimize influences on judgment and decision-making according to the affect-as-information framework proposed by Storbeck and Clore [55]. At the end of each web page experience, participants were provided with a questionnaire link that measured microaggression, public stigma, social distance from people with mental illness, perceived autonomy, and immersiveness. One week after the experimental session, participants completed a follow-up questionnaire assessing microaggression, public stigma, and social distance from people with mental illness. Finally, participants were debriefed. Participants could not reaccess the contents of the web pages after the experimental session. Screenshots of the web page interventions are provided in Multimedia Appendix 1.

Measures

Baseline Measures

At baseline, participants provided demographic and background information, including age, gender, education level, religion, and previous experience with mental illness.

Contact With People Having Mental Illness

To assess one's previous experience with mental illness, the Level of Contact Report [56] was used, where participants indicated whether they had the experiences reported in the 12 items such as "I have watched a movie or television show in which a character depicted a person with mental illness" and "I have observed persons with a severe mental illness on a frequent basis." Higher scores indicated higher levels of previous contact with people with mental illnesses.

Mental Illness Stigma Measures

Public Stigma Toward People With Mental Illness

The 21-item Public Stigma Scale-Mental Illness-Short Version [57] was used to assess public stigma regarding mental illness and personal advocacy. Each item was rated on a 6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree). Sample items included "People with mental illness are a burden to society" (public stigma) and "I wholeheartedly fight for the rights of people with mental illness" (personal advocacy). Reverse scoring was performed for personal advocacy items. Higher scores indicate higher levels of public stigma. In this study, its Cronbach α values were .93, .95, and .94 at baseline, after the experiment, and at the 1-week follow-up, respectively.

Microaggression

Microaggression was measured using the 17-item Mental Illness Microaggressions Scale [58], which covers the assumption of inferiority, patronization, and fear of mental illness. Each item was rated on a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). Sample items included "If someone I'm close to told me that they had a mental illness diagnosis, I would expect them to have trouble understanding some things" (assumption of inferiority), "If someone I'm close to told me that they had a mental illness diagnosis, I would give them advice on how to remain stable" (patronization), and "If I saw a person who I thought had a mental illness in public, I would keep my distance from them" (fear of mental illness). Higher scores indicate higher levels of microaggression. In this study, the Cronbach α values of the Mental Illness Microaggressions Scale were .78, .86, and .87 at baseline, after the experiment, and at the 1-week follow-up, respectively.

Social Distance From People With Mental Illness

The 8-item Social Distancing Scale [57] was used to measure the behavioral intention to maintain social distance from people with mental illness. Participants rated the extent to which they endorsed each item from 1 (very willing) to 6 (very unwilling) on items such as "Assuming you have children, you will let persons with mental illnesses take care of your children" and "You will work with persons with mental illnesses in the same institution." In this study, its Cronbach α values were .83, .88, and .86, at baseline, after the experiment, and at the 1-week follow-up, respectively.

Mediators Measures

Perceived Autonomy

To assess the perceived autonomy of the web page experience, the 10-item Self-Determination Scale [59] was used in the postexperiment questionnaire. Each item was a pair of opposite statements, in which participants rated their level of perceived choice and self-awareness with a slider from 1 (only A feels true) to 5 (only B feels true). Sample items included item—"A. During this web page experience, I always feel like I choose the things I do. B. During this web page experience, I sometimes feel that it's not really me choosing the things I do" (perceived choice)—and item 2—"A. During this web page experience, my emotions sometimes seem alien to me. B. During this web page experience, my emotions always seem to belong to me" (self-awareness). Reverse scoring was performed for perceived choice items. In this study, its Cronbach α was .89 after the experiment.

Immersiveness

The 15-item Transportation Scale [60] was used to assess participants' immersiveness in the web experience. It used a 4-point Likert scale ranging from 1 (very much) to 4 (not at all) for items such as "I could picture myself in the scene of the events described in the web page." The last 4 items were adapted to suit the experimental conditions. In the *combo* and *stigma* conditions, the last four items were "While reading the web page, I had a vivid image of the avatar representing me"; "While reading the web page, I had a vivid image of the host"; "While reading the web page, I had a vivid image of the journey"; and "While reading the web page, I had a vivid image of the dialogue." In the *interact* and *control* conditions, the last four items were "While reading the web page, I had a vivid image of the avatar representing me"; "While reading the web page, I had a vivid image of my home"; "While reading the web page, I had a vivid image of my breakfast"; and "While reading the web page, I had a vivid image of my office." Items 2, 5, and 9 were framed negatively. All the items were scored in the direction that higher scores indicate higher levels of immersiveness. In this study, its Cronbach α was .84 after the experiment.

Data Analysis

All analyses were conducted using SPSS (version 27.0; IBM Corporation) and the moderation and mediation plug-in PROCESS. Categorical chi-square and 1-way ANOVA were used to examine baseline differences among the experimental conditions. Repeated measures ANOVA with Bonferroni adjustment and post hoc analysis were conducted to detect significant interaction effects between condition and time to see if conditions showed significant reduction in all mental illness stigma outcomes across the 3 time points. Mediation analysis was conducted using PROCESS model 4 to investigate the relationship of possible mediators, perceived autonomy, and immersiveness, with all outcomes at follow-up assessment.

Results

Participant Characteristics

A total of 263 participants were recruited for this study and completed the experimental session, pre-experiment and postexperiment questionnaires. All but 1 participant (262/263, 99.6%) completed the 1-week follow-up questionnaire. The procedure of the study is illustrated in Figure 1. Demographics and baseline characteristics of 263 participants were analyzed, and data from 262 participants were analyzed using repeated measures ANOVA and mediation analyses. The mean age of the participants was 22.56 (SD 6.16) years. Most of the study participants were women (182/263, 69.2%). The participants were predominantly university students (227/263, 86.3%), with 74.9% (197/263) being undergraduates. Among the 197 undergraduate students, 11.8% (31/197) were in year 1, 17.1% (45/197) in year 2, 22.8% (60/197) in year 3, 20.5% (54/197) in year 4, 1.9% (5/197) in year 5, and 0.8% (2/197) in year 6. The other detailed demographics and baseline characteristics of the participants are presented in Table 1.

Across the four conditions, significant gender differences were found (χ^2_3=10.0; *P*=.02). The *control* condition had a higher women-to-men ratio than the other 3 conditions. A significant difference in age across conditions was also found in the 1-way ANOVA ($F_{3,259}$=3.28; *P*=.02). A correlation analysis was performed to investigate whether gender and age were correlated with mental illness stigma outcomes. Age was not correlated with any of the outcomes, including public stigma (*r*=−0.02; *P*=.81), microaggression (*r*=−0.02; *P*=.74), and social distance (*r*=0.08; *P*=.18). Gender was weakly correlated with public stigma (*r*=−0.14; *P*=.02) and microaggression (*r*=−0.12; *P*=.05) but not with social distance (*r*=−0.01; *P*=.84). No other significant differences in demographic characteristics were found. In terms of participants' previous experiences with mental illness, no significant differences were found across the conditions. Given the baseline difference in gender across conditions, gender was included as a covariate in subsequent analyses.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of participant recruitment. *Combo* condition: interactivity present; stigma content present. *Control* condition: interactivity absent; stigma content absent. *Interact* condition: interactivity present; stigma content absent. *Stigma* condition: interactivity absent; stigma content present.

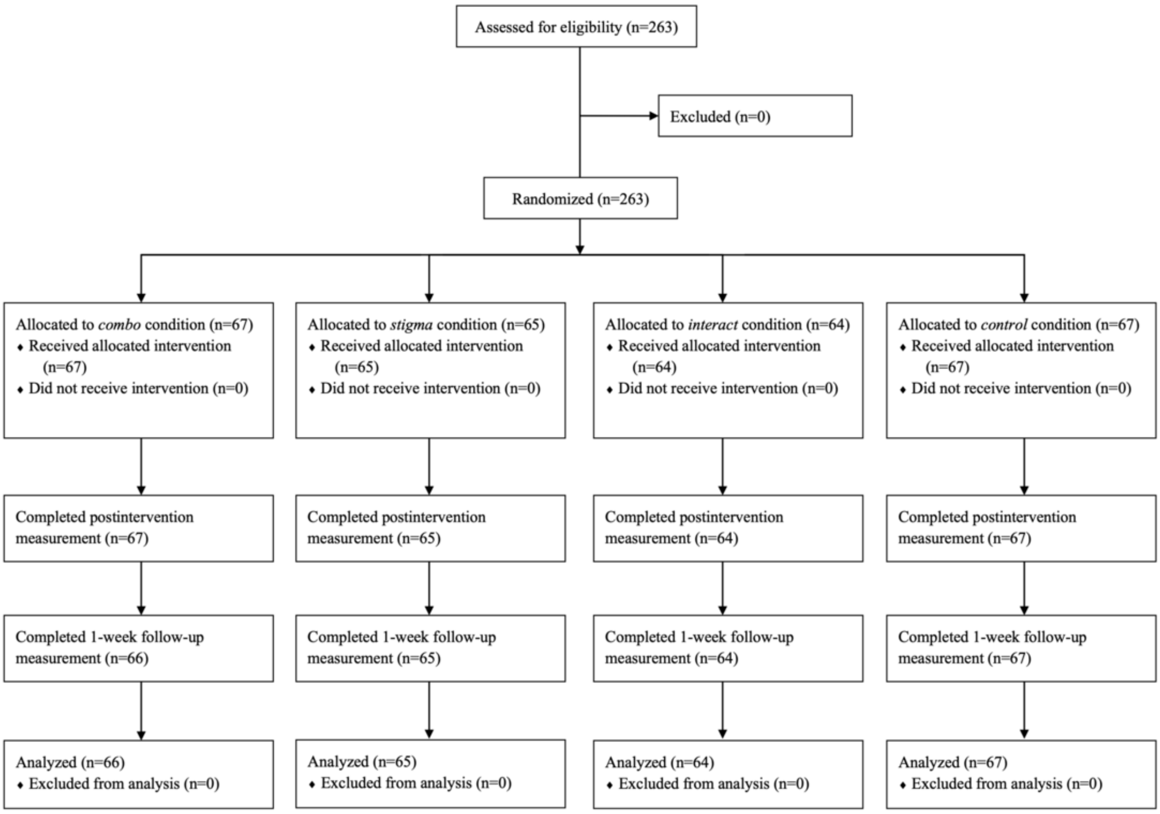


Table 1. Baseline characteristics across conditions.

Characteristics	*Combo* condition[a] (n=67)	*Stigma* condition[b] (n=65)	*Interact* condition[c] (n=64)	*Control* condition[d] (n=67)
Age (years), mean (SD)	22.49 (6.33)	24.49 (8.34)	21.89 (3.95)	21.39 (4.78)
Gender, n (%)				
Men	27 (40)	23 (35)	20 (31)	11 (16)
Women	40 (60)	42 (65)	44 (69)	56 (84)
Student: current educational level, n (%)				
Diploma, certificate, or associate degree	1 (2)	3 (5)	0 (0)	3 (5)
Bachelor's degree	52 (78)	42 (65)	48 (75)	55 (82)
Master's degree	4 (6)	6 (9)	9 (14)	2 (3)
Doctoral degree	0 (0)	1 (2)	0 (0)	1 (2)
Nonstudent: educational attainment, n (%)				
Secondary (form 1-6 or 7)	2 (3)	4 (6)	0 (0)	0 (0)
Diploma, certificate, or associate degree	3 (5)	0 (0)	0 (0)	1 (2)
Bachelor's degree	3 (5)	3 (5)	4 (6)	5 (8)
Master's degree	2 (3)	5 (8)	2 (3)	0 (0)
Doctoral degree	0 (0)	1 (2)	1 (2)	0 (0)
Religion, n (%)				
No religion	53 (79)	52 (80)	49 (77)	43 (64)
Buddhism	0 (0)	1 (2)	1 (2)	1 (2)
Catholicism	0 (0)	1 (2)	1 (2)	2 (3)
Christianity	13 (19)	9 (14)	13 (20)	20 (30)
Taoism	1 (2)	2 (3)	0 (0)	0 (0)
Others	0 (0)	0 (0)	0 (0)	1 (2)
Sexual orientation, n (%)				
Heterosexual	64 (96)	56 (86)	59 (92)	65 (97)
Homosexual	0 (0)	6 (9)	1 (2)	0 (0)
Bisexual	3 (5)	2 (3)	1 (2)	0 (0)
Pansexual	0 (0)	0 (0)	1 (2)	1 (2)
Others	0 (0)	1 (2)	2 (3)	1 (2)
Level of previous contact with people having a mental illness, mean (SD)	2.52 (1.26)	2.43 (1.00)	2.20 (0.93)	2.22 (1.00)

[a] Interactivity present; stigma content present.

[b] Interactivity absent; stigma content present.

[c] Interactivity present; stigma content absent.

[d] Interactivity absent; stigma content absent.

Mental Illness Stigma Measures

Public Stigma Toward People With Mental Illness

Results from the repeated measures ANOVA indicated a significant time×condition effect (*P*=.002; η^2=0.04), and a post hoc analysis was conducted. In the *combo* condition, public stigma significantly decreased from baseline to after the assessment (mean difference 0.61, 95% CI 0.49-0.74; *P*<.001; η^2=0.37), and the decrease was maintained at the 1-week follow-up (mean difference 0.53, 95% CI 0.37-0.69; *P*<.001; η^2=0.37). In the *stigma* condition, public stigma also significantly decreased from baseline to after the assessment (mean difference 0.42, 95% CI 0.30-0.55; *P*<.001; η^2=0.22), and the decrease was maintained at the 1-week follow-up (mean difference 0.34, 95% CI 0.18-0.50; *P*<.001; η^2=0.22). In the *interact* condition, public stigma significantly decreased from baseline to after the assessment (mean difference 0.14, 95% CI 0.02-0.26; *P*=.02; η^2=0.03), but the effect was not sustained at the 1-week follow-up (mean difference 0.12, 95% CI −0.04 to 0.28; *P*=.22; η^2=0.03). In the *control* condition, the effect was not significant from baseline to after the assessment (mean difference 0.07, 95% CI −0.06 to 0.20; *P*=.56; η^2=0.01) and from baseline to 1-week follow-up (mean difference 0.09, 95% CI −0.08 to 0.26; *P*=.57; η^2=0.01). In terms of mean difference values, the results indicated that the effect was the strongest in the *combo* condition, followed by the *stigma* and *interact* conditions. An additional post hoc analysis was carried out to compare *combo* and *stigma* conditions; the interaction effect between interactivity and stigma content was not significant (*P*=.09).

Microaggression

The results showed a significant time×condition effect (*P*<.001; η^2=0.06), and a post hoc analysis was carried out. Microaggression significantly decreased from baseline to after the assessment in both the *combo* (mean difference 0.34, 95% CI 0.25-0.42; *P*<.001; η^2=0.31) and *stigma* (mean difference 0.28, 95% CI 0.19-0.36; *P*<.001; η^2=0.24) conditions. The effects were sustained and strengthened at the 1-week follow-up in both conditions (*combo*: mean difference 0.39, 95% CI 0.29-0.49; *P*<.001; η^2=0.31; *stigma*: mean difference 0.33, 95% CI 0.23-0.43; *P*<.001; η^2=0.24). In the *interact* condition, the effect was not significant from baseline to after the assessment (mean difference 0.03, 95% CI −0.05 to 0.12; *P*=.99; η^2=0.01) and from baseline to 1-week follow-up (mean difference 0.06, 95% CI −0.04 to 0.16; *P*=.40; η^2=0.01). In the *control* condition, the effect was also not significant from baseline to after the assessment (mean difference 0.03, 95% CI −0.06 to 0.12; *P*=.99; η^2=0.01) and from baseline to 1-week follow-up (mean difference −0.04, 95% CI −0.15 to 0.07; *P*=.99; η^2=0.01). The results indicated that the effect of the *combo* condition was stronger than that of the *stigma* condition in terms of mean difference values. No significant interaction effect between interactivity and stigma content was found (*P*=.58) after performing an additional post hoc analysis to compare the *combo* and *stigma* conditions.

Social Distance From People With Mental Illness

The results showed a nonsignificant time×condition effect (*P*=.25; η^2=0.02). The additional post hoc analysis comparing the *combo* and *stigma* conditions showed no significant interaction effect between interactivity and stigma content (*P*=.46). The details of the repeated measures ANOVA are shown in Table 2.

Table 2. Imputed values of means and SDs across conditions.

	Combo condition[a] (n=66), mean (SD)	*Stigma* condition[b] (n=65), mean (SD)	*Interact* condition[c] (n=64), mean (SD)	*Control* condition[d] (n=67), mean (SD)
Public stigma				
Baseline	2.85 (0.66)	2.68 (0.75)	2.83 (0.72)	2.81 (0.72)
After the experiment	2.24 (0.66)	2.25 (0.73)	2.69 (0.80)	2.73 (0.75)
1-week follow-up	2.33 (0.78)	2.33 (0.74)	2.71 (0.74)	2.75 (0.72)
Microaggression				
Baseline	2.45 (0.37)	2.44 (0.36)	2.46 (0.32)	2.41 (0.38)
After the experiment	2.13 (0.45)	2.17 (0.44)	2.43 (0.37)	2.39 (0.42)
1-week follow-up	2.06 (0.43)	2.12 (0.49)	2.40 (0.33)	2.45 (0.40)
Social distance				
Baseline	2.39 (0.48)	2.28 (0.50)	2.35 (0.49)	2.38 (0.51)
After the experiment	1.98 (0.57)	1.97 (0.51)	2.36 (0.48)	2.35 (0.51)
1-week follow-up	2.13 (0.53)	2.07 (0.61)	2.41 (0.48)	2.40 (0.53)

[a] Interactivity present; stigma content present.

[b] Interactivity absent; stigma content present.

[c] Interactivity present; stigma content absent.

[d] Interactivity absent; stigma content absent.

Mediating Analysis

To compare the mediation effect of perceived autonomy and immersiveness between conditions with public stigma and microaggression, mediation analyses were performed by incorporating both perceived autonomy and immersiveness into PROCESS model 4. Table 3 shows the unstandardized and standardized factor loadings for the model. A mediation model of perceived autonomy and immersiveness between conditions with public stigma and microaggression is shown in Figure 2. Mediation analysis for social distance was not conducted, because no interaction effect was observed in the social distance across conditions.

We observed significant indirect effects of the *combo* (b=−0.19, bias-corrected and accelerated [BCa] CI −0.36 to −0.03), *stigma* (b=−0.15, BCa CI −0.29 to −0.02) and *interact* (b=−0.16, BCa CI −0.32 to −0.02) conditions on public stigma through perceived autonomy. The nonsignificant indirect effects of the *combo* (b=−0.13, BCa CI −0.30 to 0), *stigma* (b=−0.13, BCa CI −0.30 to 0), and *interact* (b=−0.07, BCa CI −0.16 to 0) conditions on public stigma through immersiveness were observed. The results showed that perceived autonomy was a significant mediator between the conditions and public stigma.

We observed nonsignificant indirect effects of the *combo* (b=0.07, BCa CI −0.03 to 0.17), *stigma* (b=0.05, BCa CI −0.03 to 0.14), and *interact* (b=0.06, BCa CI −0.03 to 0.15) conditions on microaggression through perceived autonomy. The indirect effects of the *combo* (b=−0.13, BCa CI −0.23 to −0.06), *stigma* (b=−0.13, BCa CI −0.23 to −0.05), and *interact* (b=−0.07, BCa CI −0.13 to −0.02) conditions on microaggression through immersiveness was significant. The results showed that immersiveness was a significant mediator between the conditions and microaggression.

Table 3. Unstandardized and standardized parameter estimates for the hypothesized model.

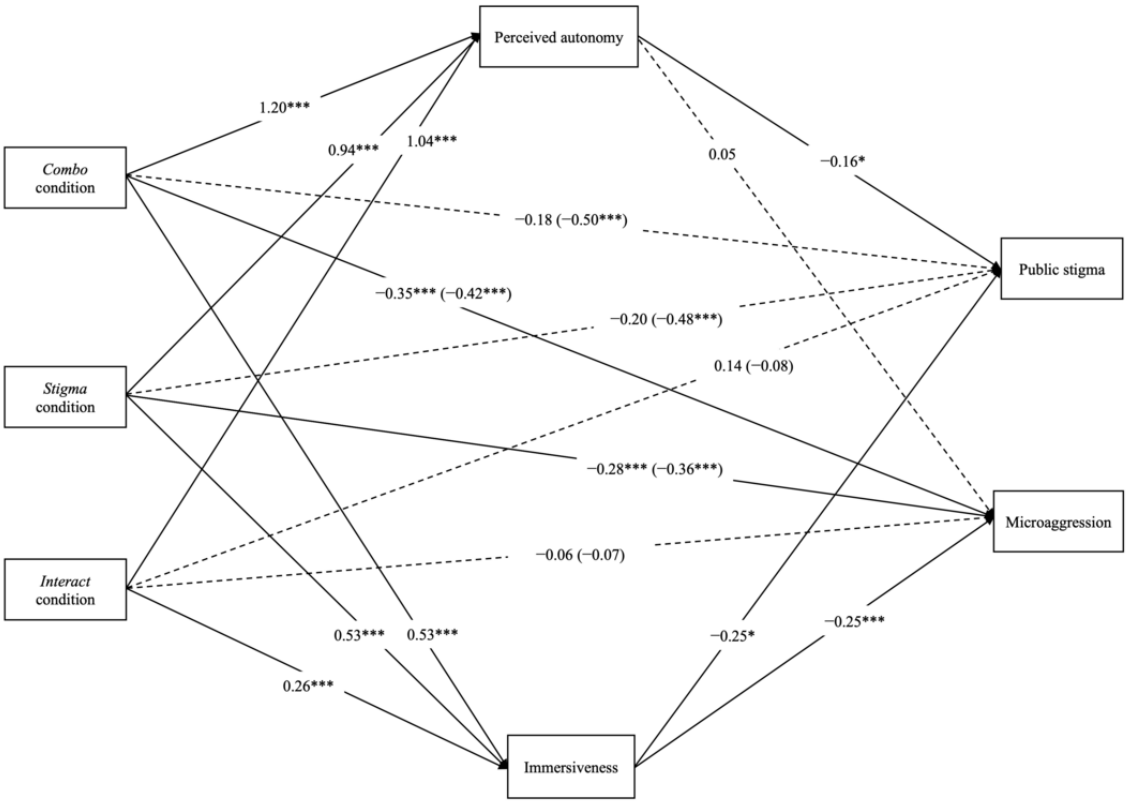
Parameter estimates (structural model)	Unstandardized B (SE)	Standardized β	*t* value (df)	*P* value
Combo[a]→perceived autonomy	1.20 (0.13)	1.42	9.53 (257)	<.001
Combo→immersiveness	0.53 (0.07)	1.21	7.81 (257)	<.001
Combo→public stigma	−0.18 (0.15)	−.23	−1.18 (255)	.24
Combo→microaggression	−0.35 (0.09)	−.78	−4.07 (255)	<.001
Stigma[b]→perceived autonomy	0.94 (0.13)	1.11	7.50 (257)	<.001
Stigma→immersiveness	0.53 (0.07)	1.20	7.80 (257)	<.001
Stigma→public stigma	−0.20 (0.15)	−.26	−1.37 (255)	.17
Stigma→microaggression	−0.28 (0.08)	−.61	−3.33 (255)	<.001
Interact[c]→perceived autonomy	1.04 (0.13)	1.23	8.30 (257)	<.001
Interact→immersiveness	0.26 (0.07)	.60	3.87 (257)	<.001
Interact→public stigma	0.14 (0.14)	.19	1.01 (255)	.31
Interact→microaggression	−0.06 (0.08)	−.14	−0.78 (255)	.44
Gender→perceived autonomy	0.25 (0.10)	.14	2.59 (257)	.01
Gender→immersiveness	0.07 (0.05)	.07	1.30 (257)	.19
Gender→public stigma	−0.24 (0.10)	−.14	−2.38 (255)	.02
Gender→microaggression	−0.09 (0.06)	−.09	−1.57 (255)	.12
Perceived autonomy→public stigma	−0.16 (0.07)	−.17	−2.27 (255)	.02
Perceived autonomy→microaggression	0.05 (0.04)	.10	1.39 (255)	.16
Immersiveness→public stigma	−0.25 (0.13)	−.14	−1.97 (255)	.05
Immersiveness→microaggression	−0.25 (0.07)	−.24	−3.46 (255)	<.001

[a] Interactivity present; stigma content present.

[b] Interactivity absent; stigma content present.

[c] Interactivity present; stigma content absent.

Figure 2. A mediation model of perceived autonomy and immersiveness among conditions. The numbers in brackets denote the total effect, solid lines indicate statistically significant paths, and dotted lines denote nonsignificant paths. Only the main variable is included in the figure for simplicity. *Combo* condition: interactivity present; stigma content present. *Interact* condition: interactivity present; stigma content absent. *Stigma* condition: interactivity absent; stigma content present. *P<.05, **P<.01, ***P<.001.



Discussion

Principal Findings

This study investigated the effect of internet-based storytelling programs with a combination of stigma content and interactivity on mental illness stigma reduction. Multiple forms of mental illness stigma were accounted for, including public stigma toward people with mental illness, microaggression, and social distance from people with mental illness.

The results supported our hypothesis that an internet-based storytelling program with a combination of stigma content and interactivity was able to significantly reduce public stigma and microaggression immediately after the experiment and at the 1-week follow-up assessment. Contrary to our hypothesis, an internet-based storytelling program with a combination of stigma content and interactivity did not significantly reduce the social distance from people with mental illness. In other words, the storytelling program was more effective in improving individuals' stigmatizing cognitions, sense of personal advocacy [4], and microaggressions centering around their everyday conversations and encounters in daily life [61] than in enhancing their willingness and intention to behaviorally interact with people with mental illness in various life domains [62].

Nevertheless, the results showed that an internet-based storytelling program with stigma content alone could also lead to a reduction in public stigma and microaggression. Comparing the significant effects elicited by the 2 different internet-based storytelling programs, the one with a combination of stigma content and interactivity produced a stronger effect after the assessment and at the 1-week follow-up assessment than that with stigma content only. These results were consistent with the findings of previous studies showing that internet-based stigma reduction interventions with a combination of interactivity and stigma content can lead to more effective stigma reduction [26]. The results reinforce the importance of correcting stigmatizing misperceptions in diminishing stigma [55,63]. Interestingly, the internet-based storytelling program with interactivity only was also found to reduce public stigma after the assessment although the effect could not be maintained after 1 week. This might support our assumption of a positive relationship between interactivity and positive affect and between positive affect and reduced prejudice [37,40]. Further studies are required to examine these relationships.

Moreover, although perceived autonomy could mediate the effect of conditions on public stigma, immersiveness could mediate the effect of conditions on microaggression. In previous findings, perceived autonomy was associated with the endorsement of the selected responses and actions and cognitive changes [46]. This might demonstrate that cognitive changes are essential for reducing public stigma. On the contrary, immersiveness could facilitate cognitive and affective changes [50-52], which might expedite people to have more intended changes in ameliorating everyday microaggressions. Social distance posited the focus on one's behavioral intention toward people with mental illness in various dimensions [62], which may have more long-term implications to their life domains (eg, friendship and employment); the short-term sense of choice and immersiveness in storytelling programs may not be sufficient to bring about changes in social distance.

In addition, this study showed that a storytelling program with stigmatized content but without interactivity could also enhance perceived autonomy and immersiveness. These findings align with the literature showing that stigma content with storytelling elements, which served as an internet-based contact with people with lived experiences, was effective in inducing immersiveness [24,28,30]. The inducement of perceived autonomy with the story was outstanding, whereas the effect might be due to the application of a conversational storyline, which allowed one to feel like interacting with the protagonist even without choosing responses and actions. Thus, storytelling in the form of conversations alone might already provide an effective means for participants to immerse themselves and feel a sense of agency.

The superior effect of an internet-based storytelling program with a combination of interactivity and stigma content over the one with stigma content only was not explained by the interaction effect between interactivity and stigma content. Stigma content was a more important criterion than interactivity in inducing stigma reduction. The addition of interactivity was only supplementary to boosting the stigma reduction effect. The results confirmed previous suggestions that enhanced motivation through perceived autonomy and immersiveness would lead to enhanced stigma reduction [43,44]. This study provides empirical support for the rationale behind the use of internet-based storytelling programs, especially for those with stigma-related stories and interactivity.

Limitations and Future Directions

This study has some limitations that deserve attention. First, our sample mainly consisted of young university students. These findings may not be generalizable to other populations. In social marketing, segmentation of our target population is essential. This study showed that internet-based antistigma storytelling programs with interactivity may be an effective tool in reducing mental illness stigma for young, educated people in the community who are comfortable and skillful in accessing information over the internet. Furthermore, because of the homogeneous nature of our sample, moderation analysis was not performed. Future studies should explore possible moderators of the effect of internet-based stigma reduction interventions, such as gender, age, and education level.

Second, this study did not include a long-term follow-up. Only a 1-week follow-up assessment was included to investigate whether the mediation effect could be sustained for a week to draw possible mediating mechanisms behind the effect of internet-based stigma reduction interventions. It could capture the short-term effect of the intervention but was incapable of predicting the long-term maintenance effect of stigma reduction [64]. However, the long-term effect must be tested to provide insights into developing sustainable and effective stigma reduction interventions. It is unknown whether an internet-based storytelling program with a combination of interactivity and stigma content could impose a more prolonged stigma reduction effect when than a program with stigma content only. To examine the stigma reduction effect of various internet-based storytelling programs over time, future research should lengthen its follow-up.

Third, this study lacked behavioral measures but solely used self-report measures that tapped into the thoughts, affect, and behavioral intentions of the participants. Although the measures used have been empirically validated, future studies aiming to examine stigma change should also use behavioral measures, such as the inclusion of offering payment for completing questionnaires and the option of donating the payment to a mental health charity, in addition to self-report measures to bolster the findings.

The necessary intensity of storytelling and interactivity to strengthen the stigma reduction effect should also be investigated. Research has shown that placing individuals in a highly immersive environment might lead to worsening attitudes [65], which varies according to the degree of identification with the embodied target [66]. Therefore, the degree of immersiveness elicited should be considered. In addition, future studies can explore how individual differences, such as the baseline level of empathy, may influence intervention benefits, as internet-based interventions with different adaptations might cater to different segments of the population. People with different levels of dispositional empathy may have variable receptivity to internet-based storytelling programs, and matching their styles with this approach may maximize the outcomes.

Finally, this study suggests that increased perceived autonomy and immersiveness could strengthen the stigma reduction effect in internet-based storytelling programs with interactivity and stigma content. With the advancement in virtual reality technology and the sense of embodiment and story transportation being found to mediate public stigma reduction [67], future studies can consider comparing virtual reality–based and internet-based stigma reduction programs to shed light on which delivery methods are more effective and cost-effective in reducing mental illness stigma.

Conclusions

In sum, this study showed that internet-based storytelling programs with a combination of interactivity and stigma content are effective in reducing mental illness stigma, and perceived autonomy and immersiveness are significant mechanisms in the stigma reduction process. The findings of this study are encouraging and support the possible mechanisms behind the effects of internet-based storytelling programs. Furthermore, the study upholds the gravity of stigma content, as the strength of interactivity in reducing mental illness stigma can only be manifested in the presence of stigma content. To leverage the power of technology in reducing mental illness stigma, it is paramount for antistigma campaigns to incorporate these active ingredients into the design of antistigma interventions to create low-cost, effective, and scalable internet-based stigma reduction programs.

Acknowledgments

The Amazing Adventure Against Stigma website is part of the Jockey Club TourHeart Project funded by the Hong Kong Jockey Club Charities Trust (reference number: 2017/0070).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the four web pages.
[PDF File (Adobe PDF File), 696 KB - jmir_v24i8e37973_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1361 KB - jmir_v24i8e37973_app2.pdf]

References

1. Sheehan L, Corrigan P. Stigma of disease and its impact on health. In: Sweeny K, Robbins ML, Cohen LM, editors. The Wiley Encyclopedia of Health Psychology. Hoboken, NJ, USA: John Wiley & Sons; 2020:57-65.
2. Sickel AE, Seacat JD, Nabors NA. Mental health stigma: impact on mental health treatment attitudes and physical health. J Health Psychol 2019 Apr;24(5):586-599. [doi: 10.1177/1359105316681430] [Medline: 27909162]
3. Stigma and discrimination. World Health Organization. 2013. URL: https://www.who.int/southeastasia/outbreaks-and-emergencies/covid-19/What-can-we-do-to-keep-safe/stigma-and-discrimination [accessed 2021-04-05]
4. Corrigan PW, Morris SB, Michaels PJ, Rafacz JD, Rüsch N. Challenging the public stigma of mental illness: a meta-analysis of outcome studies. Psychiatr Serv 2012 Oct;63(10):963-973. [doi: 10.1176/appi.ps.201100529] [Medline: 23032675]
5. Corrigan PW, Penn DL. Lessons from social psychology on discrediting psychiatric stigma. Stigma Health 2015 Aug;1(S):2-17. [doi: 10.1037/2376-6972.1.S.2]
6. Walsh DA, Foster JL. A call to action. A critical review of mental health related anti-stigma campaigns. Front Public Health 2021 Jan 8;8:569539 [FREE Full text] [doi: 10.3389/fpubh.2020.569539] [Medline: 33490010]
7. Michaels PJ, Corrigan PW, Buchholz B, Brown J, Arthur T, Netter C, et al. Changing stigma through a consumer-based stigma reduction program. Community Ment Health J 2014 May;50(4):395-401. [doi: 10.1007/s10597-013-9628-0] [Medline: 23760975]
8. Chan JY, Mak WW, Law LS. Combining education and video-based contact to reduce stigma of mental illness: "The Same or Not the Same" anti-stigma program for secondary schools in Hong Kong. Soc Sci Med 2009 Apr;68(8):1521-1526. [doi: 10.1016/j.socscimed.2009.02.016] [Medline: 19282079]
9. Clement S, van Nieuwenhuizen A, Kassam A, Flach C, Lazarus A, de Castro M, et al. Filmed v. live social contact interventions to reduce stigma: randomised controlled trial. Br J Psychiatry 2012 Jul;201(1):57-64. [doi: 10.1192/bjp.bp.111.093120] [Medline: 22157800]
10. Corrigan PW, Larson J, Sells M, Niessen N, Watson AC. Will filmed presentations of education and contact diminish mental illness stigma? Community Ment Health J 2007 Apr;43(2):171-181. [doi: 10.1007/s10597-006-9061-8] [Medline: 16988883]
11. Gronholm PC, Henderson C, Deb T, Thornicroft G. Interventions to reduce discrimination and stigma: the state of the art. Soc Psychiatry Psychiatr Epidemiol 2017 Mar;52(3):249-258 [FREE Full text] [doi: 10.1007/s00127-017-1341-9] [Medline: 28144713]
12. Han M, Cha R, Lee HA, Lee SE. Mental-illness stigma among Korean immigrants: role of culture and destigmatization strategies. Asian Am J Psychol 2017;8(2):134-141. [doi: 10.1037/aap0000074]
13. Naslund JA, Deng D. Addressing mental health stigma in low-income and middle-income countries: a new frontier for digital mental health. Ethics Med Public Health 2021 Dec;19:100719. [doi: 10.1016/j.jemep.2021.100719] [Medline: 35083375]
14. Topuzoglu A. Intervention to mental illness stigmatization in the community. Marmara Med J 2021 May 31;34(2):17-22. [doi: 10.5472/marumj.897870]
15. Collins RL, Wong EC, Roth E, Cerully JL, Marks J. Changes in mental illness stigma in California during the statewide stigma and discrimination reduction initiative. Rand Health Q 2015 Nov 30;5(2):10 [FREE Full text] [Medline: 28083386]
16. Hanisch SE, Birner UW, Oberhauser C, Nowak D, Sabariego C. Development and evaluation of digital game-based training for managers to promote employee mental health and reduce mental illness stigma at work: quasi-experimental study of

program effectiveness. JMIR Ment Health 2017 Aug 04;4(3):e31 [FREE Full text] [doi: 10.2196/mental.7600] [Medline: 28778839]

17. Tippin GK, Maranzan KA. Efficacy of a photovoice‐based video as an online mental illness anti‐stigma intervention and the role of empathy in audience response: a randomized controlled trial. J Appl Soc Psychol 2019 Mar 28;49(6):381-394. [doi: 10.1111/jasp.12590]
18. Utter K, Waineo E, Bell CM, Quaal HL, Levine DL. Instagram as a window to societal perspective on mental health, gender, and race: observational pilot study. JMIR Ment Health 2020 Oct 27;7(10):e19171 [FREE Full text] [doi: 10.2196/19171] [Medline: 33107831]
19. Williams T, Salmonsen J, Mizock L. Webinar Intervention Research to Reduce Stigma Using a Pre–Post Research Design. Thousand Oaks, CA, USA: Sage Publications; 2021.
20. Griffiths KM, Carron-Arthur B, Parsons A, Reid R. Effectiveness of programs for reducing the stigma associated with mental disorders. A meta-analysis of randomized controlled trials. World Psychiatry 2014 Jun;13(2):161-175 [FREE Full text] [doi: 10.1002/wps.20129] [Medline: 24890069]
21. Mehta N, Clement S, Marcus E, Stona AC, Bezborodovs N, Evans-Lacko S, et al. Evidence for effective interventions to reduce mental health-related stigma and discrimination in the medium and long term: systematic review. Br J Psychiatry 2015 Nov;207(5):377-384 [FREE Full text] [doi: 10.1192/bjp.bp.114.151944] [Medline: 26527664]
22. Potts LC, Bakolis I, Deb T, Lempp H, Vince T, Benbow Y, INDIGO READ Study Group. Anti-stigma training and positive changes in mental illness stigma outcomes in medical students in ten countries: a mediation analysis on pathways via empathy development and anxiety reduction. Soc Psychiatry Psychiatr Epidemiol (forthcoming) 2022 Apr 22. [doi: 10.1007/s00127-022-02284-0] [Medline: 35451604]
23. Pettigrew TF, Tropp LR. How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. Eur J Soc Psychol 2008 Sep;38(6):922-934. [doi: 10.1002/ejsp.504]
24. Davidson T, Moreland A, Bunnell BE, Winkelmann J, Hamblen JL, Ruggiero KJ. Reducing stigma in mental health through digital storytelling. In: Research Anthology on Mental Health Stigma, Education, and Treatment (3 Volumes). Hershey, PA, USA: IGI Global; 2021:909-919.
25. Ferchaud A, Seibert J, Sellers N, Escobar Salazar N. Reducing mental health stigma through identification with video game avatars with mental illness. Front Psychol 2020 Sep 9;11:2240 [FREE Full text] [doi: 10.3389/fpsyg.2020.02240] [Medline: 33013585]
26. Kim H, Stout PA. The effects of interactivity on information processing and attitude change: implications for mental health stigma. Health Commun 2010 Mar;25(2):142-154. [doi: 10.1080/10410230903544936] [Medline: 20390680]
27. Kirschner B, Goetzl M, Curtin L. Mental health stigma among college students: test of an interactive online intervention. J Am Coll Health (forthcoming) 2020 Oct 13:1-8. [doi: 10.1080/07448481.2020.1826492] [Medline: 33048656]
28. Kosyluk K, Marshall J, Conner K, Macias DR, Macias S, Michelle Beekman B, et al. Challenging the stigma of mental illness through creative storytelling: a randomized controlled trial of this is my brave. Community Ment Health J 2021 Jan;57(1):144-152. [doi: 10.1007/s10597-020-00625-4] [Medline: 32388690]
29. Friedrich B, Evans-Lacko S, London J, Rhydderch D, Henderson C, Thornicroft G. Anti-stigma training for medical students: the Education Not Discrimination project. Br J Psychiatry Suppl 2013 Apr;55:s89-s94. [doi: 10.1192/bjp.bp.112.114017] [Medline: 23553700]
30. Chew QH, Sim K. Psychiatry teaching amongst medical undergraduates: stories that matter and mediators of better learning outcome. Postgrad Med 2020 Sep;132(7):590-594. [doi: 10.1080/00325481.2020.1774178] [Medline: 32442058]
31. Corrigan PW, Kosyluk KA. Erasing the stigma: where science meets advocacy. Basic Appl Soc Psych 2013 Feb 04;35(1):131-140. [doi: 10.1080/01973533.2012.746598]
32. Corrigan P, Kosyluk K. Mental illness stigma: types, constructs, and vehicles for change. In: Corrigan PW, editor. The Stigma of Disease and Disability: Understanding Causes and Overcoming Injustices. Washington, DC, USA: American Psychological Association; 2014:35-56.
33. Warne T, McAndrew S. Passive patient or engaged expert? Using a Ptolemaic approach to enhance mental health nurse education and practice. Int J Ment Health Nurs 2007 Aug;16(4):224-229. [doi: 10.1111/j.1447-0349.2007.00471.x] [Medline: 17635621]
34. Song D, Lim HJ, Chung YJ. The stigma of mental illness and the way of destigmatization: the effects of interactivity and self-construal. World Acad Sci Eng Technol 2011;60:1286-1293.
35. Johnson GJ, Bruner II GC, Kumar A. Interactivity and its facets revisited: theory and empirical test. J Advert 2006 Dec;35(4):35-52. [doi: 10.2753/joa0091-3367350403]
36. Vallée-Tourangeau G, Vallée-Tourangeau F. Cognition beyond the classical information processing model: cognitive interactivity and the systemic thinking model (SysTM). In: Cowley SJ, Vallée-Tourangeau F, editors. Cognition Beyond the Brain: Computation, Interactivity and Human Artifice. Cham, Switzerland: Springer; 2017:133-154.
37. Cairncross S, Mannion M. Interactive multimedia and learning: realizing the benefits. Innov Educ Teach Int 2001 Jan;38(2):156-164. [doi: 10.1080/14703290110035428]

38. Liu Y, Shrum LJ. What is interactivity and is it always such a good thing? Implications of definition, person, and situation for the influence of interactivity on advertising effectiveness. J Advert 2002 Dec;31(4):53-64. [doi: 10.1080/00913367.2002.10673685]
39. MacKenzie SB, Lutz RJ, Belch GE. The role of attitude toward the ad as a mediator of advertising effectiveness: a test of competing explanations. J Market Res 1986 May 1;23(2):130-143. [doi: 10.1177/002224378602300205]
40. Tenenbaum HR, Capelos T, Lorimer J, Stocks T. Positive thinking elevates tolerance: experimental effects of happiness on adolescents' attitudes toward asylum seekers. Clin Child Psychol Psychiatry 2018 Apr;23(2):346-357. [doi: 10.1177/1359104518755217] [Medline: 29566555]
41. McMillan SJ. Exploring models of interactivity from multiple research traditions: users, documents, and systems. In: Lievrouw LA, Livingstone S, editors. Handbook of New Media: Social Shaping and Consequences of ICTs. Thousand Oaks, CA, USA: Sage Publications; 2002:163-182.
42. Sundar SS. Social psychology of interactivity in human-website interaction. In: Joinson A, McKenna K, Postmes T, Reips UD, editors. Oxford Handbook of Internet Psychology. Oxford, UK: Oxford University Press; 2007:89-102.
43. Griffin BW. Perceived autonomy support, intrinsic motivation, and student ratings of instruction. Stud Educ Eval 2016 Dec;51:116-125. [doi: 10.1016/j.stueduc.2016.10.007]
44. Sobitan A, Vlachos P. Immersive event experience and attendee motivation: a quantitative analysis using sensory, localisation, and participatory factors. J Policy Res Tourism Leisure Events 2020 Feb 09;12(3):437-456. [doi: 10.1080/19407963.2020.1721638]
45. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
46. Deci EL, Ryan RM. Facilitating optimal motivation and psychological well-being across life's domains. Can Psychol 2008 Feb;49(1):14-23. [doi: 10.1037/0708-5591.49.1.14]
47. Dede C. Immersive interfaces for engagement and learning. Science 2009 Jan 02;323(5910):66-69. [doi: 10.1126/science.1167311] [Medline: 19119219]
48. Tal-Or N, Cohen J. Understanding audience involvement: conceptualizing and manipulating identification and transportation. Poetics 2010 Aug;38(4):402-418. [doi: 10.1016/j.poetic.2010.05.004]
49. Appel M, Richter T. Transportation and need for affect in narrative persuasion: a mediated moderation model. Media Psychol 2010 May 28;13(2):101-135. [doi: 10.1080/15213261003799847]
50. Green MC, Brock TC, Kaufman GF. Understanding media enjoyment: the role of transportation into narrative worlds. Commun Theory 2004 Nov;14(4):311-327 [FREE Full text] [doi: 10.1111/j.1468-2885.2004.tb00317.x]
51. Kaufman GF, Libby LK. Changing beliefs and behavior through experience-taking. J Pers Soc Psychol 2012 Jul;103(1):1-19. [doi: 10.1037/a0027525] [Medline: 22448888]
52. Oliver MB, Dillard JP, Bae K, Tamul DJ. The effect of narrative news format on empathy for stigmatized groups. Journal Mass Commun Q 2012 Mar 14;89(2):205-224. [doi: 10.1177/1077699012439020]
53. Simmons L, Jones T, Bradley E. Reducing mental health stigma: the relationship between knowledge and attitude change. Eur J Ment Health 2017;12(1):25-40. [doi: 10.5708/ejmh.12.2017.1.2]
54. Amazing Adventure against stigma. The Hong Kong Jockey Club Charities Trust. URL: https://antistigma.psy.cuhk.edu.hk/ [accessed 2022-08-03]
55. Storbeck J, Clore GL. Affective arousal as information: how affective arousal influences judgments, learning, and memory. Soc Personal Psychol Compass 2008 Sep 01;2(5):1824-1843 [FREE Full text] [doi: 10.1111/j.1751-9004.2008.00138.x] [Medline: 25067943]
56. Holmes EP, Corrigan PW, Williams P, Canar J, Kubiak MA. Changing attitudes about schizophrenia. Schizophr Bull 1999;25(3):447-456. [doi: 10.1093/oxfordjournals.schbul.a033392] [Medline: 10478780]
57. Mak WW, Chong ES, Wong CC. Beyond attributions: understanding public stigma of mental illness with the common sense model. Am J Orthopsychiatry 2014 Mar;84(2):173-181. [doi: 10.1037/h0099373] [Medline: 24826933]
58. Gonzales L, Davidoff KC, DeLuca JS, Yanos PT. The mental illness microaggressions scale-perpetrator version (MIMS-P): reliability and validity. Psychiatry Res 2015 Sep 30;229(1-2):120-125. [doi: 10.1016/j.psychres.2015.07.061] [Medline: 26233830]
59. Sheldon KM. Creativity and self-determination in personality. Creativity Res J 1995 Jan;8(1):25-36. [doi: 10.1207/s15326934crj0801_3]
60. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 2000 Nov;79(5):701-721. [doi: 10.1037/0022-3514.79.5.701]
61. Gonzales L, Davidoff KC, Nadal KL, Yanos PT. Microaggressions experienced by persons with mental illnesses: an exploratory study. Psychiatr Rehabil J 2015 Sep;38(3):234-241. [doi: 10.1037/prj0000096] [Medline: 25402611]
62. Corrigan PW, Edwards AB, Green A, Diwan SL, Penn DL. Prejudice, social distance, and familiarity with mental illness. Schizophr Bull 2001;27(2):219-225. [doi: 10.1093/oxfordjournals.schbul.a006868] [Medline: 11354589]
63. Penn DL, Chamberlin C, Mueser KT. The effects of a documentary film about schizophrenia on psychiatric stigma. Schizophr Bull 2003;29(2):383-391. [doi: 10.1093/oxfordjournals.schbul.a007012] [Medline: 14552511]

64. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar 12;387(10023):1123-1132. [doi: 10.1016/S0140-6736(15)00298-6] [Medline: 26410341]
65. Tassinari M, Aulbach MB, Jasinskaja-Lahti I. The use of virtual reality in studying prejudice and its reduction: a systematic review. PLoS One 2022 Jul 14;17(7):e0270748 [FREE Full text] [doi: 10.1371/journal.pone.0270748] [Medline: 35834584]
66. Starr CR, Anderson BR, Green KA. "I'm a computer scientist!": virtual reality experience influences stereotype threat and STEM motivation among undergraduate women. J Sci Educ Technol 2019 May 15;28(5):493-507. [doi: 10.1007/s10956-019-09781-z]
67. Yuen AS, Mak WW. The effects of immersive virtual reality in reducing public stigma of mental illness in the university population of Hong Kong: randomized controlled trial. J Med Internet Res 2021 Jul 14;23(7):e23683 [FREE Full text] [doi: 10.2196/23683] [Medline: 34259636]

Abbreviations

BCa: bias-corrected and accelerated
CONSORT: Consolidated Standards of Reporting Trials

Edited by R Kukafka; submitted 15.03.22; peer-reviewed by C Henderson, M Heilemann; comments to author 26.04.22; revised version received 21.06.22; accepted 18.07.22; published 12.08.22.

Please cite as:
Fong THC, Mak WWS
The Effects of Internet-Based Storytelling Programs (Amazing Adventure Against Stigma) in Reducing Mental Illness Stigma With Mediation by Interactivity and Stigma Content: Randomized Controlled Trial
J Med Internet Res 2022;24(8):e37973
URL: https://www.jmir.org/2022/8/e37973
doi:10.2196/37973
PMID:35969460

Original Paper

Analyzing the Impact of Mobile App Engagement on Mental Health Outcomes: Secondary Analysis of the Unwinding Anxiety Program

William Nardi[1], MPH; Alexandra Roy[1], BA; Shira Dunsiger[1], PhD; Judson Brewer[1], MD, PhD

[1] Department of Behavioral and Social Sciences, Brown University, Providence, RI, United States

Corresponding Author:
William Nardi, MPH
Department of Behavioral and Social Sciences
Brown University
121 South Main Street
Providence, RI, 02903
United States
Phone: 1 8609649144
Email: william_nardi@brown.edu

Abstract

Background: App-based interventions provide a promising avenue for mitigating the burden on mental health services by complimenting therapist-led treatments for anxiety. However, it remains unclear how specific systems' use of app features may be associated with changes in mental health outcomes (eg, anxiety and worry).

Objective: This study was a secondary analysis of engagement data from a stage 1 randomized controlled trial testing the impact of the Unwinding Anxiety mobile app among adults with generalized anxiety disorder. The aims of this study were 2-fold: to investigate whether higher microengagement with the primary intervention feature (ie, educational modules) is associated with positive changes in mental health outcomes at 2 months (ie, anxiety, worry, interoceptive awareness, and emotional reactivity) and to investigate whether the use of adjunctive app features is also associated with changes in mental health outcomes.

Methods: We analyzed the intervention group during the stage 1 trial of the Unwinding Anxiety mobile app. The total use of specific mobile app features and the use specific to each feature were calculated. We used multivariate linear models with a priori significance of α=.05 to investigate the impact of cumulative app use on anxiety, worry, interoceptive awareness, and emotional regulation at 2 months, controlling for baseline scores, age, and education level in all models. Significant relationships between system use metrics and baseline participant characteristics were assessed for differences in use groupings using between-group testing (ie, 2-tailed *t* tests for continuous data and chi-square analyses for categorical data).

Results: The sample was primarily female (25/27, 93%), and the average age was 42.9 (SD 15.6) years. Educational module completion, the central intervention component, averaged 20.2 (SD 11.4) modules out of 32 for the total sample. Multivariate models revealed that completing >75% of the program was associated with an average 22.6-point increase in interoceptive awareness (*b*=22.6; SE 8.32; *P*=.01; 95% CI 5.3-39.8) and an 11.6-point decrease in worry (*b*=−11.6; SE 4.12; *P*=.01; 95% CI −20.2 to −3.1). In addition, a single log unit change in the total number of meditations was associated with a 0.62-point reduction in the Generalized Anxiety Disorder-7 scale scores (*b*=0.62; SE 0.27; *P*=.005; 95% CI −1.2 to −0.6), whereas a single log unit use of the stress meter was associated with an average of a 0.5-point increase in emotional regulation scores (Five Facet Mindfulness Questionnaire; *b*=0.5; SE 0.21; *P*=.03; 95% CI 0.1-0.9).

Conclusions: This study offers a clearer understanding of the impact of engagement with app features on broader engagement with the health outcomes of interest. This study highlights the importance of comprehensive investigations of engagement during the development of evidence-based mobile apps.

(J Med Internet Res 2022;24(8):e33696) doi:10.2196/33696

KEYWORDS

anxiety; worry; engagement; mobile app; mental health; mobile phone

Introduction

Background

As anxiety disorders (ADs) increase worldwide [1-5], app-based interventions offer evidence-based, high-fidelity treatment options [6] with incredible potential and comparable effects with traditional treatment models in efficacy trials [7]. Several recent meta-analyses that assessed app-based interventions for anxiety have indicated a growing evidence base [7-9]. A recent 2021 Nature review (k=22) [10] found significant but small effect sizes on anxiety symptomatology (Hedges g=0.2888; P<.001). These interventions can fill an important treatment gap, mitigate the growing burden on providers [11,12], and reduce barriers to care (eg, time, cost, and stigma) [13]. However, app-based interventions for AD often have low levels of user engagement, which reduces their efficacy and inhibits broader health care implementation [14]. To capitalize on the benefits of these interventions, engagement needs to be effectively understood and analyzed [10]. Specifically, research identifying the optimal levels of engagement with app-based interventions and the effects of various app features on anxiety-related outcomes is critical to realizing their full potential as treatments for AD [15-17].

To date, research on engagement remains limited and is often not assessed, as app-based interventions are still largely analyzed using conventional methods (eg, intent to treat) [18]. This method evaluates the effect of *assignment to treatment* (ie, the effect of being randomized to a group) instead of the direct effect associated with the app features that were actually used during the intervention [18]. Analyzing which features the user engages with is critical to designing personalized, effective interventions [14] and allows researchers to understand which in-app features drive changes in health outcomes [14,19]. Understanding engagement is also important for app development; specifically, it contributes fundamental information toward identifying the optimal intervention dose [17], streamlining inefficiencies [20], and tailoring apps to specific clinical populations [21]. Recognizing the importance of information on engagement, research is increasingly prioritizing the investigation of the relationship between engagement and improvements in mental health outcomes [10,16,22,23]. Encouragingly, in one of the first meta-analyses examining the effect of engagement on health outcomes, Gan et al [24] (k=25) found significant moderate improvements in postintervention symptomatology for participants categorized as having higher engagement than participants with lower levels of engagement (Hedges g=0.40; SE 0.16; 95% CI 0.097-0.705; P=.01). Although only 5 studies were designed to target AD, the effects indicated a significant positive association between engagement and anxiety (r=0.33; 95% CI 0.24-0.41; P<.001) [24]. This work is promising; however, the field remains preliminary and requires additional research to enhance the growing understanding of the associations between engagement with app-based interventions and anxiety [25-27].

Inconsistency in the definition and measurement of engagement is an ongoing challenge associated with analyzing the construct within this promising preliminary research [24]. Engagement can be broadly conceptualized as "(1) the extent (eg, frequency, duration) of usage; and (2) a subjective experience characterized by affect, attention, and interest" [28], as it relates to subsequent changes in the targeted health behavior (eg, anxiety and depression) [29,30]. However, different industries (eg, psychology and marketing) have historically focused on specific parts of this definition (eg, in-app use vs levels of engagement in the targeted health behavior), often failing to fully capture an understanding of the complex relationships that make up engagement [29,30]. To improve our understanding of app-based interventions and health outcomes, it is critical to understand the relationship between what is used during the intervention and subsequent changes in health outcomes [31]. One model proposed by Cole-Lewis et al [31] addresses this relationship and posits that engagement is multifaceted, encompassing multiple definitions from various disciplines. The model links both system-level and behavior-level engagement [31]. They defined engagement as a multidimensional construct encompassing a user's interactions with app features that influence specific behavioral determinants, resulting in increased engagement in the targeted health behavior [31]. The model posits that the use of in-app features (ie, microlevel engagement) is directly associated with changes in the desired health outcome (ie, macrolevel engagement). Understanding both microlevel and macrolevel measurements is necessary to evaluate the effects of app-based interventions on improvements in mental health symptomatology [31].

Applying this model of engagement, we conducted a secondary analysis using multivariate regression models to examine the associations between in-app engagement (ie, microengagement) and mental health outcomes (ie, macroengagement) for participants with generalized AD (GAD) using a targeted mobile app called Unwinding Anxiety (UA) [32]. Data were collected from the intervention group in a recently published randomized controlled trial (RCT) that tested UA versus treatment as usual [32]. The UA app is a theory-driven, multifaceted app that comprises both guided (ie, educational modules) and unguided features (eg, meditations and ecological check-ins) targeting novel reinforcement learning constructs [32,33]. The results from the primary RCT (N=65) were promising, with participants in the UA group reporting a median reduction in anxiety scores of 8.5 (IQR 6.5; P<.001) and the treatment as usual group reporting a median reduction of 1 (IQR 5; P=.01), representing a 67% versus 14% reduction at the 2-month follow-up [32].

Objectives

The aims of this study were 2-fold: (1) to investigate whether higher microengagement with the primary intervention feature (ie, educational modules) is associated with positive changes in mental health outcomes at 2 months (ie, anxiety, worry, interoceptive awareness, and emotional reactivity) and (2) to investigate whether the use of adjunctive app features is also associated with changes in mental health outcomes. We hypothesized that higher levels of microengagement with educational modules would be associated with significant changes in outcomes, which is consistent with improved mental health.

Methods

Overview

This study is a secondary data analysis of a previously described stage 1 parallel-group RCT [32]. As this secondary analysis is interested in system use data and their association with study outcomes, data were taken from the intervention arm only, and study procedures relevant to analyzing the intervention group are summarized in the following sections [32]. Roy et al [32] provide detailed information on the study design, procedures, and the results of the randomized trial.

Ethics Approval

The primary trial was registered at Clinicaltrials.gov (NCT0368472), and the Brown University Institutional Review Board approved the study procedures (reference number PV4802) [32].

Study Procedures

The data used for these analyses were from the baseline and 2-month time points for the intervention group only [32]. The inclusion criteria were as follows: (1) score ≥10 on the Generalized Anxiety Disorder 7-item scale (GAD-7), (2) owning a smartphone, (3) willingness to receive check-in calls, and (4) aged ≥18 years [32]. Participants were excluded from the study if they reported (1) dose changes of any psychoactive medication in the previous 2 months; (2) needed use of benzodiazepines and hypnotic sleep aids; (3) a history or current diagnosis of bipolar, schizophrenia or schizoaffective, or another psychotic disorder; (4) a significant medical condition that would affect the ability to complete study tasks; (5) cohabitation with someone already enrolled in the study; and (6) having a previous history of using other related apps, specifically Eat Right Now or Craving to Quit, which use similar reinforcement processes to UA [33-35]. Participants were recruited using social media largely through Facebook advertisements.

Eligible participants underwent informed consent procedures before enrolling in the study [32]. After enrollment, participants completed an in-person interview using the Mini International Neuropsychiatric Interview (MINI) International Neuropsychiatric Interview to confirm a diagnosis of GAD along with the assessment of other potential comorbid disorders (eg, depression, obsessive-compulsive disorder, and posttraumatic stress disorder) [32]. The participants were then asked to complete a web-based questionnaire using Qualtrics. Follow-up questionnaires were administered 2 months from treatment initiation using personalized email links specific to each participant's unique identification number [32].

Intervention

UA is an app-based intervention comprising educational modules that are considered the primary intervention features, consistent with recommendations from recent meta-analyses. The modules comprised instructional psychoeducational videos (5-15 minutes per day) teaching reinforcement learning concepts (Table 1). Modules are locked until the previous module is completed; however, participants can return to any of the already completed modules for review.

In addition, the app offers unguided adjunctive features divided into 2 categories: ecological features designed to synergize with skills learned in the modules and meditation practices. The ecological features included physiological check-ins and 2 types of stress evaluation: a meter that evaluates the strength and reason for stress or anxiety and a stress test that assists in familiarizing participants with practicing curiosity regarding stress or anxiety using interoceptive awareness skills learned in the program. Detailed descriptions of these features can be found in Table 1, a visual depiction of the main dashboard is shown in Figure 1 and an example of an adjunctive ecological feature is provided in Figure 2. The example is the psychological check-in feature, in which a participant is asked first to identify their current emotional state from an initial list and then rate their anxiety level at the moment, and it ends by offering a recommendation for a short practice to return to present moment awareness depicted from left to right.

For the adjunctive meditation features, participants had access to a series of 3 practices that they were encouraged but not mandated to use: Resting in Awareness, a body scan practice, and the Loving Kindness practice (Table 1) [36]. Each meditation had the option to choose from 4 lengths of time (7-30 minutes). The intent of offering varying lengths of time is that participants initially engage in shorter practices and progress to longer periods of sustained meditation. A dashboard with meditation features is shown in Figure 3. An example of meditation (ie, Loving Kindness) is depicted in Figure 4.

Table 1. Overview of the Unwinding Anxiety app engagement features with content.

Feature or day introduced	Description
Educational modules	
Modules 1-7, week 1	• Overview of the program, personalized goal setting for the program that is logged in the app, and an introduction to the modules • Topic areas focus on how worry and anxiety become habituated through reinforcement learning processes (ie, operant conditioning and reward-based learning), an overview of mindfulness and its application in identifying reinforcement patterns that result in negative health outcomes, and an introduction to curiosity as an attitudinal quality
Module 8-14, week 2	• Introduction to the application of reinforcement concepts (ie, trigger, behavior, and reward), specifically, learning how to recognize behaviors, identifying the "rewards" or outcomes of the behaviors (eg, cognitive, physical sensations, and emotions), and becoming disenchanted with these behaviors allowing for an alternative behavioral pattern to emerge • Novel to the program is the instruction to participants to not attempt to change behaviors immediately but to concentrate on the embodied experience (ie, interoceptive awareness and present moment awareness) of anxiety and the associated behaviors • The modules introduce the RAIN[a] practice and the role of curiosity in engaging with present moment experiences rather than judgment
Module 15-21, week 3	• Week 3 begins with troubleshooting and applying reinforcement lessons from the previous week • Participants are encouraged to gain acceptance of present moment experiences through resistance or unresistance, defined as the ability to engage with present moment experiences with curiosity, being aware of thoughts but not attached to them, and riding out waves of anxiety using the RAIN practice and other mindfulness exercises • Modules then focus on the detriments of "contracting" or identifying with thoughts (ie, anxiety) • Participants are asked to identify a variety of thought patterns (ie, anxiety, doubt, anger, and kindness) and observe rather than attach or react to these narratives
Module 22-30, week 4	• The week begins with explaining the science of resistance to habit change, specifically, participants are introduced to the association between anxiety and performance (ie, anxiety becomes associated with accomplishing tasks) • The previous modules regarding the application of mindfulness to unwind these associations are highlighted. Participants are guided through the importance of taking breaks when pursuing habit change, the advantages of alternate strategies to anxiety (ie, curiosity), and the ability to drop into the flow (ie, concentration and awareness focused on the present moment with a loss of reflective consciousness) • The week ends with the key elements of continued motivation and review of the effectiveness of the program, specifically improvements participants have observed, termed "evidence-based faith"
Ecological features	
Check-ins, day 1	• Select their current emotional state from a list provided (eg, happy, anxious, and relaxed) • Describe the strength of their anxiety on a 10-point Likert scale (1=low and 10=high) • Provided an exercise to complete (eg, hand awareness and breathe into anxiety)
Stress meter, day 1	• Identify the strength of their anxiety on a 10-point Likert scale (1=low and 10=high) • Identify from a list the reason for their anxiety (eg, uncompleted tasks and reliving past experiences) • Provided with a short exercise to complete (eg, breathe into anxiety)
Stress test, day 6	• Identify where anxiety is strongest in the body (eg, head, neck, and shoulders) • Select a description of the sensation from a provided list (eg, tightness, pressure, and burning) • Select the intensity from 0 to 100 on a scale (eg, 100=most stress ever) • Identify on which side of the body the anxiety is strongest (ie, left or right) • Provided with a short exercise to complete (eg, breathe into anxiety)
Meditations	
Resting in Awareness, day 1	• Participants offer gratitude to themselves for taking the time to take care of themselves • Subsequently, they are encouraged to shift their awareness to sounds in the room, then to thought processes by allowing thoughts to rise and pass away (ie, making a mental note of "thinking") • They are guided to directly observe thoughts as they arise and pass away on their own if not engaged with (eg, resisted) • Subsequently, they are instructed to open their eyes, engaging in awareness of sights, sounds, thoughts, and body sensations, whichever present moment experience is most prominent in their experience • Participants can choose from 9-, 15-, 20-, and 30-minute exercises

Feature or day introduced	Description
Body scan, day 3	• Participants bring their attention to physical sensations in their body (eg, touch and pressure) and how sensations are connected to feelings or emotions (eg, anxiety) and are guided on how to pay attention to thoughts and mental processes (eg, noticing how "busy" thoughts can get when connected with anxiety) • Participants can choose from 12-, 15-, 20-, and 30-minute exercises
Loving Kindness, day 5	• Participants notice the physical sensations associated with an imagined experience of anxiety then shift to an experience when they meet a dear friend or kind being (ie, person and animal) • From there, they are encouraged to investigate the different physical sensations of anxiety (eg, tightness and contraction) versus being with a kind person (eg, openness and warmth) • They are then asked to offer phrases of kindness to the person or being identified (eg, "May you be happy") using the phrases as mental anchors for present moment awareness • Participants can choose from 7-, 15-, 20-, and 30-minute exercises

[a] RAIN: Recognize and Relax, Allow and Accept, Investigate, and Note.

Figure 1. Unwinding Anxiety main dashboard.

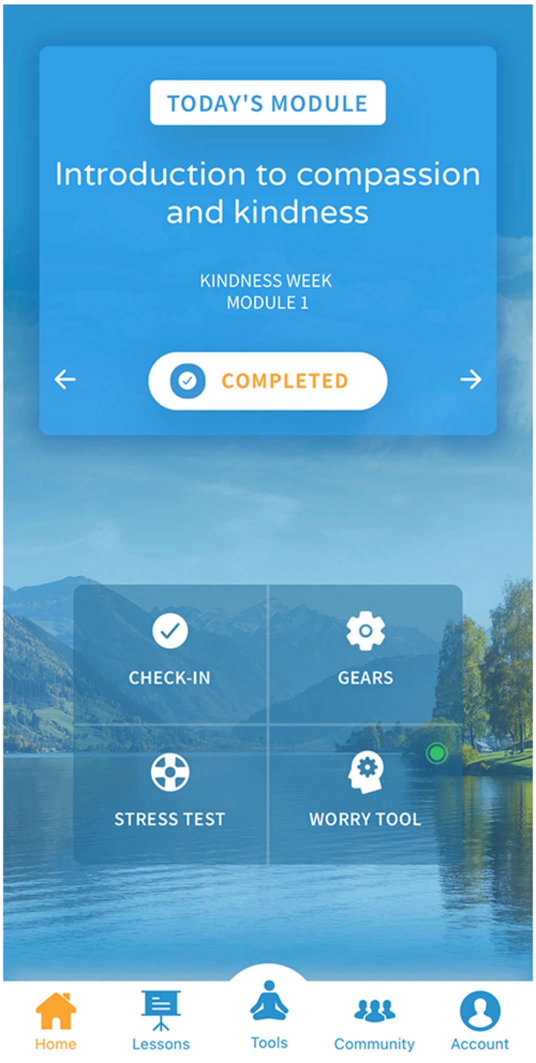
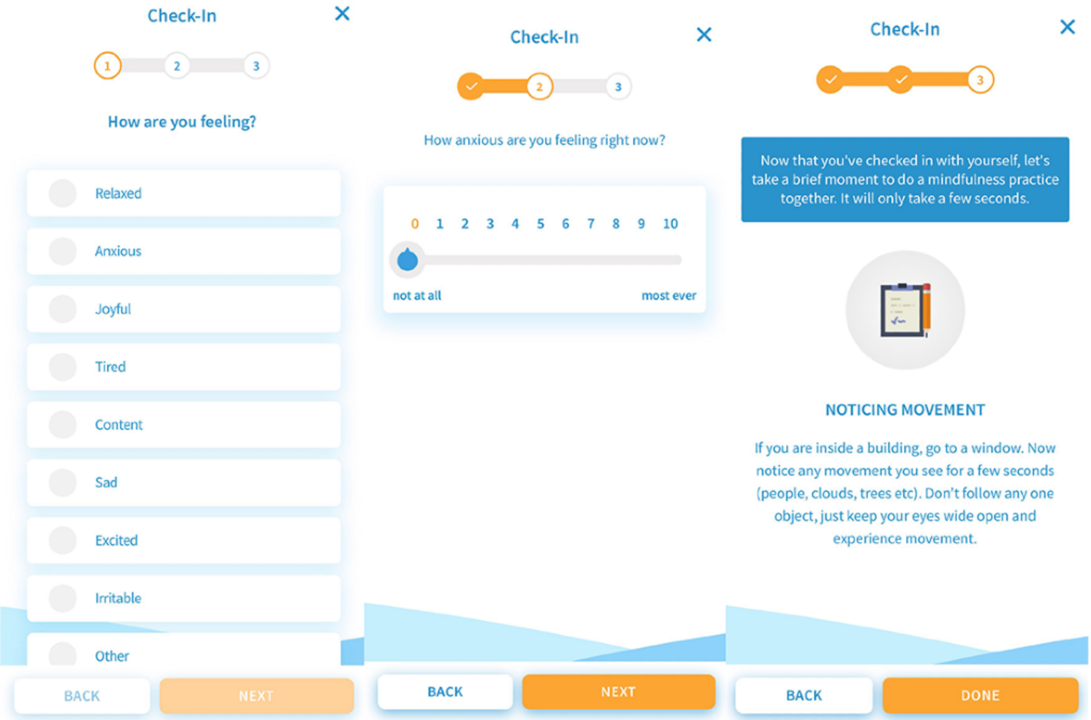


Figure 2. Adjunctive ecological feature example: check-ins.



How are you feeling?

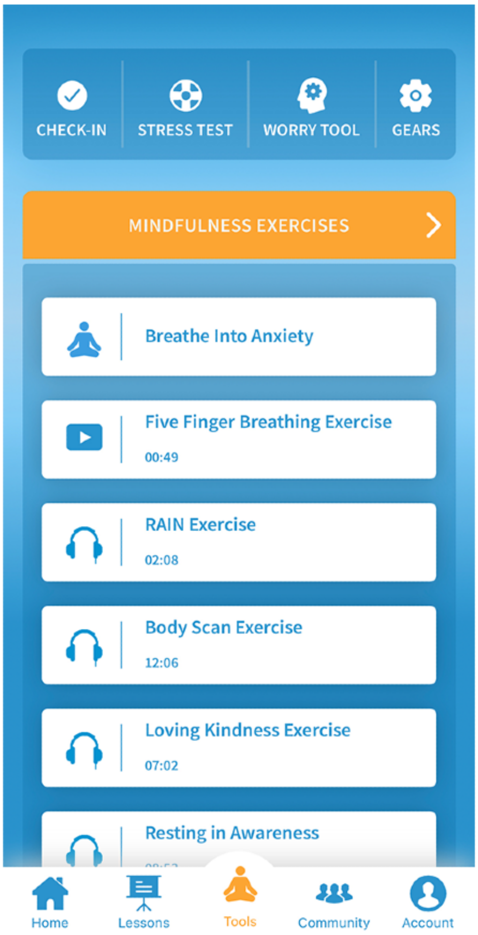
- ○ Relaxed
- ○ Anxious
- ○ Joyful
- ○ Tired
- ○ Content
- ○ Sad
- ○ Excited
- ○ Irritable
- ○ Other

How anxious are you feeling right now?

Now that you've checked in with yourself, let's take a brief moment to do a mindfulness practice together. It will only take a few seconds.

If you are inside a building, go to a window. Now notice any movement you see for a few seconds (people, clouds, trees etc). Don't follow any one object, just keep your eyes wide open and experience movement.

Figure 3. Adjunctive features dashboard.



Breathe Into Anxiety

Five Finger Breathing Exercise
00:49

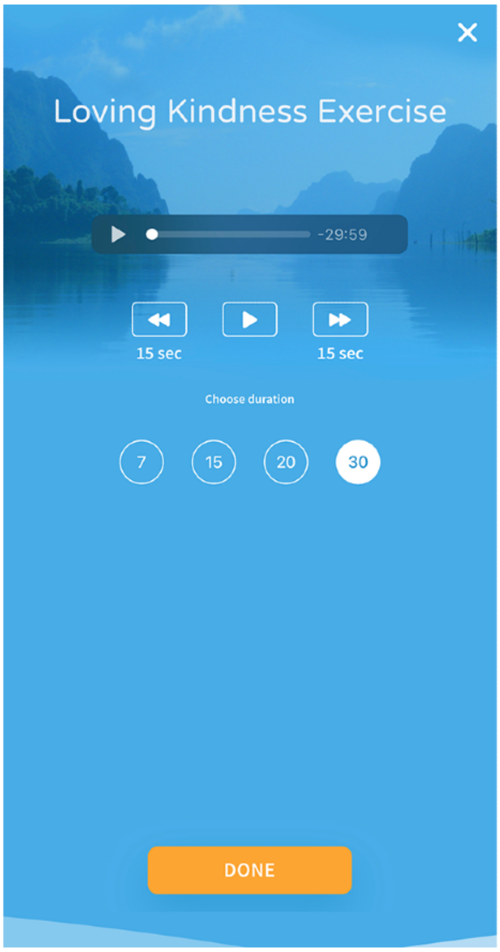
RAIN Exercise
02:08

Body Scan Exercise
12:06

Loving Kindness Exercise
07:02

Resting in Awareness

Figure 4. Adjunctive feature example: Loving Kindness meditation.



Measures

Demographics

Demographic variables collected at baseline included age, biological sex, education level, and current employment status.

Psychiatric Diagnoses

The MINI is a short (15-minute) structured diagnostic interview assessing 17 of the most prevalent mental health disorders, including depression and ADs [37]. Previous studies have validated the MINI against the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, and the International Classification of Diseases, Tenth Revision, and found the measure to be reliable and valid along with the added benefit of being more efficient [38].

Engagement Measures

The study measures were organized into 2 categories: microengagement and macroengagement. For clarity, in the model proposed by Cole-Lewis et al [31], engagement with the app-based intervention is referred to as the "little e" construct, which we categorize in these analyses as microengagement. The model then defines engagement with the targeted health outcome as the "Big E" construct, which we refer to as macroengagement [31].

Macroengagement: Mental Health Outcomes

Anxiety

Anxiety was assessed using the GAD-7, which comprises 7 items (eg, "feeling nervous, anxious, or on edge"). The GAD-7 is the most widely used self-report tool for clinical screening and tracking of GAD (sensitivity and specificity of 89% and 82%, respectively) [39]. Scores of ≥10 indicate a probable diagnosis of moderate GAD, whereas scores ≥15 indicate severe GAD [40]. In clinical testing, the scale demonstrated high reliability (Cronbach α=.88) and validity (r=0.69), correlated with the Beck Anxiety Index [41].

Emotional Regulation

The Five Facet Mindfulness Questionnaire (FFMQ) nonreactivity subscale was used to measure emotional regulation. The FFMQ is a validated 39-item questionnaire used to assess mindful awareness with high internal consistency (Cronbach α ≥.70) and an acceptable fit with a correlated 5-factor model (confirmatory fit index 0.914) [42]. It comprises 5 subscales, each validated for use independently [42,43]. The nonreactivity subscale comprises 7 questions (eg, "I perceive my feelings and emotions without having to react to them") from the 39-item FFMQ with acceptable internal consistency (Cronbach α=.75) [32].

Interoceptive Awareness

The Multidimensional Assessment of Interoceptive Awareness is a 32-item measure with response options provided on a

6-point Likert scale (0=never and 5=always) that has demonstrated moderate levels of internal consistency (Cronbach α ≥.70) and a good model fit (comparative fit index 0.886) [44]. The scale comprises eight subscales: (1) noticing, (2) not distracting, (3) not worrying, (4) attention regulation, (5) emotional awareness, (6) self-regulation, (7) body listening, and (8) trusting (eg, "I trust my body sensations") [44,45].

Worry

The Penn State Worry Questionnaire is a validated 16-item questionnaire used to assess worry (eg, "My worries overwhelm me"). It has high internal consistency (Cronbach α=.93) and validity compared with the Self-Analysis Questionnaire Tension subscale (r=0.36) and the Emotional Control Questionnaire (r=−0.53) [46]. Items are rated on a 5-point Likert scale (ie, 1=not at all typical and 5=very typical of me), and scoring for the scale is calculated as a total with possible ranges of 16 to 80 (eg, 60-80 indicating high worry) [46].

Microengagement: App Features Used

Microengagement (ie, system use) was defined as the total number of times each app feature was accessed [47]. The app features were organized into 3 categories for this study. The first category was engagement with the primary intervention feature and completion of the educational module. The other two engagement categories involved the use of the adjunctive app features: (1) ecological features and (2) meditations. These adjunctive or supportive features are components that participants could elect to use in combination with educational modules. Ecological features were considered adjunctive or supportive as they were designed as a short (<2 minutes in length) complement to the modules to assess participants' current emotional or cognitive state at the moment. Meditations were defined as features >5 minutes in length, which followed evidence-based guidelines consistent with researched meditation practices (eg, Loving Kindness meditation) [48]. Table 1 provides a detailed description of each app feature. Aggregated totals for each of the 3 categories, as well as the total use of individual features, were calculated.

Analysis

This was a secondary analysis of a previously conducted RCT. As such, the analysis was restricted to 32 participants who were randomized to the intervention group. A complete case analysis was conducted in this study. Of the 32 participants in the original intervention sample, 2 (6%) participants dropped out of the study before follow-up, and 2 (6%) participants did not complete the follow-up assessments, resulting in a final sample of 28 (88%) participants for this secondary analysis representing 88% of the original sample. For the primary RCT, a sample size of 52 was determined to have 80% power (1-sided Cronbach α=.05) to detect a statistically significant between-group difference in the primary outcome of anxiety (Cohen d=0.7). The final sample size was increased to 65 participants to account for 25% attrition.

Descriptive statistics, including means (SDs) for continuous variables and frequencies and percentages for categorical variables, were calculated for baseline variables, including demographics, anxiety, emotional regulation, and worry. Use data included ecological assessments (ie, check-ins, stress tests, and stress meters), mindfulness practices with the length of time completed (ie, Loving Kindness, Resting in Awareness, and body scan), and educational modules completed. Correlation analysis was used to identify potential confounders. Specifically, variables (eg, age, education, and income) were assessed using Spearman correlations, and those that met an a priori threshold of P≤.20 were included in subsequent models.

Microengagement was reported as the total number of times each app feature was accessed (ie, the total amount of use per item). Spearman rank correlation matrices were run to identify potential predictors to be included in the univariate linear regression models; predictors with a significance of P≤.20 were included [49]. Skewed engagement data were log transformed, except for the number of educational modules completed. Instead, the educational modules were dichotomized based on the completion of ≥75% (23/32) of the modules.

A series of multivariate linear regression models were constructed to investigate the dose-response relationships regarding the effect of use variables on psychosocial outcome measures. First, to determine predictors for inclusion in multivariate models, we ran a series of univariate linear models controlling for baseline scores to determine the impact of use variables on anxiety (ie, GAD-7), worry (ie, Penn State Worry Questionnaire), emotional regulation (ie, FFMQ), and interoceptive awareness (ie, Multidimensional Assessment of Interoceptive Awareness) at 2 months after treatment initiation. Predictors that were associated with outcomes at a modest P≤.20 level and met the assumptions testing criteria were then included in the subsequent multivariate linear regression models.

Final multivariate models investigated total meditation use, total ecological assessment use, individual meditation practices, and individual ecological assessments as predictors of the 4 outcomes (ie, anxiety, worry, interceptive awareness, and emotional regulation). All models controlled for baseline scores, as well as age and education level, as both have been shown to affect mobile app engagement in health behavior change in previous research [16]. Biological sex was not included as a covariate in the models to preserve parsimony because of a lack of correlation with outcomes (ie, P≥.20) in exploratory analyses and as only 7% (2/27) of the total sample identified as male versus 93% (25/27) as female. Analyses were run in R (version 4.0.3), and an a priori α level of .05 was set for the analyses. All models were evaluated to meet relevant model assumption criteria (eg, homoscedasticity and normality of residuals), and only models meeting the assumption criteria were reported. Summaries include R^2 statistics (measure of effect size), β coefficients, 2-tailed t test values, P values, and 95% CIs for all the included models.

Results

Demographics

The participant demographics are presented in Table 2. The sample was primarily female (25/27, 93%), and the average age was 42.9 (SD 15.6) years. Most participants reported completing some level of college education and self-reported as employed:

67% (18/27) reported having attained a bachelor's degree or higher, 59% (16/27) reported having full-time employment, and 74% (20/27) reported an annual income of ≥US $40,000 per year.

We explored potential differences in demographic information across high versus low engagement with the main intervention feature—education modules. High versus low engagement was quantified as those who completed >75% of the education modules (≥23 modules; high completion) and those who completed <75% (<23 modules; low completion). Both the high and low engagement groups averaged consistent numbers of participants who reported psychiatric conditions (Table 2). With one exception, the high engagement group had significantly more past depressive episodes (*P*=.02).

Table 2. Participant demographics: high versus low engagement and total sample (N=27).

Demographics	Low: <23 modules (n=13)	High: ≥23 modules (n=14)	*P* value	Total sample
Age (years), mean (SD)	37.7 (13.5)	47.7 (16.3)	.10	42.9 (15.6)
Sex, n (%)				
Female	13 (100)	12 (86)	.50	25 (93)
Male	0 (0)	2 (14)	.50	2 (7)
Education, n (%)				
Some college or technical school	3 (23)	4 (29)	.90	7 (26)
2-year degree	1 (78)	1 (7)	.90	2 (7)
4-year degree	2 (15)	3 (21)	.90	5 (19)
Master's degree	7 (54)	5 (36)	.90	12 (44)
Doctorate degree	0 (0)	1 (7)	.90	1 (4)
Current employment, n (%)				
Employed full-time (≥35 hours weekly)	8 (62)	8 (57)	.10	16 (59)
Employed part-time	3 (23)	0 (0)	.10	3 (11)
Not in the labor force	1 (8)	5 (36)	.10	6 (22)
Unemployed >1	1 (8)	1 (7)	.10	2 (7)
Income (US $), n (%)				
20,000-29,000	1 (8)	1 (7)	.70	2 (7)
30,000-39,000	2 (15)	1 (7)	.70	3 (11)
40,000-49,000	2 (15)	1 (7)	.70	3 (11)
50,000-59,000	1 (8)	0 (0)	.70	1 (4)
70,000-79,000	1 (8)	1 (7)	.70	2 (7)
80,000-89,000	1 (8)	1 (7)	.70	2 (7)
100,000-149,000	4 (31)	4 (29)	.70	8 (30)
>150,000	0 (0)	4 (29)	.70	4 (15)
Mini International Neuropsychiatric Interview diagnostic criteria, n (%)				
Major depressive episode, current	3 (23)	0 (0)	.10	3 (11)
Major depressive episode, past	2 (15)	9 (65)	.02[a]	11 (41)
Panic disorder, current	2 (15)	1 (7)	.99	3 (11)
Panic disorder, past	2 (15)	2 (14)	.99	4 (15)
Posttraumatic stress disorder, met criteria	1 (8)	1 (7)	.99	2 (7)
Obsessive-compulsive disorder, met criteria	2 (15)	2 (14)	.99	4 (15)
Social anxiety disorder, met criteria	2 (15)	2 (14)	.99	4 (15)
Agoraphobia, met criteria	1 (8)	4 (29)	.30	5 (19)

[a] Significance set at a priori α level of .05 (ie, $P \leq .05$).

The baseline anxiety, emotional regulation, worry, and interoceptive awareness scores are reported in Table 3. Overall, the sample had average anxiety scores of 13.0 (SD 4.9) and high average worry scores of 65.5 (SD 7.1). The average interoceptive awareness score was 78.7 (SD 22.4), and participants had an average low emotional regulation score of 15.2 (SD 4.2). No significant differences were observed between the high and low engagement groups in the primary outcomes at baseline.

Participants completed an average of 20.2 (SD 11.4) modules of the primary intervention feature for the total sample. Of these, 52% (14/27) were considered "high completers" (ie, ≥23 modules). Use data of the adjunctive features over the 2-month intervention period are presented separately for high versus low engagement in Table 4. For the ecological features, participants' total engagement averaged 38.1 (SD 52.6) uses. Regarding specific ecological features, the average check-in use accounted for most of the use at 31.1 (SD 45), whereas stress tests and stress meters averaged <5 uses for the 2-month period. When assessing adjunctive ecological feature engagement across high versus low primary intervention engagement, the high engagement group used adjunctive ecological tools more across all types of features than the low engagement group (Table 4).

Participants averaged 7.2 (SD 7.2) mindfulness practice uses over the 2-month period. The range of uses varied for meditations, although not as broadly as ecological tool use, with some individuals never using the meditations and other participants using the meditations more frequently, with median meditation use of 7 (IQR 8). In addition, when assessing meditation use across intervention engagement categories (ie, high vs low), those who completed a high level of the primary intervention also used meditation practices more on average (Table 4). Specifically, those who had a higher engagement with the primary intervention also used significantly more meditations overall, as well as the Loving Kindness meditation and Resting in Awareness meditation (Table 4).

Summary statistics of meditation practice use for the full sample, and for high versus low engagement, are presented in Table 3. The completion of cumulative mindfulness practices across the entire sample averaged 7.2 (SD 7.2) over the 2-month period. The range of uses varied for meditations, although not as broadly as the ecological tool uses, with some individuals never using the meditations and other participants using the meditations more frequently, with median meditation use of 7 (IQR 8). In reviewing differences in average meditation use between high and low engagement, the group with high engagement also used meditation practices more on average, and significant differences were observed for cumulative meditations used, Loving Kindness, and Resting in Awareness but not for the body scan meditations (Table 4).

Table 3. Average outcome measure scores at baseline.

Measure	Low: <23 modules, mean (SD)	High: ≥23 modules, mean (SD)	*P* value	Total sample, mean (SD)
Anxiety (Generalized Anxiety Disorder-7 scale)	13.4 (5.9)	12.6 (3.9)	.60	13.0 (4.9)
Emotional regulation (Five Facet Mindfulness Questionnaire, Nonreactivity Subscale only)	14.9 (4.2)	15.4 (4.4)	.70	15.2 (4.2)
Worry (Penn State Worry Questionnaire)	66.8 (6.6)	64.2 (7.6)	.30	65.5 (7.1)
Interoceptive awareness (Multidimensional Assessment of Interoceptive Awareness)	76.5 (19.6)	80.7 (25.2)	.60	78.7 (22.4)

Table 4. Average number of tool uses by segment over 2 months: high versus low engagement and total sample.

App components	Low: <23 modules, mean (SD)	High, ≥23 modules, mean (SD)	*P* value	Total sample, mean (SD)
Cumulative ecological tool use	12.9 (14.4)	61.6 (64.1)	.02[a]	38.1 (52.6)
Check-ins	10.5 (12.0)	50.4 (55.5)	.02	31.1 (45.0)
Stress test	1.2 (1.9)	4.9 (6.2)	.04	3.1 (4.9)
Stress meter	0.5 (0.7)	2.2 (2.2)	.01	1.4 (1.9)
Breath awareness	0.9 (1.9)	4.1 (5.4)	.05	2.5 (4.3)
Cumulative mindfulness practice use	3.1 (2.9)	10.9 (7.9)	<.01	7.2 (7.2)
Loving Kindness Practice	1.0 (1.1)	4.2 (2.6)	<.01	2.7 (2.6)
Body scan	1.1 (1.1)	3.4 (4.5)	.09	2.3 (3.5)
Resting in Awareness	1.0 (1.4)	3.4 (2.8)	.01	2.2 (2.5)

[a] Significance set at a priori α level of .05 (ie, $P \leq .05$).

Multivariate Regression Results

Overview

Analyses were conducted to identify associations between microengagement, quantified as specific features use tools, and macroengagement, defined as worry, anxiety, interoceptive awareness, and emotional regulation. Predictors (ie, microengagement metrics) were entered into univariate models, controlling for baseline scores on the outcome of interest, and assumption testing was performed on models that met the a priori significance threshold ($P \leq .20$). However, after review, models that violated assumptions were excluded, and those that met the criteria were consolidated for subsequent multivariate testing.

Using the results from the univariate models, multivariate linear regression models were built to answer the primary research question testing the association of (1) total meditation practice, (2) total ecological assessment, (3) individual meditation practices, and (4) individual ecological assessments with changes in psychosocial outcomes at 2 months (ie, anxiety, worry, emotional regulation, and interoceptive awareness). Only models that met a priori significance levels and all model assumptions are reported (Table 5) by system use categorization (ie, educational modules, ecological tools, and meditation practices). Parameter estimates, effect sizes (ie, coefficient of determination; R^2), 95% CIs, and additional relevant statistics for all the included models are outlined in Table 5.

Table 5. Multivariate linear regression analyses of use metrics and change in psychosocial measures at 2 months.

Use metric[a]	Outcome	R^2	*P* value	*t* test (*df*)	*b* (95% CI)
Low vs high module	Interoception[b]	0.22	.01	2.72 (26)	22.6 (5.3 to 39.8)
Low vs high module	Worry[c]	0.25	.01	−2.83 (26)	−11.6 (−20.2 to −3.1)
Total meditations[d]	Anxiety[e]	0.21	.03	−3.16 (26)	−0.6 (−1.2 to −0.6)
Stress meter[d]	Mindfulness[f]	0.29	.03	2.36 (26)	0.5 (0.1 to 0.9)

[a] All models were adjusted for age, educational attainment, and baseline outcome measure total score.

[b] Multidimensional Assessment of Interoceptive Awareness.

[c] Penn State Worry Questionnaire.

[d] Variable was log transformed, and the results are reported on a log scale.

[e] Generalized Anxiety Disorder-7 scale.

[f] Five Facet Mindfulness Questionnaire, Nonreactivity Subscale only.

Education Modules

Completion of ≥75% of the educational modules was significantly associated with increases in interoceptive awareness and decreases in worry. More specifically, it was associated with an average 22.6-point increase in interoceptive awareness (SE 8.32; *P*=.01; 95% CI 5.3-39.8) and an 11.6-point decrease in worry (SE 4.12; *P*=.01; 95% CI −20.2 to −3.1) when holding age, education level, and baseline worry and interoceptive awareness scores constant.

Meditation Practices

Total meditation practice was associated with a significant reduction in anxiety scores. For each log unit change in the total number of meditations, there was a reduction of 0.62 in anxiety scores (SE 0.27; *P*=.005; 95% CI −1.2 to −0.6) after controlling for age, educational level, and baseline anxiety scores.

Ecological Tools

Stress meter use was associated with significant changes in emotional regulation. Specifically, a difference of 1 unit (on the log scale) in the stress meter was associated with a 0.5-unit increase on average in emotional regulation scores (SE 0.21; *P*=.03; 95% CI 0.1-0.9). Both associations were adjusted for age, education level, and baseline emotional regulation scores.

Discussion

Principal Findings

Comprehensive app-based interventions offer promising, high-fidelity treatments for AD, which are critical for addressing the growing treatment needs of the health care system [7,9,10,50]. To fully realize their potential, research will need to maximize engagement through a comprehensive understanding of how microengagement with in-app features affects macrolevel engagement in the target health outcome [31]. Our analyses sought to build on existing evidence by using multivariate linear regression models to examine the associations between improvements in mental health outcomes and the use of specific features within the UA mobile app. The results indicated that microengagement with app features was associated with significant changes in macroengagement in health outcomes (ie, interoceptive awareness, anxiety, emotional regulation, and worry), consistent with the engagement model proposed by Cole-Lewis et al [31]. Consistent with our hypothesis, engagement with the primary intervention feature (ie, completing at least 75% of the modules) was associated with an average increase of 22 points in interoceptive awareness scores and an 11-point decrease in clinical measures of worry at the 2-month follow-up. Associations were also found between the adjunctive app features (ie, ecological features and meditations) and improvements in health outcomes. Total meditation use was associated with a highly significant 0.62-point average reduction in anxiety (GAD-7), and the use of the Stress Meter, a tool for recognizing and investigating stressful situations in the moment, was associated with an average increase of 0.5 points in emotional regulation.

Moving forward, this research has significant implications for testing, developing, and customizing app-based interventions that target AD. First, this study offers important evidence that engagement with in-app features is a critical mechanism for increasing treatment benefits. Specifically, our findings showed a strong association between module completion and mental health outcomes, offering important evidence within the limited field of work on engagement with app-based interventions for AD. Previous research has largely focused on module completion as the primary engagement metric, finding small but significant associations with improvements in mental health outcomes for all digital interventions (eg, internet based and app based) [51-55]. However, these studies were mostly internet-based interventions, with few apps clinically tested for the effects of engagement on AD-related outcomes [23,24].

In addition, we found that intervention effects were driven by more than module completion, with findings indicating that adjunctive features (eg, ecological check-ins) were associated with important changes in health outcomes. The adjunctive or supportive components of complex app-based interventions have rarely been investigated. Instead, to date, most research has favored the analysis of completed modules or using a total frequency metric for all features used [24]. Focusing singularly on these adjunctive components offers 2 important insights: a clear indication of how those additional features are associated with changes in outcomes and information on potentially unnecessary or ineffective tools, which may needlessly complicate the app. Future research building on this study, focusing on understanding how specific supportive features function mechanistically to affect clinical outcomes, will be important as this information allows developers to streamline digital interventions [31]. By prioritizing streamlining app-based interventions, highlighting the effective features and removing ineffective features we can increase intervention efficiency as well as the likelihood of initiating and sustaining health behavior changes [22,56].

Although our analyses offer important novel insights into engagement and health outcomes, to fully capitalize on the potential of app-based interventions for sustained behavior change engagement, research needs to move beyond considering engagement as an aggregated total. Summed user engagement metrics, such as those used in these analyses, offer important insights but provide a limited view when considering the practical reality of engagement as a dynamic interaction between the participant and the app over time [22,23,47,57]. In the case of UA specifically, the app is designed to develop knowledge (eg, retraining reinforcement learning, education on understanding anxiety, and goal-directed behavior training) and promote skills maintenance (eg, meditation and emotional regulation), which can be applied in everyday life. The eventual goal of the app is actually "off ramping" participants or effectively reducing microengagement with the app while sustaining the associated health outcome improvements (eg, reductions in anxiety and increased emotional regulation). In this case, the pattern of engagement may change; for example, a participant may start at a higher level of engagement but begin to titrate how often they use the app over time. As such, the goal for UA is that health behavior change would be sustained despite lower engagement over time, which cross-sectional or sum engagement metrics cannot reflect. Thus, future analyses will need to move beyond aggregated engagement and analyze individual time series data to capture the dynamic, longitudinal relationship between in-app engagement and sustained health behavior changes.

Comparison With Previous Work

Although the analysis presented is the first to investigate associations between the use of UA tools and mental health outcomes, the underlying reinforcement learning theory and curriculum for the primary intervention feature have been associated with behavioral reductions in our previous work on smoking cessation [58,59]. Specifically, our findings regarding module engagement build on results from an RCT (n=225) of the Craving to Quit program, a mobile platform designed to help people quit smoking, which uses a similar suite of educational modules grounded in the reinforcement learning theory and application of meditation practices on which the UA program is founded. These previous results indicated that educational module engagement, categorized as low (0-14 modules), medium (15-41 modules), and high (≥42 modules) in linear mixed models, was associated with significant reductions in the relationship between craving and cigarette smoking at 6 months ($F_{1,104}$=4.44; *P*=.04) in the intervention group (n=182) [58].

In addition, in a previous trial of an app-based intervention targeting eating behaviors and craving-related eating—Eat Right Now—the results indicated significant associations between ecological feature engagement and craving-related eating [60]. Eat Right Now also comprises ecological features, which guide participants by investigating cravings, in addition to mindful eating exercises, to reduce the reward value of overeating [60,61]. The results showed significant effects of ecological feature use on reductions in craving-related eating, anticipated reward value of craving-related eating, and the likelihood of engaging in craving-related eating within an initial sample of 46 participants, which was then replicated among a larger sample (n=1119) using Rescorla-Wagner computational modeling [61].

This study, coupled with previous work, demonstrates the important role of multiple engagement metrics associated with positive changes in health outcomes among a range of clinical populations.

Limitations

This study has some limitations. First, engagement data were calculated as the total number of uses. It is possible that a feature was opened, closed, or stopped midway. It is also possible that the participant was not actively engaged with the feature (eg, distracted or the participant was not the one using the app). In future studies, the length of use time should be considered for analysis to mitigate these limitations. In addition, this analysis considered specific features that operate singularly on health outcomes. However, this is unlikely to reflect the engagement patterns of participants who use a combination of features in pursuit of their health goals. Aggregation of microengagement patterns with the inclusion of the total time of engagement may offer windows into specific "usage profiles," which could more effectively reflect the interdependent nature of many of these features. Another important limitation was the small sample size available for analysis. In addition, the sample was overwhelmingly female, White, and highly educated, making the generalizability of these findings to broader populations questionable. The length of follow-up may also have been a limiting factor for the analysis as there was variability regarding module completion, which may not have been aligned with the 2-month assessment window. Future analyses should use larger, more diverse samples to test the impact of microengagement patterns in the broader population.

Conclusions

This secondary analysis offers evidence of associations linking the use of in-app features in UA to improvements in mental health outcomes; however, these results are preliminary and exploratory. The work presented offers a clearer understanding of the impact of how microengagement with the app features affects macroengagement with health outcomes of interest, consistent with the model proposed by Cole-Lewis et al [31]. This study highlights the importance of comprehensive investigations of engagement during the development of evidence-based app-based interventions. Future research would benefit from comprehensive, longitudinal analyses, as well as primary studies that specifically assess the effects of varying levels of engagement trials on health outcomes (eg, microrandomized trials, rapid optimization methods, and multiphase optimization strategies) [18,62-64]. As research continues, understanding the effects of engagement with app-based intervention features on clinical outcomes will be critical to designing targeted interventions needed to increase patient health and support accessible, comprehensive care.

Conflicts of Interest

JB and AR are paid advisers to Sharecare, the company that owns the mindfulness app used in this study. This financial interest has been disclosed to and is being managed by Brown University, in accordance with its Conflict of Interest and Conflict of Commitment policies, including being restricted from recruitment, being blinded to the study group until after analysis, and not having access to data or performing analyses.

References

1. Managing stress and anxiety. Anxiety and Depression Association of America. URL: https://adaa.org/living-with-anxiety/managing-anxiety [accessed 2022-07-21]
2. Liu C, Stevens C, Conrad R, Hahm H. Evidence for elevated psychiatric distress, poor sleep, and quality of life concerns during the COVID-19 pandemic among U.S. young adults with suspected and reported psychiatric diagnoses. Psychiatry Res 2020 Oct;292:113345 [FREE Full text] [doi: 10.1016/j.psychres.2020.113345] [Medline: 32745794]
3. Jewell JS, Farewell CV, Welton-Mitchell C, Lee-Winn A, Walls J, Leiferman JA. Mental health during the COVID-19 pandemic in the United States: online survey. JMIR Form Res 2020 Oct 23;4(10):e22043 [FREE Full text] [doi: 10.2196/22043] [Medline: 33006939]
4. Vissink CE, van Hell HH, Galenkamp N, van Rossum IW. The effects of the COVID-19 outbreak and measures in patients with a pre-existing psychiatric diagnosis: a cross-sectional study. J Affect Disord Rep 2021 Apr;4:100102 [FREE Full text] [doi: 10.1016/j.jadr.2021.100102] [Medline: 33558866]
5. Hao F, Tan W, Jiang L, Zhang L, Zhao X, Zou Y, et al. Do psychiatric patients experience more psychiatric symptoms during COVID-19 pandemic and lockdown? A case-control study with service and research implications for immunopsychiatry. Brain Behav Immun 2020 Jul;87:100-106 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.069] [Medline: 32353518]
6. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun 15;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
7. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone apps for the treatment of mental disorders: systematic review. JMIR Mhealth Uhealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
8. Mor S, Grimaldos J, Tur C, Miguel C, Cuijpers P, Botella C, et al. Internet- and mobile-based interventions for the treatment of specific phobia: a systematic review and preliminary meta-analysis. Internet Interv 2021 Dec;26:100462 [FREE Full text] [doi: 10.1016/j.invent.2021.100462] [Medline: 34646752]
9. Saramago P, Gega L, Marshall D, Nikolaidis GF, Jankovic D, Melton H, et al. Digital interventions for Generalized Anxiety Disorder (GAD): systematic review and network meta-analysis. Front Psychiatry 2021;12:726222 [FREE Full text] [doi: 10.3389/fpsyt.2021.726222] [Medline: 34938209]
10. Wu A, Scult MA, Barnes ED, Betancourt JA, Falk A, Gunning FM. Smartphone apps for depression and anxiety: a systematic review and meta-analysis of techniques to increase engagement. NPJ Digit Med 2021 Feb 11;4(1):20 [FREE Full text] [doi: 10.1038/s41746-021-00386-8] [Medline: 33574573]
11. Devane C, Chiao E, Franklin M, Kruep E. Anxiety disorders in the 21st century: status, challenges, opportunities, and comorbidity with depression. Am J Manag Care 2005 Oct;11(12 Suppl):S344-S353 [FREE Full text] [Medline: 16236016]
12. Ormel J. Common mental disorders and disability across cultures. JAMA 1994 Dec 14;272(22):1741. [doi: 10.1001/jama.1994.03520220035028]
13. Hom MA, Stanley IH, Joiner TE. Evaluating factors and interventions that influence help-seeking and mental health service utilization among suicidal individuals: a review of the literature. Clin Psychol Rev 2015 Aug;40:28-39. [doi: 10.1016/j.cpr.2015.05.006] [Medline: 26048165]

14. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
15. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
16. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
17. McVay MA, Bennett GG, Steinberg D, Voils CI. Dose-response research in digital health interventions: concepts, considerations, and challenges. Health Psychol 2019 Dec;38(12):1168-1174 [FREE Full text] [doi: 10.1037/hea0000805] [Medline: 31580127]
18. Kwasnicka D, Keller J, Perski O, Potthoff S, Ten Hoor GA, Ainsworth B, et al. White paper: open digital health - accelerating transparent and scalable health promotion and treatment. Health Psychol Rev 2022 Mar 14:1-17. [doi: 10.1080/17437199.2022.2046482] [Medline: 35240931]
19. Smith VA, Coffman CJ, Hudgens MG. Interpreting the results of intention-to-treat, per-protocol, and as-treated analyses of clinical trials. JAMA 2021 Aug 03;326(5):433-434. [doi: 10.1001/jama.2021.2825] [Medline: 34342631]
20. Gomolin A, Lebouché B, Engler K, Vedel I. Optimizing smartphone intervention features to improve chronic disease management: a rapid review. Health Informatics J 2020 Sep 12;26(3):1795-1809 [FREE Full text] [doi: 10.1177/1460458219891377] [Medline: 31826703]
21. Satre DD, Meacham MC, Asarnow LD, Fisher WS, Fortuna LR, Iturralde E. Opportunities to integrate mobile app-based interventions into mental health and substance use disorder treatment services in the wake of COVID-19. Am J Health Promot 2021 Nov;35(8):1178-1183. [doi: 10.1177/08901171211055314] [Medline: 34652971]
22. Molloy A, Anderson PL. Engagement with mobile health interventions for depression: a systematic review. Internet Interv 2021 Dec;26:100454 [FREE Full text] [doi: 10.1016/j.invent.2021.100454] [Medline: 34621626]
23. 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]
24. Gan DZ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079 [FREE Full text] [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
25. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
26. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015 May 19;6(1):27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
27. Lee J, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
28. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
29. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
30. O'Brien HL, Morton E, Kampen A, Barnes SJ, Michalak EE. Beyond clicks and downloads: a call for a more comprehensive approach to measuring mobile-health app engagement. BJPsych Open 2020 Aug 11;6(5):e86 [FREE Full text] [doi: 10.1192/bjo.2020.72] [Medline: 32778200]
31. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
32. Roy A, Hoge EA, Abrante P, Druker S, Liu T, Brewer JA. Clinical efficacy and psychological mechanisms of an app-based digital therapeutic for generalized anxiety disorder: randomized controlled trial. J Med Internet Res 2021 Dec 02;23(12):e26987 [FREE Full text] [doi: 10.2196/26987] [Medline: 34860673]
33. Roy A, Druker S, Hoge EA, Brewer JA. Physician anxiety and burnout: symptom correlates and a prospective pilot study of app-delivered mindfulness training. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e15608 [FREE Full text] [doi: 10.2196/15608] [Medline: 32234708]
34. 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(1):312. [doi: 10.1186/1471-244x-13-312]

35. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided Acceptance and Commitment Therapy for social anxiety disorder and/or panic disorder provided via the internet and a smartphone application: a randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
36. Bishop SR, Lau M, Shapiro S, Carlson L, Anderson ND, Carmody J, et al. Mindfulness: a proposed operational definition. Clin Psychol Sci Pract 2004;11(3):230-241. [doi: 10.1093/clipsy.bph077]
37. Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, et al. DSM-5 criteria for substance use disorders: recommendations and rationale. Am J Psychiatry 2013 Aug;170(8):834-851 [FREE Full text] [doi: 10.1176/appi.ajp.2013.12060782] [Medline: 23903334]
38. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
39. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
40. 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]
41. Johnson SU, Ulvenes PG, Øktedalen T, Hoffart A. Psychometric properties of the general anxiety disorder 7-item (GAD-7) scale in a heterogeneous psychiatric sample. Front Psychol 2019 Aug 6;10:1713 [FREE Full text] [doi: 10.3389/fpsyg.2019.01713] [Medline: 31447721]
42. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, et al. Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. Assessment 2008 Sep;15(3):329-342. [doi: 10.1177/1073191107313003] [Medline: 18310597]
43. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
44. Mehling WE, Price C, Daubenmier JJ, Acree M, Bartmess E, Stewart A. The multidimensional assessment of interoceptive awareness (MAIA). PLoS One 2012;7(11):e48230 [FREE Full text] [doi: 10.1371/journal.pone.0048230] [Medline: 23133619]
45. Mehling WE, Gopisetty V, Daubenmier J, Price CJ, Hecht FM, Stewart A. Body awareness: construct and self-report measures. PLoS One 2009;4(5):e5614 [FREE Full text] [doi: 10.1371/journal.pone.0005614] [Medline: 19440300]
46. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
47. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
48. Crane B. Full catastrophe living; using the wisdom of your body and mind to face stress, pain and illness. Jon Kabat-Zinn, 1996 Publisher: Piatkus, London ISBN: 0-749-915-854. Spirituality Health 2006 Jun;3(2):52. [doi: 10.1002/shi.88]
49. Bendel RB, Afifi AA. Comparison of stopping rules in forward "stepwise" regression. J Am Stat Assoc 1977 Mar;72(357):46-53. [doi: 10.1080/01621459.1977.10479905]
50. Lemon C, Huckvale K, Carswell K, Torous J. A narrative review of methods for applying user experience in the design and assessment of mental health smartphone interventions. Int J Technol Assess Health Care 2020;36(1):64-70. [doi: 10.1017/S0266462319003507] [Medline: 31973787]
51. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. Psychotherapy (Chic) 2014 Jun;51(2):207-219. [doi: 10.1037/a0032527] [Medline: 24041199]
52. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]
53. Hedman E, Andersson E, Lekander M, Ljótsson B. Predictors in Internet-delivered cognitive behavior therapy and behavioral stress management for severe health anxiety. Behav Res Ther 2015 Jan;64:49-55. [doi: 10.1016/j.brat.2014.11.009] [Medline: 25540862]
54. Hedman E, Lindefors N, Andersson G, Andersson E, Lekander M, Rück C, et al. Predictors of outcome in Internet-based cognitive behavior therapy for severe health anxiety. Behav Res Ther 2013 Oct;51(10):711-717. [doi: 10.1016/j.brat.2013.07.009] [Medline: 23994252]
55. Oromendia P, Orrego J, Bonillo A, Molinuevo B. Internet-based self-help treatment for panic disorder: a randomized controlled trial comparing mandatory versus optional complementary psychological support. Cogn Behav Ther 2016 Jun;45(4):270-286. [doi: 10.1080/16506073.2016.1163615] [Medline: 27007256]
56. 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]

57. Torous J, Powell AC, Rodriguez-Villa E. Health information technology resources to support measurement-based care. Child Adolesc Psychiatr Clin N Am 2020 Oct;29(4):763-773. [doi: 10.1016/j.chc.2020.06.011] [Medline: 32891375]
58. Garrison KA, Pal P, O'Malley SS, Pittman BP, Gueorguieva R, Rojiani R, et al. Craving to quit: a randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2020 Mar 16;22(3):324-331 [FREE Full text] [doi: 10.1093/ntr/nty126] [Medline: 29917096]
59. Janes AC, Datko M, Roy A, Barton B, Druker S, Neal C, et al. Quitting starts in the brain: a randomized controlled trial of app-based mindfulness shows decreases in neural responses to smoking cues that predict reductions in smoking. Neuropsychopharmacology 2019 Aug 30;44(9):1631-1638 [FREE Full text] [doi: 10.1038/s41386-019-0403-y] [Medline: 31039580]
60. Mason AE, Jhaveri K, Cohn M, Brewer JA. Testing a mobile mindful eating intervention targeting craving-related eating: feasibility and proof of concept. J Behav Med 2018 Apr;41(2):160-173 [FREE Full text] [doi: 10.1007/s10865-017-9884-5] [Medline: 28918456]
61. Taylor VA, Moseley I, Sun S, Smith R, Roy A, Ludwig VU, et al. Awareness drives changes in reward value which predict eating behavior change: probing reinforcement learning using experience sampling from mobile mindfulness training for maladaptive eating. J Behav Addict 2021 Oct 05;10(3):482-497 [FREE Full text] [doi: 10.1556/2006.2021.00020] [Medline: 34264854]
62. Morton K, Ainsworth B, Miller S, Rice C, Bostock J, Denison-Day J, et al. Adapting behavioral interventions for a changing public health context: a worked example of implementing a digital intervention during a global pandemic using rapid optimisation methods. Front Public Health 2021 Apr 26;9:668197 [FREE Full text] [doi: 10.3389/fpubh.2021.668197] [Medline: 33981669]
63. Guastaferro K, Collins LM. Achieving the goals of translational science in public health intervention research: the multiphase optimization strategy (MOST). Am J Public Health 2019 Feb;109(S2):S128-S129. [doi: 10.2105/AJPH.2018.304874] [Medline: 30785800]
64. Dimairo M, Pallmann P, Wason J, Todd S, Jaki T, Julious SA, et al. The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. BMJ 2020 Jun 17;369:m115 [FREE Full text] [doi: 10.1136/bmj.m115] [Medline: 32554564]

Abbreviations

AD: anxiety disorder
FFMQ: Five Facet Mindfulness Questionnaire
GAD: generalized anxiety disorder
GAD-7: Generalized Anxiety Disorder 7-item scale
MINI: Mini International Neuropsychiatric Interview
RCT: randomized controlled trial
UA: Unwinding Anxiety

Edited by G Eysenbach; submitted 19.09.21; peer-reviewed by N Nazari, H Mohamadian; comments to author 06.10.21; revised version received 18.11.21; accepted 13.07.22; published 15.08.22.

Please cite as:
Nardi W, Roy A, Dunsiger S, Brewer J
Analyzing the Impact of Mobile App Engagement on Mental Health Outcomes: Secondary Analysis of the Unwinding Anxiety Program
J Med Internet Res 2022;24(8):e33696
URL: https://www.jmir.org/2022/8/e33696
doi:10.2196/33696
PMID:35969440

Original Paper

Effects of a WeChat-Based Life Review Program for Patients With Digestive System Cancer: 3-Arm Parallel Randomized Controlled Trial

Meihua Zheng[1,2*], MM; Xiaoling Zhang[1*], MM; Huimin Xiao[1], PhD

[1]School of Nursing, Fujian Medical University, Fuzhou City, Fujian Province, China
[2]The Second Affiliated Hospital of Fujian Medical University, Quanzhou City, Fujian Province, China
[*]these authors contributed equally

Corresponding Author:
Huimin Xiao, PhD
School of Nursing
Fujian Medical University
No 1 Xuefu North Road, University Town
Shangjie Town, Minhou County
Fuzhou City, Fujian Province, 350122
China
Phone: 86 13860692061
Email: huimin_xiao@126.com

Abstract

Background: Patients with digestive system cancer often experience psychospiritual distress. Life review is an evidence-based psychological intervention for patients with cancer, but the effects of digital life review programs are unclear, especially for patients with digestive system cancer.

Objective: We examined the effects of a WeChat-based life review program on the psychospiritual well-being of patients with digestive system cancer.

Methods: This study was a 3-arm parallel randomized controlled trial. Eligible patients with digestive system cancer were recruited from a university hospital in Fujian, China. They were randomized to a life review group and 2 control groups. All participants received routine care, and the life review group also received the 4-week WeChat-based life review program. Control group 1 also received a 4-week program of friendly visiting. Anxiety, depression, hope, and self-transcendence were measured at baseline and 2 days, 1 month, and 6 months after the intervention.

Results: A total of 150 participants were randomly allocated to the WeChat-based life review group (n=50), control group 1 (n=50), or control group 2 (n=50). The overall dropout rate was 10% (15/150), and 92% (46/50) of participants in the the life review group completed the intervention. Significant interaction effects for time and group membership were found for anxiety (*P*<.001), depression (*P*<.001), hope (*P*<.001), and self-transcendence (*P*<.001) at all follow-up time points. For anxiety and depression, the scores did not differ significantly between the life review group and control group 1 on day 2 (*P*=.80 for anxiety, *P*=.51 for depression), but the scores were significantly lower in the life review group at month 1 and month 6 (*P*=.02 for anxiety at both months 1 and 6; *P*=.003 and *P*<.001 for depression at months 1 and 6, respectively). Significant increases in hope and self-transcendence were revealed in the life review group compared to control group participants at all follow-up sessions.

Conclusions: The WeChat-based life review program was effective in reducing anxiety and depressive symptoms and in improving the level of hope and self-transcendence among patients with digestive system cancer. Though friendly visiting can also help to relieve anxiety, its effects are short-term.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IOR-17011998; https://tinyurl.com/5acycpd4

(J Med Internet Res 2022;24(8):e36000) doi:10.2196/36000

KEYWORDS

digestive system cancer; life review; digital technology; anxiety; depression; hope; self-transcendence; cancer; randomized controlled trial; distress; psychological; digestive system

Introduction

Cancer is one of the leading causes of mortality and morbidity in the world; approximately 19 million new cases and 10 million deaths occurred in 2020, and these numbers are predicted to increase by 50% over the next 20 years. In China, digestive system cancers, including cancers of the colon, rectum, stomach, liver, and esophagus, are ranked within the top 5 diagnoses, accounting for 41% of new cancer cases and 49% of mortality [1]. A recent systematic review and meta-analysis reported a high prevalence of anxiety and depressive symptoms among patients with digestive system cancer, ranging from 50% among patients with hepatic and pancreatic cancer to 70% among patients with colorectal, esophageal, and gastric cancer [2]. Hopelessness, meaninglessness, and despair are also reported often, along with negative emotions triggered by concerns about death, seeking meaning in life, or unresolved life events associated with regret or pity [3,4].

A systematic review revealed the importance of psychological interventions in palliative care, as they can specifically address patients' emotional difficulties and spiritual concerns [5]. Life review has been recognized as an effective psychological intervention. It is a process of recalling, evaluating, and integrating life experiences to facilitate the achievement of ego integrity [6]. Ego integrity is a state of achieving a sense of meaning and acceptance of past life events that has been found to relate to higher levels of mental health and well-being among patients in a palliative care setting [7]. Life review enables patients to express their emotions, confirm their roles in life, reassess their attitudes toward death, reorganize their perspectives toward life, and finally integrate their entire life into a more acceptable or meaningful whole [8]. Originally, life review targeted older adults' psychosocial crises, but it has since been applied to palliative care. Accumulated evidence suggests that life review could reduce anxiety and depression, elevate hope and meaning in life, and improve self-transcendence and the quality of life of patients with cancer [9-11].

Digital technologies are increasingly being used to promote life review interventions via mobile phones, computers, wearable devices, and social media or applications [12-14]. Wise et al [15] first designed a telephone-based life review and illness narrative intervention with online resources for patients with cancer to share their personal stories and establish social networks. Afterwards, Wise et al [16] further demonstrated the effectiveness of life review in increasing feelings of peace and decreasing negative mood in patients with stage III or IV cancer after 4 months of the program. However, telephone-based life review interviews did not provide the opportunity to observe nonverbal cues, such as patients' facial expressions and body language. Additionally, Wise reported a high dropout rate in a sample that was predominantly White, female, and had high income and high education. Recently, Dang et al [17] tested an avatar-facilitated life review intervention to reconstruct the self and identity of patients with cancer through performativity. Patients were given full-body movement devices that captured their motions and synchronized their voices onto an avatar in a virtual environment. Although the virtual environment induced a sense of immersion during the therapeutic interaction and, therefore, enhanced patients' engagement and self-expression, the program was expensive, and there were hardware limitations [17-19].

Social media sites, such as Facebook, Instagram, and YouTube, and mobile applications have also been used to conduct life review interventions, because they enable patients to share photos, videos, and life stories [20]. WeChat is a social media platform with high popularity in 200 countries, especially China, due to its simplicity, convenience, efficiency, and mobility [21]. It allows users to interact asynchronously with each other through text messaging, voice messaging, video conferencing, and other means, as well as obtain information and resources from various WeChat platforms at any time. In 2018, our research team developed a WeChat-based life review program for patients with cancer, consisting of e-life review interviews, memory prompts, review extraction, mind space, and e-legacy products [22]. A preliminary study found that the program was acceptable, feasible, and promising in improving the psychospiritual well-being of patients with cancer [23]. Thus, this study aimed to robustly evaluate the effectiveness of the WeChat-based life review program in improving the psychospiritual well-being of patients with digestive system cancer using a 3-arm parallel randomized controlled trial.

Methods

Study Design and Setting

A randomized, controlled, single-blinded, 3-group pretest and repeated posttest experimental trial was conducted at the oncology department of a university-affiliated general hospital in Fujian, Southeast China. This study was performed in accordance with the CONSORT-EHEALTH checklist (Multimedia Appendix 1) [24] and was registered with the Chinese Clinical Trial Registry (ChiCTR-IOR-17011998).

Participants

Participants were recruited from June 2019 to October 2020, with follow-up ending in April 2021. Inclusion criteria were as follows: (1) diagnosis of digestive system cancer, (2) age ≥ 18 years, (3) awareness of diagnosis and treatment, (4) ability to use WeChat, and (5) no cognitive or verbal communication impairments. The exclusion criteria included (1) current use of anxiolytics or antidepressants, (2) participation in other psychotherapeutic programs, and (3) severe disability or diagnosis with a rapid-deterioration disease (Karnofsky performance status <40%).

Sample Size

Power analysis was used to estimate the sample size. Assuming a power of 90%, a 2-tailed test, and an effect size of 0.33 for anxiety and 0.43 for depression, 38 and 24 participants were needed to detect changes in anxiety and depression scores, respectively [25]. For hope (effect size 0.68) and self-transcendence (effect size 0.39), sample sizes of 10 and 28 participants were needed, respectively [26]. Anticipating a 20% attrition rate, we aimed to recruit 46 participants for each study group. A final total of 50 participants was recruited for each group.

Randomization and Blinding

A research assistant who was not involved in subject recruitment, data collection, or the interventions conducted the randomization schedule. A research randomizer website [27] was used to generate 150 nonrepeating random number sequences. The numbers ranged from 1 to 150, with 1 to 50, 51 to 100, and 101 to 150 corresponding to the life review group, control group 1, and control group 2, respectively. Each number was separately packaged in a sequentially sealed, opaque envelope to ensure allocation concealment. In this study, the recruited participants and facilitator (the first author) were not blinded to the group assignment; another research assistant, who was blinded to group allocation, conducted the data collection and analysis.

Interventions

All participants received routine care from medical staff at the oncology department. In addition, participants in the life review group received the 4-week WeChat-based life review program and those in control group 1 received the 4-week friendly visiting program.

Life Review Group

The life review group received the WeChat-based life review program along with routine care. The program consisted of a synchronous e-life review interview and asynchronous communication modules (Multimedia Appendix 2). The e-life review consisted of an individual, online, 40-to-60-minute nurse interview on WeChat, including 4 sections: present life (cancer experience); adulthood; childhood and adolescence; and summary of life. The asynchronous communication involved 4 modules. “Memory prompts” presented a set of images, music, videos, and audio-picture books relevant to each life section to trigger the participants’ memories and facilitate the life review process. “Review extraction” was a summary of meaningful events in which participants could view or leave comments. “Mind space” enabled participants to express emotions, hand down wishes, or reveal their true feelings to anyone who was important at that stage. “E-legacy product” was a digital booklet reflecting participants’ significant experiences, which they could transfer to their offspring.

The WeChat-based life review program was conducted weekly and facilitated by the first author, a registered nurse with more than 25 years of experience in clinical cancer care and 50 hours of life review training. Before the intervention, participants in the life review group installed WeChat and created a personal account. They accessed the memory prompts module to obtain an overview of the current session. Then, an e-life review interview was arranged by means of a video call with additional use of instant texts, voice messages, and emoticons. During the life review process, the facilitator monitored participants’ physical condition, emotional status, and responses to the guiding questions. Participants were also encouraged to access the asynchronous communication modules, which were available 24 hours a day, to freely review their interview content, express feelings and blessings, and provide important pictures and e-legacy products.

Control Group 1

Control group 1 received 4 sessions on an individual basis that provided social contact by engaging participants in daily conversation without reviewing the past. For consistency with the life review group, the 4 friendly visiting sessions were conducted by the first author through WeChat. Each visit lasted about 40 minutes, depending on the participants’ preference.

Control Group 2

Control group 2 received routine care, including drug treatment, nutritional support, symptom management, health education, and functional exercise.

Measures

A self-designed questionnaire by the first author was used to collect participants’ sociodemographic information and clinical characteristics. Sociodemographic data included age, gender, marital status, education, monthly income, and religion. Clinical characteristics included the specific diagnosis; the presence or absence of chronic disease and metastasis; the use or nonuse of surgery, chemotherapy, targeted therapy, radiotherapy, and immunotherapy; and Karnofsky performance status. Karnofsky performance status was used to evaluate participants’ physical function on an 11-point scale, with 0 indicating death and a score of less than 40% indicating severe disability and rapidly progressing disease. This study only included participants with a score higher than 40%.

Psychological outcomes included anxiety and depression symptoms; these were measured by the Hospital Anxiety and Depression Scale [28]. This is a 14-item scale divided into anxiety and depressive subscales, with each of 7 items rated on a 4-point Likert scale (higher scores represent increased anxiety or depression). The Chinese version of this scale has good sensitivity and specificity [29].

Spiritual outcomes consisted of hope and self-transcendence. The 12-item Herth Hope Scale [30] was used to assess participants’ hope on a 4-point Likert scale (range 12-48). Higher scores indicate higher levels of hope. The scale has been extensively used for assessment of hope in Chinese patients with cancer; it has a Cronbach α of .87 and a construct validity of .85, indicating good reliability and validity [31]. The 15-item self-transcendence scale assessed participants’ self-transcendence [32]. Each item was rated from 1 (“not at all”) to 4 (“almost always”), with the total score ranging from 15 to 60 and a higher score indicating a higher level of self-transcendence. The Chinese version of the scale has shown good reliability (Cronbach α=.83-.87) [33].

Data Collection and Analysis

A trained research assistant who was blinded to group assignments conducted all data collection. Outcome data were collected at baseline (T0) and 2 days (T1), 1 month (T2), and 6 months (T3) after the intervention. Statistical analysis was performed using R for Windows (version 3.5; R Foundation for Statistical Computing), with statistical significance set at $P<.05$. Normally distributed continuous variables were expressed as the mean and SD, nonnormally distributed continuous variables were presented as the median and range, and categorical variables were expressed as numbers (percentages). The Little test was used to check whether the missing data were missing completely at random. An intention-to-treat analysis was employed. Hypothesis testing used the chi-square test, the Mann-Whitney *U* test, the Fisher exact test, or a 1-way ANOVA to compare baseline data among groups. Since hierarchical linear models have more flexible data requirements and account for individual changes relative to group differences [34], they were employed for repeated measures. Both the baseline scores (intercepts) and change in scores (linear slopes) for each outcome within the groups were estimated in this model [35]. Time was represented as a dummy-coded variable to compare the outcomes at T1 to T0, T2 to T0, and T3 to T0. The effect of life review was examined based on the 3 dummy-coded time variables and the interaction effects between groups.

Ethical Considerations

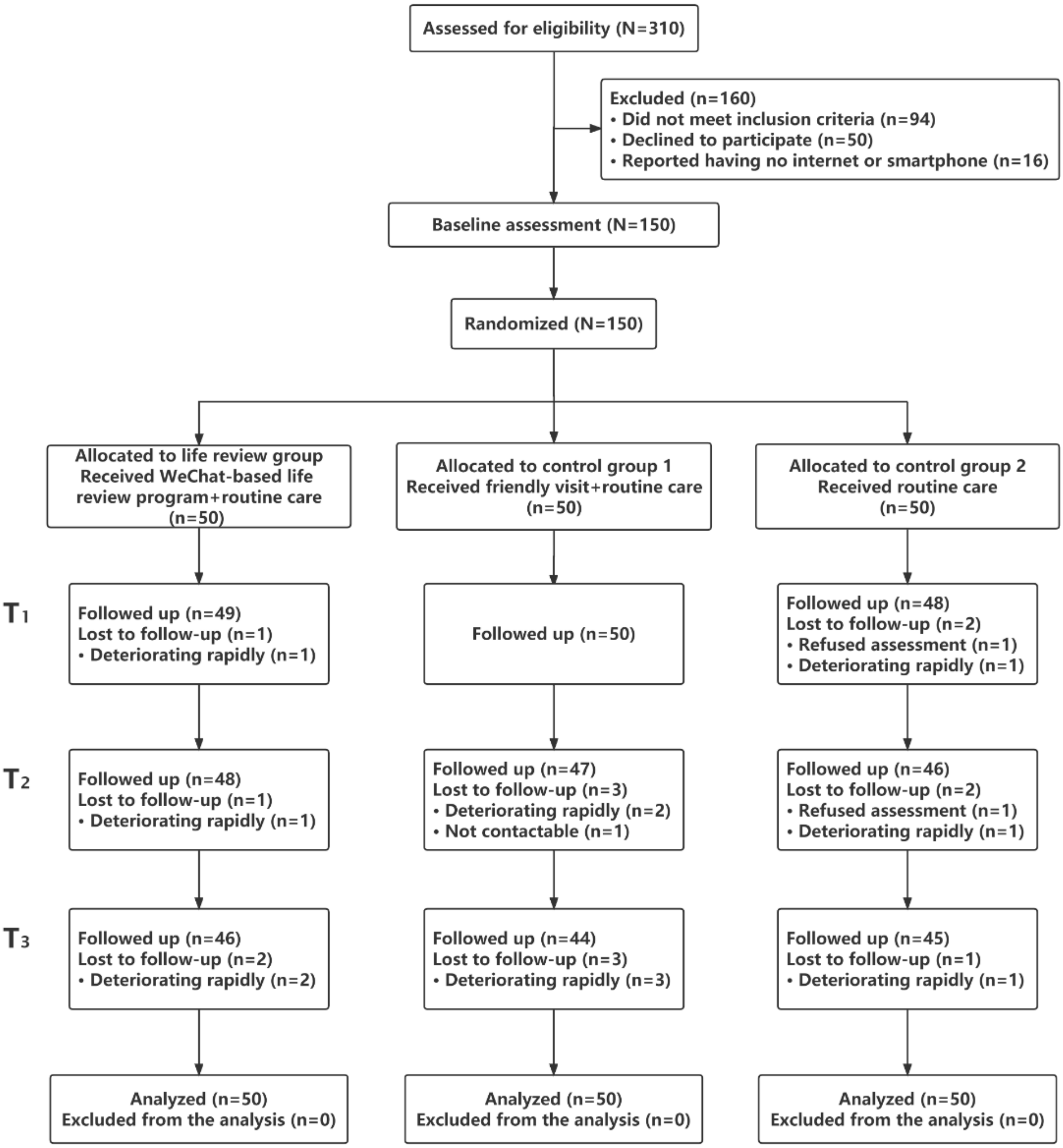
This study was approved by the Ethics Committee of Fujian Medical University (2016/00020) and the study hospital. All participants were provided with detailed information about the study, and written informed consent was obtained from each participant prior to data collection. Importantly, the data collected were kept confidential and anonymous and were used exclusively for this research.

Results

Participant Recruitment and Retention

During the period of this study, 310 patients with digestive system cancer were assessed for eligibility; only 150 patients met the inclusion criteria and consented to participate in this study. They were randomly allocated to 3 groups: the life review group (n=50), control group 1 (n=50), and control group 2 (n=50). Fifteen participants withdrew from the study because their disease progressed (n=12), they refused to participate (n=2), or because they could not be contacted (n=1). Four of these participants were from the life review group, 6 from control group 1, and 5 from control group 2. A final total of 135 participants completed the intervention and measurements. A flowchart of the study is shown in Figure 1.

Figure 1. Flowchart of the study.



Participant Characteristics

Table 1 shows the participants’ sociodemographic characteristics, clinical characteristics, and the baseline outcome variables across the study groups. The participants’ mean age was 58.48 (SD 9.96) years, and the majority were male (123/150, 82%), married (144/150, 96%), and affiliated with a religion (118/150, 78.7%). Less than half of the participants (74/150, 49.3%) had a primary school education level or lower and 57/150 (38%) had an average monthly household income per capita of RMB 1000 (US $148) or less. Among these participants, cancer in the digestive tract (114/150, 76%) was more common than cancer in the digestive glands (36/150, 24%); 48/150 (32%) patients had metastasis. Most patients had undergone surgery (124/150, 82.7%) or chemotherapy (102/150, 68%), and the average Karnofsky performance status was more than 60% (99/150, 66%). The groups’ demographic and clinical details were broadly comparable between the 3 groups. Interestingly, baseline anxiety and depression scores were up to 12% higher in control groups 1 and 2, but this difference could have arisen by chance alone (*P*=.89 and *P*=.17, respectively). Furthermore, this small difference would only have biased the overall estimate of effectiveness toward the null hypothesis, as it would have plausibly been slightly easier to reduce depression and anxiety scores in the control groups, as they started with a higher baseline.

Table 1. Baseline characteristics of participants.

Variables	Total (N=150)	Life review (n=50)	Control 1 (n=50)	Control 2 (n=50)	F/χ^2 (*df*)	*P* value
Age (years), mean (SD)	58.48 (9.96)	57.48 (9.29)	59.50 (10.67)	58.46 (9.96)	0.512	.60[a]
Gender, n (%)					1.084 (2)	.58[b]
Male	123 (82)	43 (35)[c]	41 (33)[c]	39 (38)[c]		
Female	27 (18)	7 (26)[c]	9 (33)[c]	11 (41)[c]		
Marital status, n (%)					1.261 (1)	.32[d]
Married	144 (96)	48 (96）	48 (96)	48 (96)		
Unmarried/ widowed/ divorced/ separated	6 (4）	2 (4）	2 (4)	2 (4)		
Monthly household income (US $), (n %)					6.128 (6)	.42[b]
≤15	57 (38)	19 (33)[c]	20 (35)[c]	18 (32)[c]		
15-44	33 (22)	8 (24)[c]	12 (36)[c]	13 (39)[c]		
44-88	36 (24)	12 (33)[c]	9 (25)[c]	15 (42)[c]		
>88	24 (16)	11 (46)[c]	9 (38)[c]	4 (17)[c]		
Religion, n (%)					1.986 (2)	.37[b]
Yes	118 (78.7)	41 (82）	36 (72）	41 (82）		
No	32 (21.3)	9 (18）	14 (28）	9 (18）		
Education level, n (%)					7.494 (6)	.28[b]
Primary school or below	74 (49.3）	25 (34)[c]	23 (31)[c]	26 (35)[c]		
Junior middle school	45 (30）	11 (24)[c]	15 (33)[c]	19 (42)[c]		
Senior high school	21 (14）	10 (47)[c]	7 (33)[c]	4 (19)[c]		
Tertiary or above	10 (6.7）	4 (40)[c]	5 (50)[c]	1 (10)[c]		
Chronic disease, n (%)					0.273 (2)	.87[b]
Yes	40 (26.7)	12 (24)	14 (28)	14 (28)		
No	110 (73.3)	38 (76)	36 (72)	36 (72)		
Diagnosis, n (%)					2.851 (2)	.24[b]
Cancer in digestive tract	114 (76)	35 (70)	42 (84)	37 (74)		
Cancer in digestive glands	36 (24)	15 (30)	8 (16)	13 (26)		
Metastasis, n (%)					2.206 (2)	.33[b]
Yes	48 (32)	18 (36)	18 (36)	12 (24)		
No	102 (68)	32 (64)	32 (64)	38 (76)		
Surgery, n (%)					4.001 (2)	.14[b]
Yes	124 (82.7)	37 (74)	44 (88)	43 (86)		
No	26 (17.3)	13 (26)	6 (12)	7 (14)		
Chemotherapy, n (%)					2.206 (2)	.33[b]
Yes	102 (68)	32 (64)	38 (76)	32 (64)		
No	48 (32)	18 (36)	12 (24)	18 (36)		
Targeted therapy, n (%)					2.542 (2)	.37[d]
Yes	6 (4)	1 (17)	4 (67)	1 (17)		

Variables	Total (N=150)	Life review (n=50)	Control 1 (n=50)	Control 2 (n=50)	F/χ^2 (*df*)	*P* value
No	144 (96)	49 (34)	46 (32)	49 (34)		
Radiotherapy, n (%)					2.990 (2)	.22[b]
Yes	21 (14)	7 (14)	10 (20)	4 (8)		
No	129 (86)	43 (86)	40 (80)	46 (92)		
Immunotherapy, n (%)					0.398 (2)	>.99[d]
Yes	9 (6）	3 (6)	3 (6)	3 (6)		
No	141 (94）	47 (94）	47 (94)	47 (94)		
Karnofsky performance status, n (%)					2.317 (2)	.35[b]
≤60	51 (34)	13 (26)	18 (35)	20 (39)		
>60	99 (66)	37 (37)	32 (32)	30 (60)		
Baseline outcome scores, mean (SD)						
Anxiety	3.81 (3.72)	3.74 (3.92)	3.68 (3.38)	4.02 (3.90)	0.118	.89[a]
Depression	4.13 (3.83)	3.60 (3.53)	5.06 (4.30)	4.02 (3.97)	1.817	.17[a]
Hope	36.57 (3.80)	36.76 (4.53)	36.50 (3.64)	36.46 (3.17)	0.091	.91[a]
Self-transcendence	45.97 (6.28)	46.40 (6.46)	46.02 (6.02)	45.44 (6.39)	0.295	.75[a]

[a] Calculated with ANOVA.

[b] Calculated with the X^2 test.

[c] The denominator used to calculate these percentages is the value for n in the “Total” column of the same row.

[d] Calculated with the Fisher exact test.

Effects on Outcome Variables

Table 2 shows the mean (SD) for the outcome variables at baseline and at the 3 follow-up sessions. A hierarchical linear model was employed to examine the change in outcome variables at each time point (Table 3). Overall, the interaction effects of the intervention on anxiety, depression, hope, and self-transcendence between groups over time were statistically significant.

Figure 2 shows the change over time in the mean (SD) scores for anxiety, depression, hope, and self-transcendence. Specific comparisons of outcome variables between groups at each time point and within groups are presented in Multimedia Appendix 3 and Multimedia Appendix 4. For anxiety, there was a significant decrease in the life review group at T1, T2, and T3 compared to baseline (*P*<.001, *P*<.001, and *P*=.002, respectively), indicating that the scores remained stable after the intervention. In the control groups, the anxiety score tended to show an overall upward trend, except for a decrease from baseline to T1 in control group 1. No significant difference in anxiety score was found between participants in the life review group and control group 1 at T1 (*P*=.80). However, the anxiety score was significantly lower in the life review group than in control group 1 at T2 and T3 (*P*=.02 for both). Compared with control group 2, the scores significantly decreased in the life review group at all follow-up sessions (*P*=.01, *P*=.02, and *P*=.01 at T1, T2, and T3, respectively).

A similar tendency was found in the depression score. There was a significant decrease in the life review group at all periods, and an increase in control group 2 from baseline (*P*=.02, *P*<.001, and *P*=.002 for T1, T2, and T3, respectively). As for control group 1, depression decreased significantly at T1 (*P*<.001) and increased at T2 and T3 (*P*=.07 and *P*=.37, respectively). The depression scores did not differ significantly between the life review group and control group 1 at T1 (*P*=.51), but depression was significantly lower in the life review group than in control group 1 at T2 and T3 (*P*=.003 and *P*<.001, respectively). There was also a significant difference in the depression score between the life review group and control group 2 at all follow-up sessions (*P*=.02 for both T1 and T2, *P*=.004 for T3).

A significant difference was observed in the hope score between the life review group and the 2 control groups at all follow-up sessions. Intragroup comparisons showed a significant increase in hope in the life review group after the intervention at T1, T2, and T3 (all *P*<.001). No significant differences were found over time in control group 1 (*P*=.55, *P*=.32, and *P*=.46 for T1, T2, and T3, respectively), while significant decreases were found in control group 2 at T1, T2, and T3 (*P*=.02 for both T1 and T2, *P*=.002 for T3).

In terms of self-transcendence, there was a significant difference between the life review group and the 2 control groups at T1, T2, and T3. Intragroup comparisons showed a significant increase in self-transcendence in the life review group after the intervention at T1, T2, and T3 (all *P*<.001). No statistically significant differences were found in self-transcendence over time for control group 1 (*P*=.46, *P*=.51, and *P*=.24 for T1, T2, and T3, respectively), while significant decreases were found in control group 2 at T1, T2, and T3 (P=.01, P=.04, P=.01 for T1, T2, and T3, respectively).

Table 2. Outcome variables at baseline and posttests (N=150; n=50 in each group).

Outcome variables	T0,[a] mean (SD)	T1,[b] mean (SD)	T2,[c] mean (SD)	T3,[d] mean (SD)
Anxiety				
Life review	3.74 (3.92)	2.84 (2.61)	3.00 (3.18)	2.98 (2.71)
Control 1	3.68 (3.39)	2.68 (2.55)	4.56 (2.43)	4.40 (2.37)
Control 2	4.02 (3.90)	4.50 (4.06)	4.58 (4.22)	4.50 (3.54)
Depression				
Life review	3.52 (3.40)	2.86 (2.80)	2.86 (2.87)	2.80 (1.91)
Control 1	4.84 (4.04)	3.32 (3.05)	5.26 (4.36)	5.08 (3.62)
Control 2	4.02 (3.97)	4.58 (4.39)	4.80 (4.66)	4.88 (4.51)
Hope				
Life review	36.76 (4.53)	38.58 (4.12)	38.12 (3.75)	38.04 (3.58)
Control 1	36.50 (3.64)	36.62 (3.49)	36.28 (3.31)	36.30 (3.54)
Control 2	36.46 (3.17)	35.98 (3.25)	35.94 (3.40)	35.60 (3.05)
Self-transcendence				
Life review	46.40 (6.46)	49.04 (5.80)	48.62 (6.03)	48.88 (5.44)
Control 1	46.08 (6.06)	46.30 (5.36)	45.86 (5.84)	45.26 (5.43)
Control 2	45.44 (6.39	44.68 (6.27)	44.74 (6.25)	44.52 (5.95)

[a] T0: baseline.

[b] T1: 2 days postintervention.

[c] T2: 1 month postintervention.

[d] T3: 6 months postintervention.

Table 3. Parameter estimates of the models with random intercept and random slope.

Model (random intercept and slope; fixed effects)	Anxiety				Depression				Hope				Self-transcendence			
	Estimate	Standard error	*t* test (*df*)	*P* value	Estimate	Standard error	*t* test (*df*)	*P* value	Estimate	Standard error	*t* test (*df*)	*P* value	Estimate	Standard error	*t* test (*df*)	*P* value
Intercept	4.020	0.468	8.599 (170.11)	<.001	4.020	0.526	7.642 (168.22)	<.001	36.460	0.508	71.817 (168.78)	<.001	45.440	0.842	53.992 (162.09)	<.001
LRG[a]	–0.280	0.661	–0.424 (170.11)	.67	–0.500	0.744	–0.672 (168.22)	.50	0.300	0.718	0.418 (168.78)	.68	0.960	1.190	0.807 (162.09)	.42
CG1[b]	–0.340	0.661	–0.514 (170.11)	.61	0.820	0.744	1.102 (168.22)	.27	0.040	0.718	0.056 (168.78)	.96	0.640	1.190	0.538 (162.09)	.59
T1[c]	0.480	0.204	2.354 (441)	.02	0.560	0.221	2.538 (441)	.01	–0.480	0.216	–2.227 (441)	.03	–0.760	0.301	–2.522 (441)	.01
T2[d]	0.560	0.204	2.747 (441)	<.001	0.780	0.221	3.535 (441)	<.001	–0.520	0.216	–2.413 (441)	.02	–0.700	0.301	–2.323 (441)	.02
T3[e]	0.480	0.204	2.354 (441)	.02	0.860	0.221	3.898 (441)	<.001	–0.860	0.216	–3.990 (441)	<.001	–0.920	0.301	–3.053 (441)	<.001
LRG:T1	–1.380	0.288	–4.786 (441)	<.001	–1.220	0.312	–3.910 (441)	<.001	2.300	0.305	7.546 (441)	<.001	3.400	0.426	7.977 (441)	<.001
CG1:T1	–1.480	0.288	–5.133 (441)	<.001	–2.080	0.312	–6.666 (441)	<.001	0.600	0.305	1.969 (441)	.05	0.980	0.426	2.299 (441)	.02
LRG:T2	–1.300	0.288	–4.509 (441)	<.001	–1.440	0.312	–4.615 (441)	<.001	1.880	0.305	6.168 (441)	<.001	2.920	0.426	6.851 (441)	<.001
CG1:T2	0.320	0.288	1.110 (441)	.27	–0.360	0.312	–1.154 (441)	.25	0.300	0.305	0.984 (441)	.33	0.480	0.426	1.126 (441)	.26
LRG:T3	–1.240	0.288	–4.301 (441)	<.001	–1.580	0.312	–5.064 (441)	<.001	2.140	0.305	7.021 (441)	<.001	3.400	0.426	7.977 (441)	<.001
CG1:T3	0.240	0.288	0.832 (441)	.41	–0.620	0.312	–1.987 (441)	.05	0.660	0.305	2.165 (441)	.03	0.100	0.426	0.235 (441)	.82
Random intercept variance	9.889	3.145	12.619	3.552	N/A[f]	N/A	N/A	N/A	11.726	3.424	N/A	N/A	33.145	5.757	N/A	N/A
Residual variance	1.039	1.019	1.217	1.103	N/A	N/A	N/A	N/A	1.161	1.078	N/A	N/A	2.271	1.507	N/A	N/A

[a] LRG: life review group.

[b] CG1: control group 1.

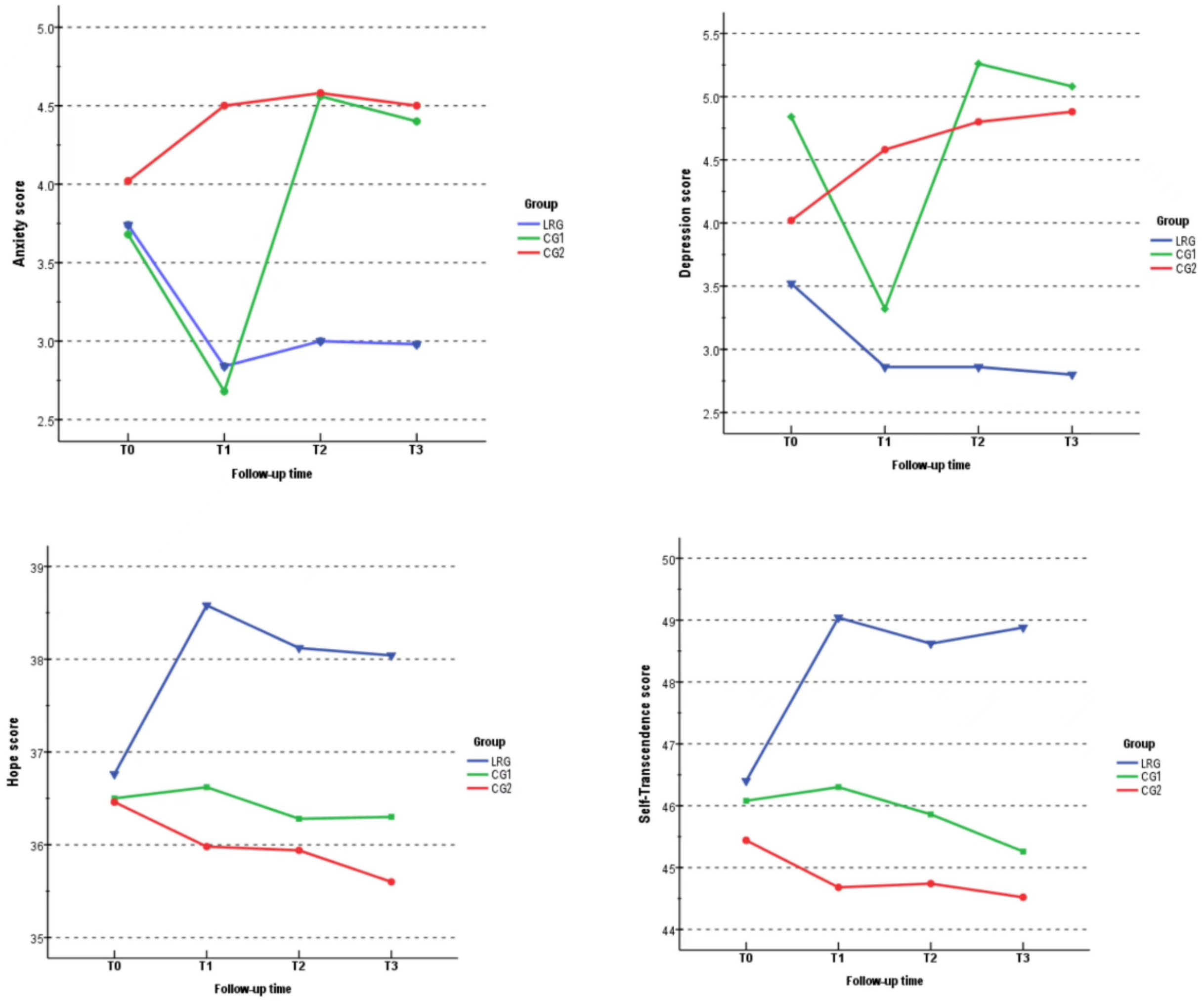
[c] T1: 2 days postintervention.

[d] T2: 1 month postintervention.

[e] T3: 6 months postintervention.

[f] N/A: not applicable.

Figure 2. Changes in the mean scores for anxiety, depression, hope, and self-transcendence in the patients over time. CG1: control group 1; CG2: control group 2; LRG: life review group; T0: baseline; T1: 2 days postintervention; T2: 1 month postintervention; T3: 6 months postintervention.



Discussion

Primary Findings

This is the first study to evaluate the effects of an online life review intervention on the psychospiritual well-being of patients with digestive system cancer; this study adopted a rigorous randomized controlled trial design with a very large sample size and long follow-up time. Our results show that the WeChat-based life review program could reduce anxiety and depressive symptoms and improve feelings of hope and self-transcendence among patients with digestive system cancer for a period of at least 6 months after the intervention. Our findings also revealed that friendly visiting might reduce anxiety symptoms in the short term; however, it did not improve depressive symptoms, feelings of hope, or self-transcendence among patients with digestive system cancer.

Participants' Adherence

Fifteen of 150 participants (10%) withdrew after 6 months of follow-up, which is an attrition rate lower than that of previous online life review studies [16,36]. Specifically, in the life review group, 46 out of 50 patients completed the whole program, indicating that the WeChat-based life review program was well implemented. A possible reason may be that life review is a naturally occurring, universal mental process among patients with cancer in the final life stage [37]. Patients with deteriorating health or low functionality can still participate in life review, even when lying in bed [38]. The low dropout rate may also be due to the simplicity, convenience, efficiency, and mobility of the WeChat-based program, especially among patients with cancer [39]. Considering the time and space limitations, it provided a range of ways to communicate, including text and pictures, voice messages, and video calls, allowing patients to participate in the program at any convenient time and location.

Patients' Outcomes

Among patients with digestive system cancer, anxiety symptoms significantly decreased from baseline up to 6 months after the life review intervention, which is consistent with previous research findings [40]. It is also worth noting that friendly visiting might help reduce anxiety symptoms in the short term, but not the long term. Scholars have put forward the idea that expression is an effective way of regulating emotions, allowing patients to feel supported by others, sort out their thoughts, and release negative emotions [41,42]. In this study, both friendly visiting and life review were conducted in a virtual, individual session where patients could feel safe and comfortable and reveal their innermost feelings in a familiar environment. Friendly visiting allowed patients to express their complaints about the disease and helped them divert their attention to other achievable things, leading to a temporary decrease in anxiety. Conversely, the WeChat life review program's long-term effectiveness may be due to opportunities for patients to retrieve positive thoughts, express and re-evaluate negative emotions, focus on the balance of positive and negative reminiscences, and integrate memories into a meaningful whole. Though painful memories may be picked up during the life review process, the facilitator offers guidance to consider these memories from other perspectives. Accordingly, patients are able to let go, accept, or even gain fresh insights into their lives, finally achieving self-integrity [8]. Meanwhile, the friendly visiting intervention focused on daily conversations without reviewing the past and with no guidance from the facilitator; thus, its effect on anxiety was unstable, with patients' anxiety scores going back to baseline or increasing 1 and 6 months after the intervention.

Our study results further confirmed the long-term benefits of the life review intervention for patients with digestive system cancer. The WeChat-based life review program significantly decreased depressive symptoms long-term, for at least 6 months after the intervention. This is in line with the findings of Lamers et al [43], who reported positive effects 6 months after the implementation of a similar program to ours among adults with moderate depressive symptoms.

The decrease in depression scores might be due to the accumulation of positive thoughts [44]. The WeChat-based life review program facilitated the retrieval of happy feelings from positive memories, prompted by means of appreciating images, music, videos, and audio-picture books. It also provided an opportunity for patients to learn from the past and affirm their contributions to families and society, which may induce positive emotions. On the other hand, reconciling negative experiences contributes to relieving depressive symptoms and improving one's emotional state [45]. In the process of life review, patients were encouraged to optimistically interpret the negative experiences in their own way to give positive meaning to the unpleasant stories, difficulties, and disappointments in their lives. From different perspectives, those negative experiences were reconstructed to bring about desired changes in the patients' views of themselves and their world. Finally, various life experiences were integrated into an acceptable whole and the patients moved toward acceptance of their lives.

Significant improvements in hope were also perceived among patients with cancer who participated in the WeChat-based life review program, consistent with previous research [11,46]. It might be that life review helps patients collate and learn from their pasts and reaffirm their contributions and accomplishments to their families and society, strengthening their awareness of their existence. During the life review process, patients may also perceive support from the facilitator and their family, since positive correlations have been observed between social support and increased hope [47]. Alternatively, patients could have set goals that matched their ability, making them more likely to be successful, thus increasing their feelings of hope. Through the life review intervention, patients become systematically aware of their life trajectory, gain a better understanding of their current situation, and take action congruent with their palliative situation. In addition, the e-legacy products may be beneficial to increase the patients' hope. A systematic review has reported that patients are in a positive state when reviewing their lives, especially when they view, touch, and appreciate the e-legacy product made in the life review process [8]; such feelings are maintained for a period of time.

Significant increases in self-transcendence were observed among patients with digestive system cancer who participated in the WeChat-based life review program, which is consistent with previous findings [23,48]. According to Reed [49], self-transcendence is an expansion of one's conceptual boundaries; inwardly, through introspective activities, outwardly, through concerns about others' welfare, and temporally, by integrating perceptions of one's past and future to enhance the present. The following reasons explain how our WeChat-based life review program could improve self-transcendence. First, during life review, patients recall and evaluate life experiences, and they are encouraged to express their feelings, reorganize their perspectives, and reconstruct the meaning of their lives, which can strengthen the inward domain of self-transcendence. Second, the WeChat-based life review program helps patients connect with their surroundings, which can improve outward transcendence by engaging in reciprocal relationships. While reviewing their lives, patients can reconsider and reflect on their connections to family and society, thereby discovering important emotional support around them. Third, the WeChat-based life review program integrates patients' past and future to improve their present, which is helpful in enhancing the temporal domain of self-transcendence. In sum, the WeChat-based life review program enabled patients to gradually focus on caring for others, transcending the present and achieving self-transcendence by introspection and gaining a harmonious view of the past, present, and future.

This study explored new possibilities for psychological interventions in oncology. The program took advantage of WeChat's increased availability and scalability to life review interventions, which is expected to overcome the obstacles of geographic distance and traffic issues, benefiting more patients, especially remotely located individuals. WeChat also promises to reduce personnel resources for delivery compared with face-to-face interventions. We recommend that future studies also examine the cost-effectiveness of this program, which could convince facilitators to engage in practice and integrate this intervention into transitional care in the community for patients with cancer.

Limitations

The limitations of this study should be noted. First, the WeChat-based life review program may not be suitable for people with poor literacy skills, because they may encounter difficulties in completing the 4 life review modules. Second, participant recruitment took place in only 1 hospital; hence, the generalizability of the findings may be limited. Multicenter and transregional research with a rigorous design is recommended in future research. In addition, this study covered multiple types of cancer diagnoses; future studies may consider selecting patients with the same cancer diagnosis.

Conclusions

Our WeChat-based life review program showed short- and long-term effectiveness in reducing anxiety and depressive symptoms and improving hope and self-transcendence among patients with digestive system cancer. By contrast, friendly visiting might reduce anxiety symptoms, but did not influence depression, hope, or self-transcendence. Accordingly, this WeChat-based life review program should be integrated into transitional care for digestive system cancer.

Acknowledgments

This study was supported by the Humanities and Social Science Fund of the Ministry of Education (grant 18YJA190017) and the Natural Science Foundation of Fujian Province (grant 2020J01644). We are grateful to all patients who participated in the study. We also thank the nurses and doctors from the study hospital's oncology department for their great support during this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 663 KB - jmir_v24i8e36000_app1.pdf]

Multimedia Appendix 2
Screenshot of life review.
[PNG File , 418 KB - jmir_v24i8e36000_app2.png]

Multimedia Appendix 3
Comparison of four outcome variables between groups at each time point.
[DOC File , 40 KB - jmir_v24i8e36000_app3.doc]

Multimedia Appendix 4
Comparison of four outcome variables within groups.
[DOC File , 37 KB - jmir_v24i8e36000_app4.doc]

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021 May;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Zamani M, Alizadeh-Tabari S. Anxiety and depression prevalence in digestive cancers: a systematic review and meta-analysis. BMJ Support Palliat Care 2021 Aug 20:bmjspcare-2021-003275. [doi: 10.1136/bmjspcare-2021-003275] [Medline: 34417285]
3. Fu J, Zeng Y, Tan Y, Fu B, Qiu H. Effects of hospice care on quality of life and negative emotions in patients with advanced tumor: A protocol for systematic review and meta-analysis. Medicine (Baltimore) 2020 Jul 02;99(27):e20784 [FREE Full text] [doi: 10.1097/MD.0000000000020784] [Medline: 32629661]
4. Connolly M, Timmins F. Spiritual Care for Individuals with Cancer: The Importance of Life Review as a Tool for Promoting Spiritual Well-Being. Semin Oncol Nurs 2021 Oct;37(5):151209 [FREE Full text] [doi: 10.1016/j.soncn.2021.151209] [Medline: 34465498]
5. Teo I, Krishnan A, Lee GL. Psychosocial interventions for advanced cancer patients: A systematic review. Psychooncology 2019 Jul;28(7):1394-1407. [doi: 10.1002/pon.5103] [Medline: 31077475]
6. Butler RN. The life review: an interpretation of reminiscence in the aged. Psychiatry 1963 Feb;26:65-76. [doi: 10.1080/00332747.1963.11023339] [Medline: 14017386]
7. Vuksanovic D, Dyck M, Green H. Development of a brief measure of generativity and ego-integrity for use in palliative care settings. Pall Supp Care 2015 Oct;13(5):1411-1415. [doi: 10.1017/S1478951515000206] [Medline: 25672884]
8. Zhang X, Xiao H, Chen Y. Effects of life review on mental health and well-being among cancer patients: A systematic review. Int J Nurs Stud 2017 Sep;74:138-148. [doi: 10.1016/j.ijnurstu.2017.06.012] [Medline: 28692880]

9. Wang C, Chow AY, Chan CL. The effects of life review interventions on spiritual well-being, psychological distress, and quality of life in patients with terminal or advanced cancer: A systematic review and meta-analysis of randomized controlled trials. Palliat Med 2017 Dec;31(10):883-894. [doi: 10.1177/0269216317705101] [Medline: 28488923]
10. Huang M, Wang R, Wang H. Effect of Life Review on Quality of Life in Terminal Patients: A Systematic Review and Meta-Analysis. J Nurs Res 2020 Apr;28(2):e83. [doi: 10.1097/JNR.0000000000000335] [Medline: 31246646]
11. Chen Y, Xiao H, Zheng J, Zhang X, Lin X. Effects of a mind map-based life review programme on psychospiritual well-being in cancer patients undergoing chemotherapy: A randomised controlled trial. Eur J Cancer Care (Engl) 2020 May;29(3):e13221. [doi: 10.1111/ecc.13221] [Medline: 31908102]
12. Lazar A, Thompson H, Demiris G. A systematic review of the use of technology for reminiscence therapy. Health Educ Behav 2014 Oct;41(1 Suppl):51S-61S [FREE Full text] [doi: 10.1177/1090198114537067] [Medline: 25274711]
13. Steindal SA, Nes AAG, Godskesen TE, Dihle A, Lind S, Winger A, et al. Patients' Experiences of Telehealth in Palliative Home Care: Scoping Review. J Med Internet Res 2020 May 05;22(5):e16218 [FREE Full text] [doi: 10.2196/16218] [Medline: 32369037]
14. Wongvibulsin S, Habeos EE, Huynh PP, Xun H, Shan R, Porosnicu Rodriguez KA, et al. Digital Health Interventions for Cardiac Rehabilitation: Systematic Literature Review. J Med Internet Res 2021 Feb 08;23(2):e18773 [FREE Full text] [doi: 10.2196/18773] [Medline: 33555259]
15. Wise M, Marchand L, Aeschlimann E, Causier D, Cleary J. Integrating a narrative medicine telephone interview with online life review education for cancer patients: lessons learned and future directions. J Soc Integr Oncol 2009;7(1):19-25 [FREE Full text] [Medline: 19476731]
16. Wise M, Marchand LR, Roberts LJ, Chih M. Suffering in Advanced Cancer: A Randomized Control Trial of a Narrative Intervention. J Palliat Med 2018 Feb;21(2):200-207 [FREE Full text] [doi: 10.1089/jpm.2017.0007] [Medline: 29135330]
17. Dang M, Noreika D, Ryu S, Sima A, Ashton H, Ondris B, et al. Feasibility of Delivering an Avatar-Facilitated Life Review Intervention for Patients with Cancer. J Palliat Med 2021 Apr;24(4):520-526. [doi: 10.1089/jpm.2020.0020] [Medline: 32896200]
18. Ryu S, Price SK. Embodied storytelling and meaning-making at the end of life: VoicingHan avatar life-review for palliative care in cancer patients. Arts Health 2021 Jun 23:1-15. [doi: 10.1080/17533015.2021.1942939] [Medline: 34160335]
19. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]
20. Thomas L, Briggs P. Reminiscence through the Lens of Social Media. Front Psychol 2016;7:870 [FREE Full text] [doi: 10.3389/fpsyg.2016.00870] [Medline: 27378971]
21. 9 Powerful WeChat Statistics You Need to Know in 2022. Brew Interactive. URL: https://blog.brewinteractive.com/wechat-statistics [accessed 2022-05-20]
22. Zhang X, Xiao H. Development and evaluation of a WeChat-based life review programme for patients with cancer: protocol for a randomised controlled trial. BMJ Open 2018 Dec 06;8(12):e020239 [FREE Full text] [doi: 10.1136/bmjopen-2017-020239] [Medline: 30530470]
23. Zhang X, Xiao H, Chen Y. Evaluation of a WeChat-based life review programme for cancer patients: A quasi-experimental study. J Adv Nurs 2019 Jul;75(7):1563-1574. [doi: 10.1111/jan.14018] [Medline: 30950535]
24. 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]
25. Steinhauser KE, Alexander SC, Byock IR, George LK, Olsen MK, Tulsky JA. Do preparation and life completion discussions improve functioning and quality of life in seriously ill patients? Pilot randomized control trial. J Palliat Med 2008 Nov;11(9):1234-1240. [doi: 10.1089/jpm.2008.0078] [Medline: 19021487]
26. Chen SH. Effectiveness of life review in increasing self-transcendence, hope and spiritual well-being of patients with cancer. Department of Nursing, Fooying University. URL: https://hdl.handle.net/11296/f97c4e [accessed 2022-08-12]
27. Research Randomizer: Random Sampling and Random Assignment Made Easy!. Research Randomizer. URL: https://www.randomizer.org [accessed 2022-08-12]
28. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
29. Annunziata MA, Muzzatti B, Bidoli E, Flaiban C, Bomben F, Piccinin M, et al. Hospital Anxiety and Depression Scale (HADS) accuracy in cancer patients. Support Care Cancer 2020 Aug;28(8):3921-3926. [doi: 10.1007/s00520-019-05244-8] [Medline: 31858249]
30. Herth K. Development and refinement of an instrument to measure hope. Sch Inq Nurs Pract 1991;5(1):39-51; discussion 53. [Medline: 2063043]
31. An N, Ye J, Huang Q. Correlation between levels of hope and coping styles of oral cancer patients after surgery (In Chinese). Chinese Gen Pract Nurs 2016:331-334 [FREE Full text]
32. Reed PG. Self-transcendence and mental health in oldest-old adults. Nurs Res 1991;40(1):5-11. [Medline: 1987557]

33. Zhang J, Sun JP, Zhang L. The reliability and validity of the Chinese version of the self-transcendence scale in the elderly (in Chinese). Chinese Journal of Gerontology 2014 Apr;34(07):1910-1911 [FREE Full text]
34. Shin JH, Shin I. Investigation of Longitudinal Data Analysis: Hierarchical Linear Model and Latent Growth Model Using a Longitudinal Nursing Home Dataset. Res Gerontol Nurs 2019 Nov 01;12(6):275-283. [doi: 10.3928/19404921-20191024-02] [Medline: 31755964]
35. Ventura F, Sawatzky R, Öhlén J, Karlsson P, Koinberg I. Challenges of evaluating a computer-based educational programme for women diagnosed with early-stage breast cancer: a randomised controlled trial. Eur J Cancer Care (Engl) 2017 Sep;26(5):e12534. [doi: 10.1111/ecc.12534] [Medline: 27339805]
36. Westerhof G, Lamers S, Postel M, Bohlmeijer E. Online Therapy for Depressive Symptoms: An Evaluation of Counselor-Led and Peer-Supported Life Review Therapy. Gerontologist 2019 Jan 09;59(1):135-146. [doi: 10.1093/geront/gnx140] [Medline: 28961958]
37. Kwan CWM, Ng MSN, Chan CH. The use of life review to enhance spiritual well-being in patients with terminal illnesses: An integrative review. J Clin Nurs 2017 Dec;26(23-24):4201-4211. [doi: 10.1111/jocn.13977] [Medline: 28748593]
38. Ando M, Tsuda A, Morita T. Life review interviews on the spiritual well-being of terminally ill cancer patients. Support Care Cancer 2007 Feb;15(2):225-231. [doi: 10.1007/s00520-006-0121-y] [Medline: 16967303]
39. Lin J, Li H. The Impact of Wechat Platform Interaction on Anxiety and Depression of Cancer Patients After Discharge (in Chinese). China Health Standard Management 2016;7(05):13-15 [FREE Full text]
40. Thronicke A, Kröz M, Merkle A, Matthes H, Herbstreit C, Schad F. Psychosocial, Cognitive, and Physical Impact of Elaborate Consultations and Life Review in Female Patients with Non-Metastasized Breast Cancer. Complement Med Res 2018;25(2):92-101 [FREE Full text] [doi: 10.1159/000486672] [Medline: 29656290]
41. Mikkelsen MB, Elkjær E, Mennin DS, Fresco DM, Zachariae R, Applebaum A, et al. The impact of emotion regulation therapy on emotion differentiation in psychologically distressed caregivers of cancer patients. Anxiety Stress Coping 2021 Jul;34(4):479-485 [FREE Full text] [doi: 10.1080/10615806.2021.1929934] [Medline: 34047220]
42. Brandão T, Tavares R, Schulz MS, Matos PM. Measuring emotion regulation and emotional expression in breast cancer patients: A systematic review. Clin Psychol Rev 2016 Feb;43:114-127. [doi: 10.1016/j.cpr.2015.10.002] [Medline: 26520599]
43. Lamers SMA, Bohlmeijer ET, Korte J, Westerhof GJ. The efficacy of life-review as online-guided self-help for adults: a randomized trial. J Gerontol B Psychol Sci Soc Sci 2015 Jan;70(1):24-34. [doi: 10.1093/geronb/gbu030] [Medline: 24691155]
44. Dainer-Best J, Shumake JD, Beevers CG. Positive imagery training increases positive self-referent cognition in depression. Behav Res Ther 2018 Dec;111:72-83 [FREE Full text] [doi: 10.1016/j.brat.2018.09.010] [Medline: 30321746]
45. Cova F, Garcia F, Oyanadel C, Villagran L, Páez D, Inostroza C. Adaptive Reflection on Negative Emotional Experiences: Convergences and Divergence Between the Processing-Mode Theory and the Theory of Self-Distancing Reflection. Front Psychol 2019;10:1943 [FREE Full text] [doi: 10.3389/fpsyg.2019.01943] [Medline: 31551854]
46. Wang YC, Xu YR. Effects on self-esteem and hope among cervical cancer patients with life review intervention (in Chinese). Modern Nurse 2015:73-75 [FREE Full text]
47. Goldzweig G, Baider L, Andritsch E, Rottenberg Y. Hope and social support in elderly patients with cancer and their partners: an actor-partner interdependence model. Future Oncol 2016 Dec;12(24):2801-2809. [doi: 10.2217/fon-2016-0267] [Medline: 27712084]
48. Chen Y, Zheng J, Xiao H, Lin H, Zhang X. Effects of a Mind Map-Based Life Review Program on Anxiety and Depressive Symptoms on Cancer Patients Undergoing Chemotherapy: A Randomized Controlled Trial. Cancer Nurs 2022;45(1):E116-E123. [doi: 10.1097/NCC.0000000000000874] [Medline: 32740327]
49. Reed PG. Demystifying self-transcendence for mental health nursing practice and research. Arch Psychiatr Nurs 2009 Oct;23(5):397-400. [doi: 10.1016/j.apnu.2009.06.006] [Medline: 19766931]

Abbreviations

T0: baseline
T1: 2 days postintervention
T2: 1 month postintervention
T3: 6 months postintervention

Edited by R Kukafka; submitted 06.03.22; peer-reviewed by T Tillmann, P Chow; comments to author 07.05.22; revised version received 22.05.22; accepted 29.07.22; published 25.08.22.

Please cite as:
Zheng M, Zhang X, Xiao H
Effects of a WeChat-Based Life Review Program for Patients With Digestive System Cancer: 3-Arm Parallel Randomized Controlled Trial
J Med Internet Res 2022;24(8):e36000
URL: https://www.jmir.org/2022/8/e36000
doi:10.2196/36000
PMID:36006665

Original Paper

Chronic Kidney Disease Awareness Campaign and Mobile Health Education to Improve Knowledge, Quality of Life, and Motivation for a Healthy Lifestyle Among Patients With Chronic Kidney Disease in Bangladesh: Randomized Controlled Trial

Mohammad Habibur Rahman Sarker[1,2], MBBS, MPH, PhD; Michiko Moriyama[2], RN, MSN, PhD; Harun Ur Rashid[3], MBBS, PhD, FRCP, FCPS; Md Moshiur Rahman[2], MBBS, MPH, PhD; Mohammod Jobayer Chisti[1], MBBS, MMed, PhD; Sumon Kumar Das[4], MBBS, PhD; Samir Kumar Saha[5,6], BSc, MSc, PhD; Shams El Arifeen[7], MBBS, MPH, DrPH; Tahmeed Ahmed[1], MBBS, PhD; A S G Faruque[1], MBBS, MPH

[1]Nutrition and Clinical Services Division, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh
[2]Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan
[3]Kidney Foundation Hospital and Research Institute, Dhaka, Bangladesh
[4]Menzies - School of Health Research, Charles Darwin University, Darwin, Australia
[5]Child Health Research Foundation, Dhaka, Bangladesh
[6]Dhaka Shishu Hospital, Dhaka, Bangladesh
[7]Maternal and Child Health Division, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh

Corresponding Author:
Mohammad Habibur Rahman Sarker, MBBS, MPH, PhD
Nutrition and Clinical Services Division
International Centre for Diarrhoeal Disease Research, Bangladesh
68, Shaheed Tajuddin Ahmed Sarani, Mohakhali
Dhaka, 1212
Bangladesh
Phone: 880 1775170533
Email: habibur.rahman@icddrb.org

Abstract

Background: Chronic kidney disease (CKD) is linked to major health consequences and a poor quality of life. Despite the fact that CKD is becoming more prevalent, public knowledge of the disease remains low.

Objective: This study aimed to evaluate the outcome of a health education intervention designed to enhance knowledge, health-related quality of life (QOL), and motivation about healthy lifestyle among adults with CKD.

Methods: This study was a parallel-group (1:1), randomized controlled trial in the Mirzapur subdistrict of Bangladesh that compared 2 groups of patients with CKD. Adults with CKD (stages 1-3) were enrolled in November 2020 and randomly assigned the intervention or control group. The intervention group received health education through a CKD awareness campaign and mobile health technologies and was observed for 6 months, whereas the control group received standard treatment. The primary outcome was the evaluation of improved scores on the CKD knowledge questionnaire, and the secondary outcomes were improved QOL and changes in the levels of blood pressure (BP), BMI, serum creatinine, fasting blood sugar (FBS), hemoglobin, cholesterol, high-density lipoprotein cholesterol, triglyceride, serum uric acid, blood urea nitrogen (BUN), and albumin-to-creatinine ratio.

Results: The study enrolled 126 patients (control: n=63; intervention: n=63) and performed intention-to-treat analysis. The analyses included repeated measures ANOVA, and the results were observed to be significantly different from within groups (*P*<.001), between groups (*P*<.001), and the interaction of group × time factor (*P*<.001) for knowledge score. Diastolic BP and BMI showed significant differences arising from within groups (*P*<.001 and *P*=.01, respectively) and the interaction of group × time factor (*P*=.001 and *P*=.02, respectively); food salinity and hip circumferences showed significant differences arising from within groups (*P*=.001 and *P*=.03, respectively) and between groups (*P*=.001 and *P*=.02, respectively). Moreover, systolic BP and waist circumference showed significant differences from within groups (*P*<.001 and *P*=.003, respectively). However, no significant differences were found arising from within groups, between groups, and the interactions of group × time for QOL, urine salinity, and mid-upper arm circumference. Regarding the laboratory findings, from baseline to 6 months, the mean (SD) FBS decreased by 0.51 (3.77) mmol/L in the intervention group and 0.10 (1.44) mmol/L in the control group (*P*=.03); however, blood urea nitrogen increased by 3.64 (7.17) mg/dL in the intervention group and 1.68 (10.10) mg/dL in the control group (*P*=.01).

Conclusions: The health education strategy, which included a campaign and mobile health, showed promise for enhancing CKD knowledge among patients with CKD. This strategy may also aid patients with CKD in controlling their FBS and BP. The combined health education initiatives give evidence for scaling them up in Bangladesh and possibly other low- and middle-income countries, particularly in rural and peri-urban settings.

Trial Registration: ClinicalTrials.gov NCT04094831; https://clinicaltrials.gov/ct2/show/NCT04094831.

International Registered Report Identifier (IRRID): RR2-10.2196/30191

(J Med Internet Res 2022;24(8):e37314) doi:10.2196/37314

KEYWORDS

Bangladesh; health education; health knowledge; quality of life; motivation; randomized controlled trial; RCT; campaign; chronic kidney disease; knowledge; mobile health; mHealth; kidney; chronic disease; chronic condition; patient education; patient knowledge; low- and middle-income countries; LMIC

Introduction

Chronic kidney disease (CKD) is responsible for poor health outcomes, low quality of life (QOL), and high health care expenses [1]. Globally, the rising trend of CKD is being recognized as a future public health threat [2]. The prevalence of CKD in stages 1-3 has been documented at 8.9% of the global population [3], with rates higher in low-income nations such as India (15.6%) [4] and Bangladesh (21.33%) [5]. Early stage CKD is generally asymptomatic, and diagnosis is usually made through serum creatinine and albumin-to-creatinine ratio tests [6]. If left undetected and untreated, CKD can proceed to end-stage renal disease, which requires expensive renal replacement therapy such as dialysis or kidney transplantation to save the patient's life [7]. Over the last decade, the economic burden of renal replacement therapy has increased dramatically and is substantially higher in low- and middle-income countries (LMIC) than in high-income countries [8]. Although primary renal disease causes CKD, the great majority of patients with CKD have concomitant health conditions such as diabetes, hypertension, and older age [9]. The majority of individuals with nonprimary renal disease are treated for associated risk factors such as diabetes and hypertension rather than the CKD itself [10]. Glomerular filtration rate declines slowly in most patients with CKD in stages 1-3; however, the declining trend varies among individuals and is influenced by a variety of factors such as diabetes, high blood pressure (BP), and older age, etc [6]. Individuals with CKD who receive proper information and knowledge about CKD and its risk factors are more likely to engage in health-promoting behaviors and lifestyle changes [11]. CKD early diagnosis and prevention strategies, such as a CKD preventive campaign, are currently being applied in a number of high-income countries. Increasing knowledge about CKD and its risk factors is a crucial strategy for slowing the disease' progression.

In Bangladesh, community health workers (CHWs) are health cadres who work in public-sector health facilities. In recent years, CHWs in some places have begun to provide services for noncommunicable diseases, including health education and counseling [12]. The use of CHWs to deliver an education and monitoring intervention has been found to be effective in noncommunicable diseases such as reducing BP and has the potential to be scaled up in resource-limited settings [13,14]. They can make a substantial difference in the health of patients living with CKD. CHWs can educate patients to help protect their own kidneys and improve their QOL [15]. Mobile health (mHealth) is still in its implementation phase in the field of nephrology. However, mobile phone call–based health education has great potential to provide CKD knowledge and improve QOL, because it relies on basic mobile technology and requires limited literacy or numeracy skills [15]. A nephrologist-facilitated CKD health campaign also has the potential to improve patients' knowledge and awareness [16,17].

The education of patients with CKD may increase perceived kidney disease knowledge among patients, improve QOL, and delay the progression of kidney disease [18,19]. A community-based screening revealed a high prevalence of CKD in stages 1-3, with only around 7% of the people being aware of their condition prior to the study; however, no health education intervention for these stages was investigated [5]. Most studies on the education of patients with CKD have focused on individuals with end-stage disease and shown improved outcomes [11]. However, a CKD campaign and mHealth-based health education in the early stages could be an integral part of patient management and the reduction of the related risk factors, slowing down the progression of the disease, and the need for such education is greater in rural and peri-urban areas. Thus, this study aimed to evaluate the outcome of a health education intervention designed to enhance knowledge, health-related QOL, and motivation about healthy lifestyle among rural and peri-urban adults with CKD (stages 1-3).

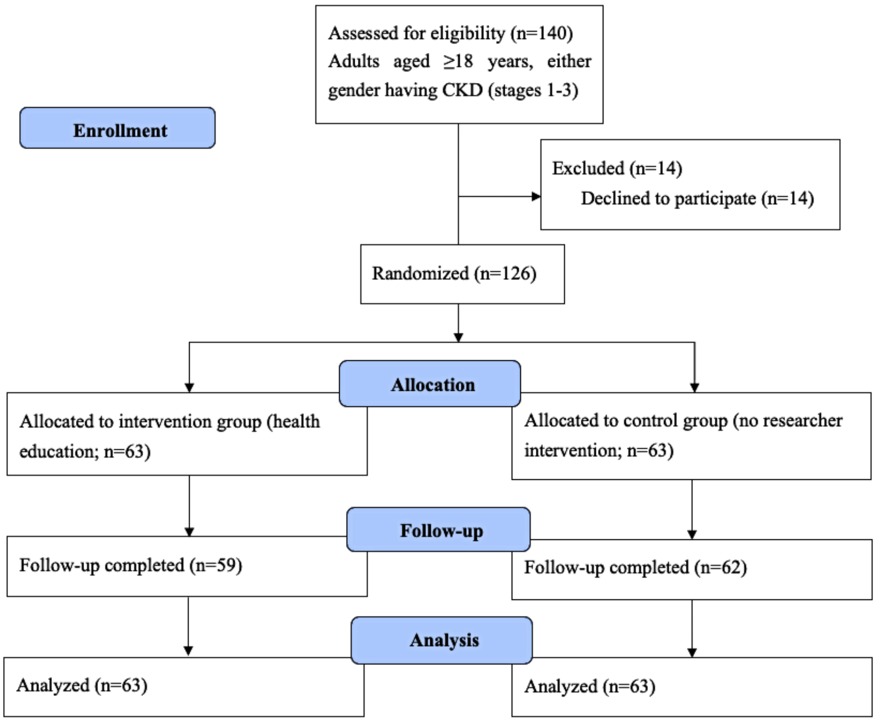
Methods

Design

This study was a community-based, single-center, prospective, open-label, parallel-group (1:1), randomized control trial (RCT) involving patients with CKD, conducted in a rural and peri-urban population of Bangladesh. This study was designed in accordance with the CONSORT (Consolidated Standards of Reporting Trials) [20] and SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) [21] guidelines.

The study flowchart is shown in Figure 1. The total study duration was 1 year; during that 1 year, the intervention duration was 6 months, starting from mid-November 2020 and completed in May 2021.

Figure 1. Flowchart of the participants throughout the study. CKD: chronic kidney disease.



Study Population and Sampling

This study was conducted in the demographic surveillance system (DSS) area of the Mirzapur subdistrict under the Tangail district, which is 60 km north of Dhaka, the capital city of Bangladesh. The details of the Mirzapur DSS have been described in a previous publication [5]. The research team completed a CKD prevalence study before implementing this RCT [5]. The Chronic Kidney Disease Epidemiology Collaboration equation was used to estimate glomerular filtration rate [22], and the National Kidney Foundation Kidney Disease Outcomes Quality Initiative guidelines were used to define and indicate the stages of CKD [23]. We randomly enrolled study patients from the residents of the DSS who have been diagnosed with CKD. The inclusion and exclusion criteria are shown in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion Criteria

- Adults (aged ≥18 years) of either gender, diagnosed with chronic kidney disease in stages 1-3 (any one of the following: estimated glomerular filtration rate = 30-59 mL/min/1.73 m^2 and albumin-to-creatinine ratio ≥30 mg/g for more than 3 months) [24]
- Lived continuously in the locality for at least 5 years
- Has a personal cell phone or access to a shared phone
- Gave written informed consent to participate in the study

Exclusion Criteria

- Individuals diagnosed with chronic kidney disease in stages 4 and 5
- Hospitalized at the time of enrollment
- Having any known, serious illness with questionable prognosis; for example, malignancy, mental illness, congenital disease, or gross physical disability (if they have prescriptions)
- Declined to give consent to participate in the study

Randomization

Permuted block randomization technique was performed using a block size of 6 based on a computer-generated random number sequence. An experienced statistician, who is not involved in the study in any way, prepared the randomization table and listed the study patients’ numbers with the corresponding intervention allocations for patients with CKD in sequentially numbered, sealed envelopes according to the randomization schedule to correspond to the serial numbers of the patients with CKD. These envelopes were kept in an office locker. Allocation was concealed in identical sealed envelopes that were only opened when the study patient was ready for enrollment under the supervision of the principal investigator. This step took place after a patient with CKD had been enrolled in the study, after obtaining voluntary informed written consent and having been assigned a study number.

Study Activities and Contents

CHWs conducted baseline home visits during the first week and received written informed consent before interviewing research patients and administering field-tested standardized questionnaires to collect information on socioeconomic and demographic characteristics, level of knowledge, and QOL. Next, they advised the study patients to visit Kumudini Women's Medical College and Hospital's laboratory for study-related investigations. After evaluating the study patients' eligibility, the statistician performed the randomization. The responsible research staff then opened the sealed envelope, revealing the patient's random allocation, and handed the patient's ID number to the CHWs. Staff in this trial were not blinded to the intervention or control groups; however, they were blinded to the baseline interview. The CHWs called the enrolled trial patients assigned to the intervention group and invited them to the CKD awareness campaign.

In addition to baseline, at 3 and 6 months, the CHWs visited the patients' homes for an interview and physical examination using the same questionnaire. After 6 months, the CHWs advised the study patients to visit the laboratory for study-related investigations.

During the home visits, the CHWs used a Portable Health Clinic box with the essential diagnostic equipment for this research (the details were described in our previous paper) [25]. At the baseline, they collected sociodemographic information such as age, gender, marital status, religion, occupation, educational background, income per month, patient's current medical history including medication use, past medical history, and family history (3 generations) including current and immediate past medical history. The same information was collected at 3 and 6 months in case of any changes from the baseline. They also administered a questionnaire on CKD knowledge and QOL. Physical examinations were performed to measure BP, pulse, height, weight, waist and hip circumferences, and mid-upper arm circumference (MUAC). In addition, blood and spot urine samples were taken to estimate the kidney function status and their related risk factors such as serum creatinine, fasting blood sugar (FBS), hemoglobin, cholesterol, high-density lipoprotein cholesterol, triglyceride, serum uric acid, blood urea nitrogen (BUN), and albumin-to-creatinine ratio (Multimedia Appendix 1).

Intervention Group

A CKD awareness campaign and mHealth technologies were used to provide health education to the intervention group.

CKD Awareness Campaign

A nephrologist conducted the CKD awareness campaign in the native language (Bengali). Important information related to CKD, such as the basics of kidney function and kidney diseases, stages, risk factors, and preventive measures were discussed during the half day campaign. Health education materials (leaflet, short textbook, and recording notebook) were provided to the study patients. The details of the CKD campaign have been discussed in the protocol paper [25].

mHealth Technology

A nephrologist trained the CHWs to conduct the mHealth education. The CHWs were trained over 3 days. The training was facilitated by a nephrologist and coordinated by the project's principal investigator. The program included lectures, discussions, question-and-answer sessions, and role-playing. We developed training materials that were tied to the study objectives to ensure that the CHWs acquired the skills needed to deliver the education to the target patients. The CHWs conducted the health education over a mobile phone call using mHealth technology once every 2 weeks during the intervention period. They discussed with the study participants in the intervention group about fundamental kidney disease, risk factors, and CKD prevention methods. Over a 10-minute session, the patients were free to discuss their health-related concerns with the CHWs. The details of the contents have been described in the protocol paper [25] (Multimedia Appendix 2).

BP Check

During the intervention phase, CHWs performed weekly home visits to the patients belonging to the intervention group and measured their BP.

Control Group

The control group received usual care and was observed without intervention throughout the trial period. At 3 and 6 months, an interview and physical examination were conducted for these patients. Furthermore, they returned to the laboratory for study-related investigations after 6 months.

Sample Size

The sample size was calculated based on the primary outcome—the enhanced knowledge of the study patients following the intervention of the health education program. We assumed that the proportion of existing knowledge among patients with CKD (stages 1-3) was 40% [26], with the percentage of predicted knowledge increasing to 70% following the intervention. As a result, with 90% power and 20% loss to follow-up, the total sample size was 126 (63 in each group). The details of the sample size calculation have been described in the protocol paper [25].

Study Outcomes

The primary and secondary outcomes were measured at baseline, 3 months, and 6 months for both the intervention and control groups (Textbox 2). The laboratory parameters were measured at baseline and 6 months.

Textbox 2. Primary and secondary outcomes.

Primary Outcome

- Evaluation of improved scores on the Chronic Kidney Disease Knowledge Questionnaire [26]

Secondary Outcomes

- Improved quality of life; measured by the EQ-5D-5L quality of life questionnaire [27]
- Changes in the levels of blood pressure, BMI, serum creatinine, fasting blood sugar, hemoglobin, cholesterol, high-density lipoprotein cholesterol, triglyceride, serum uric acid, blood urea nitrogen, and albumin-to-creatinine ratio

Measurements of Knowledge and QOL

The evaluation of knowledge was measured using the Kidney Knowledge questionnaire, a 24-item scale designed to assess the CKD knowledge of patients in stages 1-3. A more in-depth description of the method has been mentioned elsewhere [25]. QOL was measured using the standardized EQ-5D-5L questionnaire [28]. In this study, we used the Japanese region’s tariff to define the standard value (tariff) for EQ-5D-5L to assess the impact of the health care interventions on QOL.

Statistical Analysis

In this study, intention-to-treat analysis was used to compare the outcomes of the intervention and control groups. Missing data were dealt with by using the last observation carried forward method [29]. All baseline indicators at the time of registration were analyzed to ensure the comparability of the randomized samples. For the baseline assessment, continuous variables were compared by 2-tailed *t* test or Mann-Whitney *U* test, and categorical variables were compared by chi-square test. Multiple comparisons were performed by 2-way repeated measures ANOVA for the evaluation of the outcome variables such as CKD knowledge questionnaire, physical measurements, and QOL at baseline, 3 months, and 6 months. In addition, generalized estimating equations were used to estimate the effect of the health education after adjusting for relevant covariates. However, outcome variables for laboratory findings were measured at baseline and 6 months. Changes in the laboratory profiles were compared between the intervention group and control group by 2-tailed *t* test or Mann-Whitney *U* test after checking the data normality. The significance level was set at $P<.05$. The statistical analyses were conducted using SPSS statistical software (version 20.0; IBM Corp).

Consent to Participate

All study patients provided written informed consent and participated entirely voluntarily. The patients received a copy of the consent form. Each study patient received written information about the aim of the study and was informed that they could decide to leave the study at any time and for any reason. Patients in the intervention group were informed that they would get additional health education alongside the usual care, whereas patients in the control group were informed that they would not receive any interventions. Patients were assured that their information would not be disclosed, and they were informed about the use of data for analysis and the use of the results for enhancing patient care activities, conducting research, and publication without disclosing their name or identity.

Ethics Approval

Ethical approval was obtained by the Research Review Committee and Ethical Review Committee of the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) (PR 19081). The study was registered at ClinicalTrials.gov (NCT04094831).

Results

A total of 126 patients (control group: n=63; intervention group: n=63) were enrolled in the study. Of the 126 patients, 5 withdrew from further participation during the follow-up period—4 from the intervention group and 1 from the control group. Of the 4 patients who withdrew in the intervention group, 2 did not participate in the health campaign, and the other 2 did not continue their health education. The patient in the control group, on the other hand, left the area after enrollment into the study. The analyzable study patients, however, still comprised 126 patients, including 63 receiving mHealth education and 63 receiving usual care. Among these study patients, the mean (SD) age was 57.97 (15.03) years and 57.32 (14.37) years for the control and intervention groups, respectively. Among the 63 study patients in each group, 71% (n=45) were female in the control group, whereas 60% (n=38) were female in the intervention group; 67% (n=42) were housewives in the control group, whereas 56% (n=35) were housewives in the intervention group; and 79% (n=50) were married in the control group, whereas 71% (n=45) were married in the intervention group. Furthermore, comparisons between the control and intervention groups in 4 other categories show the following differences: being illiterate at 40% (n=25) versus 48% (n=30), income <US $100/month at 14% (n=9) versus 22% (n=14), present tobacco user at 13% (n=8) versus 16% (n=10), and current smokeless tobacco user at 43% (n=27) versus 30% (n=19). These characteristics did not differ significantly between the control and intervention groups. Except for hip circumference, no outcome indicators differed significantly at baseline. Baseline details are given in Multimedia Appendix 3, and patient flow through in the trial is shown in Figure 1.

The analyses included repeated measures such as ANOVA, and the results were observed to be significantly different for within groups, between groups, and the interaction of group × time factor in terms of knowledge score. QOL on average changed more favorably in the intervention group than in the control group, but the difference largely failed to achieve statistical significance. Diastolic BP and BMI showed significant differences arising from within groups and the interaction of group × time factor; food salinity and hip circumferences showed significant differences arising from within groups and between groups. Moreover, Systolic BP and waist circumference showed significant differences from within groups. However, no significant differences were found arising from within groups, between groups, and the interactions of group × time in terms of urine salinity and MUAC (Table 1).

Using the generalized estimating equation, knowledge score was considerably increased; however, food salinity and hip circumference were significantly decreased in the intervention group compared to the control group in both the unadjusted and adjusted models. However, waist circumference was considerably lower in the intervention group than the control group in the adjusted model (Table 2).

Regarding the laboratory findings, from baseline to 6 months, the mean (SD) FBS decreased by 0.51 (3.77) mmol/L in the intervention group and by 0.10 (1.44) mmol/L in the control group (*P*=.03); however, BUN increased by 3.64 (7.17) mg/dL in the intervention group and by 1.68 (10.10) mg/dL in the control group (*P*=.01). No other laboratory parameters showed any significant changes over the 6-month duration of the study (Table 3).

Table 1. Changes in the study outcomes between the intervention group and control group over time (2-way repeated measures ANOVA test was performed).

Characteristic, group	Baseline, mean (SD)	At 3 months, mean (SD)	At 6 months, mean (SD)	*P* value		
				Within groups	Between groups	Interaction
Knowledge score (%)				<.001	<.001	<.001
Control	29.89 (18.81)	45.96 (19.30)	42.06 (17.06)			
Intervention	27.78 (18.34)	68.98 (14.25)	70.76 (13.12)			
Quality of life (EQ-5D-5L score)				.83	.21	.91
Control	0.75 (0.14)	0.76 (0.14)	0.75 (0.14)			
Intervention	0.78 (0.14)	0.78 (0.13)	0.78 (0.12)			
Systolic blood pressure (mmHg)				<.001	.18	.05
Control	143.73 (24.22)	139.10 (21.89)	131.25 (19.05)			
Intervention	143.43 (24.12)	130.96 (18.21)	126.55 (18.18)			
Diastolic blood pressure (mmHg)				<.001	.79	.001
Control	88.37 (12.80)	88.20 (12.88)	82.19 (11.27)			
Intervention	91.38 (12.93)	84.55 (10.70)	81.31 (11.99)			
BMI (kg/m^2)				.01	.13	.02
Control	25.13 (3.49)	24.71 (3.60)	24.86 (3.73)			
Intervention	23.83 (4.43)	23.84 (4.40)	23.77 (4.45)			
Urine salinity (g)				.96	.57	.40
Control	10.27 (2.79)	10.20 (2.73)	10.71 (3.24)			
Intervention	10.64 (3.26)	10.80 (2.97)	10.40 (3.65)			
Food salinity (g)				.001	.001	.19
Control	0.65 (0.23)	0.69 (0.30)	0.62 (0.18)			
Intervention	0.62 (0.20)	0.61 (0.22)	0.49 (0.15)			
Mid-upper arm circumference (cm)				.11	.17	.21
Control	28.11 (3.17)	27.89 (3.20)	28.00 (3.21)			
Intervention	27.36 (3.39)	27.30 (3.40)	27.03 (3.16)			
Waist circumference (cm)				.003	.15	.51
Control	87.12 (11.03)	86.88 (10.95)	86.89 (11.01)			
Intervention	84.18 (11.37)	84.08 (11.35)	84.01 (11.32)			
Hip circumference (cm)				.03	.02	.34
Control	93.72 (6.96)	93.55 (6.88)	93.57 (6.91)			
Intervention	90.61 (8.75)	90.53 (8.71)	90.29 (8.89)			

Table 2. Changes in the study outcomes between the intervention group and control group over time (using generalized estimating equation).

Characteristic	Unadjusted coefficient (95% CI)	*P* value	Adjusted[a] Coefficient (95% CI)	*P* value
Knowledge score (%)	16.53 (12.06-21.00)	<.001	15.95 (11.76-20.14)	<.001
Quality of life (EQ-5D-5L score)	0.02 (–0.01 to 0.07)	.20	0.019 (–0.02 to 0.05)	.36
Systolic blood pressure (mmHg)	–4.37 (–10.71 to 1.95)	.17	–4.14 (–10.43 to 2.14)	.19
Diastolic blood pressure (mmHg)	–0.5 (–4.19 to 3.170)	.78	–1.05 (–4.67 to 2.56)	.56
BMI (kg/m^2)	–1.08 (–2.47 to 0.30)	.12	–1.01 (–2.26 to 0.23)	.11
Urine salinity (g)	0.22 (–0.53 to 0.97)	.56	0.15 (–0.61 to 0.92)	.69
Food salinity (g)	–0.08 (–0.12 to –0.03)	.001	–0.08 (–0.12 to –0.03)	.001
MUAC[b] (cm)	–0.77 (–1.87 to 0.32)	.16	–0.86 (–1.86 to 0.14)	.09
Waist circumference (cm)	–2.87 (–6.74 to 0.99)	.14	–3.83 (–7.21 to –0.46)	.02
Hip circumference (cm)	–3.13 (–5.86 to –0.40)	.02	–3.09 (–5.60 to –0.58)	.01

[a] Adjusted in a generalized estimating equation model for age, gender, education, marital status, and occupation.

[b] MUAC: mid-upper arm circumference.

Table 3. Changes of the laboratory values, from baseline to 6 months, between the intervention group and control group (intention-to-treat analysis was performed).

Variable	Intervention (N=63)		Control (N=63)		*P* value
	Mean (SD)	95% CI	Mean (SD)	95% CI	
FBS[a] (mmol/L)	–0.51 (3.77)	–1.46 to 0.44	–0.10 (1.44)	–0.46 to 0.26	.03[b]
Serum cholesterol (mg/dL)	–14.22 (27.58)	–21.17 to –7.27	–9.76 (24.57)	–15.95 to –3.57	.34[c]
Serum creatinine (mg/dL)	0.11 (0.18)	0.07-0.16	0.11 (0.21)	0.06-0.17	.63[b]
eGFR[d] (mL/min/1.73 m^2)	–6.62 (9.81)	–9.09 to –4.15	–6.21 (8.95)	–8.46 to –3.95	.80[c]
Serum HDL-c[e] (mg/dL)	–1.36 (8.94)	–3.62 to 0.89	–2.16 (9.12)	–4.46 to 0.14	.58[b]
Serum triglyceride (mg/dL)	–18.82 (149.36)	–56.44 to 18.79	2.71 (83.51)	–18.32 to 23.75	.75[b]
Serum albumin (g/dL)	0.07 (0.18)	0.02-0.12	0.07 (0.17)	0.03-0.11	.89[b]
Hemoglobin (g/dL)	–0.37 (0.82)	–0.58 to –0.16	–0.52 (0.67)	–0.69 to –0.35	.65[b]
BUN[f] (mg/dL)	3.64 (7.17)	1.84-5.45	1.68 (10.10)	–0.86 to 4.22	.01[b]
Serum uric acid (mg/dL)	0.30 (0.94)	0.06-0.53	0.19 (1.01)	–0.06 to 0.45	.35[b]
ACR[g] (mg/g)	–86.80 (462.59)	–203.30 to 29.70	–78.08 (212.29)	–131.55 to –24.62	.27[b]

[a] FBS: fasting blood sugar.

[b] Mann-Whitney *U* test.

[c] Independent 2-tailed *t* test.

[d] eGFR: estimated glomerular filtration rate.

[e] HDL-c: high-density lipoprotein cholesterol.

[f] BUN: blood urea nitrogen.

[g] ACR: albumin-to-creatinine ratio.

Discussion

Principal Findings

To our knowledge, this study is the first trial assessing the outcomes of a health education intervention using a nephrologist-facilitated health campaign and CHW-conducted health education using mHealth technology on patients with CKD in Bangladesh. In this single-center, randomized trial of patients with CKD in stages 1-3, health education through a nephrologist-facilitated health campaign and CHW-conducted mHealth education improved the patient knowledge status when compared with usual patient care. In addition, mHealth can significantly increase disease knowledge in patients with CKD. The effectiveness of CKD campaigns in raising CKD awareness and boosting motivation for healthy lifestyle changes to reduce

CKD-related complications seems promising. Patients can rely on nephrologists as they are a trusted channel for delivering health education on CKD and related risk factors. Most LMIC have a scarcity of nephrologists, particularly in rural areas; therefore, policy makers should prioritize this leading health problem while formulating appropriate intervention strategies. On the other hand, CHWs play an important role in delivering health education through mHealth [30], because they have the ability to develop interventions and sustain community well-being [31], especially in areas where there are a prevailing shortage of registered physicians and nurses.

It is crucial for patients with CKD to keep their BP under control as hypertension is a major risk factor for the development and progression of CKD. In our analysis, the intervention group had decreased systolic and diastolic BP at the end of the study. Weekly BP measurements, in addition to health education, might also have been influential for this study's patients. Studies have documented that regular BP monitoring reduces systolic and diastolic BP significantly when compared to usual care [32]. Our findings are comparable with other similar studies that have been undertaken in a range of settings [13,32]. An RCT in rural India showed that a 3-month health education intervention reduced BP in the intervention group compared to the control group [13]. According to a research study, one-third of Bangladeshis have never monitored their BP and have no idea how to control it [14]. Weekly BP checks may alert patients to their BP status, motivating them to better control their diastolic BP. It has also been hypothesized that if patients are aware of their weekly BP levels and know the risk of hypertension with CKD, then they may be more likely to comply with medical therapy in the longer term. In the intervention group, urine salinity remained unchanged despite considerable reductions in BP and dietary salinity. To decrease salt intake, people must regulate their daily lives once they have gained knowledge, which necessitates ongoing community education as well as changes in food business policies.

Patients with diabetes and CKD are at a higher risk of cardiovascular disease and renal failure. Comprehensive education is essential for empowering patients with diabetes and CKD to self-manage their health status [33]. Our effective health education improved patients' knowledge and awareness about bringing changes in lifestyle and maintaining healthy dietary practices in particular. We found that FBS level was significantly reduced among the intervention group at the end of the 6-month intervention period. According to a study conducted among patients with diabetes in rural China [34], health education enhanced diabetes knowledge and significantly reduced FBS in the intervention group compared to the control group. Related to the BMI, waist and hip circumferences also showed a decreasing trend at the end of 6 months.

The intervention group's BUN level increased significantly when compared to the control group, which is a notable finding in our study. In patients with CKD, BUN is a marker of the retention of nitrogenous uremic solutes, which are predominantly obtained from protein consumption [35]. Increased protein consumption is the most common extrarenal cause of BUN elevation for patients in the early stage of CKD. Nutritional health education is strongly recommended to achieve a positive outcome.

This study observed a favorable trend of QOL improvement in the intervention group compared to the control group; however, no significant improvement was observed. Health education has a strong favorable influence on all health-related QOL metrics in patients undergoing hemodialysis. However, in our analysis, patients with CKD in stages 1-3 were included, and the 6-month trial duration could be the primary explanation for these poor outcomes. Overall, our data are in line with the concept that health education may improve QOL, even though our study patients in the intervention group were illiterate. These, once again, could be linked to poor urine salinity and MUAC outcomes. Improving QOL, urine salinity, and MUAC in this population may be a challenging goal, and perhaps, multiple interventions with longer duration would be the best approach to the problem.

Limitations and Strengths

This study's limitations include a small sample size and a 6-month intervention and follow-up period, in which a longer follow-up period could generate more accurate results. Our study patients were selected randomly from the 3 unions of the Mirzapur subdistrict, and this does not represent the entire rural and peri-urban population with CKD. Moreover, data contamination by neighbors and family members was very plausible; however, the CHWs obtained a verbal agreement from the study patients not to share any study details with their neighbors or other family members. Furthermore, since the participants provided their own data, the outcome assessment was not blinded. The strengths of the study include the facilitation of the health campaign by a nephrologist, and the CHWs' provision of health education through mobile phone calls in the patients' native language (Bengali). Health education materials were developed using the same native language for better understanding even with the patients' minimum technical knowledge and skills. Furthermore, the study's strengths are the unbiased systematic sampling approach used to recruit patients and the standard laboratory facility used to identify patients with CKD.

Practical Implications

Integrating nephrologists and CHWs for health education may enhance the knowledge, glycemic control, and hypertension care of patients with CKD in rural and peri-urban Bangladesh. A CKD campaign and the use of mHealth technology would be substantial advantages for the target populations to minimize CKD risk factors. BP monitoring should be included in the routine care of patients with hypertensive CKD to control their BP.

Conclusions

The health education intervention through a campaign and mHealth technologies demonstrated the potential for improving CKD knowledge among patients with CKD. Education campaigns may have potential for improving FBS and BP among patients with CKD. Both the positive outcomes of health education and weekly BP monitoring interventions on patients with CKD provide evidence for the potential scaling up of these interventions in Bangladesh and possibly other LMIC, especially in rural and peri-urban settings.

Acknowledgments

Japan's Grants-in-Aid for Scientific Research Program has provided funding for this study (grant 18H03113). We would like to express our gratitude to the study patients for their generous support. We are grateful to Dr Yuko Ito for her enormous support and Dr Ashir Ahmed (Kyushu University, Japan) for providing the Portable Health Clinic box. We would like to express our heartfelt gratitude to the nephrologist, community health workers, and laboratory staff for their unconditional support. The International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) is also grateful to the governments of Bangladesh, Canada, Sweden, and the United Kingdom for providing core and unrestricted support.

Authors' Contributions

MHRS, MM, and ASGF designed the study. MHRS, MM, and ASGF performed the main statistical analysis. HUR, MMR, SKS, SKD, SEA, TA, and MJC helped with the main statistical analysis and data interpretation. MM, MMR, MJC, SKS, SEA, TA, ASGF, and MHRS helped with the material support during the study. All authors contributed to the drafting and review of the manuscript and approved the final version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study activities.
[DOCX File , 20 KB - jmir_v24i8e37314_app1.docx]

Multimedia Appendix 2
Contents of the mobile health education (over a mobile phone call lasting approximately 10 minutes).
[DOCX File , 20 KB - jmir_v24i8e37314_app2.docx]

Multimedia Appendix 3
Demographic characteristics and outcome variables among the study participants at baseline.
[DOCX File , 20 KB - jmir_v24i8e37314_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 84 KB - jmir_v24i8e37314_app4.pdf]

References

1. Couser WG, Remuzzi G, Mendis S, Tonelli M. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. Kidney Int 2011 Dec;80(12):1258-1270 [FREE Full text] [doi: 10.1038/ki.2011.368] [Medline: 21993585]
2. Luyckx VA, Tonelli M, Stanifer JW. The global burden of kidney disease and the sustainable development goals. Bull World Health Organ 2018 Jun 01;96(6):414-422D [FREE Full text] [doi: 10.2471/BLT.17.206441] [Medline: 29904224]
3. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2020 Feb 29;395(10225):709-733 [FREE Full text] [doi: 10.1016/S0140-6736(20)30045-3] [Medline: 32061315]
4. Varma PP. Prevalence of chronic kidney disease in India - where are we heading? Indian J Nephrol 2015;25(3):133-135 [FREE Full text] [Medline: 26060360]
5. Sarker MHR, Moriyama M, Rashid HU, Chisti MJ, Rahman MM, Das SK, et al. Community-based screening to determine the prevalence, health and nutritional status of patients with CKD in rural and peri-urban Bangladesh. Ther Adv Chronic Dis 2021;12:20406223211035281 [FREE Full text] [doi: 10.1177/20406223211035281] [Medline: 34377387]
6. Kidney Disease: Improving Global Outcomes (KDIGO). Chapter 1: definition and classification of CKD. Kidney Int Suppl (2011) 2013 Jan;3(1):19-62 [FREE Full text] [doi: 10.1038/kisup.2012.64] [Medline: 25018975]
7. Stevens PE, Levin A, Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. Ann Intern Med 2013 Jun 04;158(11):825-830 [FREE Full text] [doi: 10.7326/0003-4819-158-11-201306040-00007] [Medline: 23732715]

8. George C, Mogueo A, Okpechi I, Echouffo-Tcheugui JB, Kengne AP. Chronic kidney disease in low-income to middle-income countries: the case for increased screening. BMJ Glob Health 2017;2(2):e000256 [FREE Full text] [doi: 10.1136/bmjgh-2016-000256] [Medline: 29081996]
9. Vallianou NG, Mitesh S, Gkogkou A, Geladari E. Chronic kidney disease and cardiovascular disease: is there any relationship? Curr Cardiol Rev 2019;15(1):55-63 [FREE Full text] [doi: 10.2174/1573403X14666180711124825] [Medline: 29992892]
10. Chen TK, Knicely DH, Grams ME. Chronic kidney disease diagnosis and management: a review. JAMA 2019 Oct 01;322(13):1294-1304 [FREE Full text] [doi: 10.1001/jama.2019.14745] [Medline: 31573641]
11. Narva AS, Norton JM, Boulware LE. Educating patients about CKD: the path to self-management and patient-centered care. Clin J Am Soc Nephrol 2016 Apr 07;11(4):694-703 [FREE Full text] [doi: 10.2215/CJN.07680715] [Medline: 26536899]
12. Rawal L, Jubayer S, Choudhury SR, Islam SMS, Abdullah AS. Community health workers for non-communicable diseases prevention and control in Bangladesh: a qualitative study. Glob Health Res Policy 2020 Dec 24;6(1):1 [FREE Full text] [doi: 10.1186/s41256-020-00182-z] [Medline: 33407942]
13. Gamage DG, Riddell MA, Joshi R, Thankappan KR, Chow CK, Oldenburg B, et al. Effectiveness of a scalable group-based education and monitoring program, delivered by health workers, to improve control of hypertension in rural India: a cluster randomised controlled trial. PLoS Med 2020 Jan;17(1):e1002997 [FREE Full text] [doi: 10.1371/journal.pmed.1002997] [Medline: 31895945]
14. Jahan Y, Rahman MM, Faruque ASG, Chisti MJ, Kazawa K, Matsuyama R, et al. Awareness development and usage of mobile health technology among individuals with hypertension in a rural community of Bangladesh: randomized controlled trial. J Med Internet Res 2020 Dec 07;22(12):e19137 [FREE Full text] [doi: 10.2196/19137] [Medline: 33284129]
15. Tuot DS, Boulware LE. Telehealth applications to enhance CKD knowledge and awareness among patients and providers. Adv Chronic Kidney Dis 2017 Jan;24(1):39-45 [FREE Full text] [doi: 10.1053/j.ackd.2016.11.017] [Medline: 28224941]
16. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al. A predictive model for progression of chronic kidney disease to kidney failure. JAMA 2011 Apr 20;305(15):1553-1559. [doi: 10.1001/jama.2011.451] [Medline: 21482743]
17. Wagner M, Wanner C, Schich M, Kotseva K, Wood D, Hartmann K, et al. Patient's and physician's awareness of kidney disease in coronary heart disease patients - a cross-sectional analysis of the German subset of the EUROASPIRE IV survey. BMC Nephrol 2017 Oct 25;18(1):321 [FREE Full text] [doi: 10.1186/s12882-017-0730-3] [Medline: 29070030]
18. Finkelstein FO, Story K, Firanek C, Barre P, Takano T, Soroka S, et al. Perceived knowledge among patients cared for by nephrologists about chronic kidney disease and end-stage renal disease therapies. Kidney Int 2008 Nov;74(9):1178-1184 [FREE Full text] [doi: 10.1038/ki.2008.376] [Medline: 18668024]
19. Zimbudzi E, Lo C, Ranasinha S, Kerr PG, Usherwood T, Cass A, et al. Self-management in patients with diabetes and chronic kidney disease is associated with incremental benefit in HRQOL. J Diabetes Complications 2017 Feb;31(2):427-432. [doi: 10.1016/j.jdiacomp.2016.10.027] [Medline: 27914731]
20. Grant S, Mayo-Wilson E, Montgomery P, Macdonald G, Michie S, Hopewell S, CONSORT-SPI Group. CONSORT-SPI 2018 explanation and elaboration: guidance for reporting social and psychological intervention trials. Trials 2018 Jul 31;19(1):406 [FREE Full text] [doi: 10.1186/s13063-018-2735-z] [Medline: 30060763]
21. Chan A, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013 Feb 05;158(3):200-207 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00583] [Medline: 23295957]
22. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med 2009 May 05;150(9):604-612 [FREE Full text] [doi: 10.7326/0003-4819-150-9-200905050-00006] [Medline: 19414839]
23. Kidney Disease: Improving Global Outcomes (KDIGO) Blood Pressure Work Group. KDIGO 2021 clinical practice guideline for the management of blood pressure in chronic kidney disease. Kidney Int 2021 Mar;99(3S):S1-S87 [FREE Full text] [doi: 10.1016/j.kint.2020.11.003] [Medline: 33637192]
24. Inker LA, Astor BC, Fox CH, Isakova T, Lash JP, Peralta CA, et al. KDOQI US commentary on the 2012 KDIGO clinical practice guideline for the evaluation and management of CKD. Am J Kidney Dis 2014 May;63(5):713-735. [doi: 10.1053/j.ajkd.2014.01.416] [Medline: 24647050]
25. Sarker MHR, Moriyama M, Rashid HU, Rahman MM, Chisti MJ, Das SK, et al. Health education through a campaign and mHealth to enhance knowledge and quality of life among patients with chronic kidney disease in Bangladesh: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Nov 19;10(11):e30191 [FREE Full text] [doi: 10.2196/30191] [Medline: 34806998]
26. Gheewala PA, Peterson GM, Zaidi STR, Jose MD, Castelino RL. Public knowledge of chronic kidney disease evaluated using a validated questionnaire: a cross-sectional study. BMC Public Health 2018 Mar 20;18(1):371 [FREE Full text] [doi: 10.1186/s12889-018-5301-4] [Medline: 29554891]
27. van Hout B, Janssen MF, Feng Y, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-715 [FREE Full text] [doi: 10.1016/j.jval.2012.02.008] [Medline: 22867780]

28. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
29. Lachin JM. Fallacies of last observation carried forward analyses. Clin Trials 2016 Apr;13(2):161-168 [FREE Full text] [doi: 10.1177/1740774515602688] [Medline: 26400875]
30. Thondoo M, Strachan DL, Nakirunda M, Ndima S, Muiambo A, Källander K, InSCALE Study Group. Potential roles of mHealth for community health workers: formative research with end users in Uganda and Mozambique. JMIR mHealth uHealth 2015 Jul 23;3(3):e76 [FREE Full text] [doi: 10.2196/mhealth.4208] [Medline: 26206419]
31. Javanparast S, Windle A, Freeman T, Baum F. Community health worker programs to improve healthcare access and equity: are they only relevant to low- and middle-income countries? Int J Health Policy Manag 2018 Oct 01;7(10):943-954 [FREE Full text] [doi: 10.15171/ijhpm.2018.53] [Medline: 30316247]
32. Verberk WJ, Kroon AA, Lenders JWM, Kessels AGH, van Montfrans GA, Smit AJ, Home Versus Office Measurement, Reduction of Unnecessary Treatment Study Investigators. Self-measurement of blood pressure at home reduces the need for antihypertensive drugs: a randomized, controlled trial. Hypertension 2007 Dec;50(6):1019-1025. [doi: 10.1161/HYPERTENSIONAHA.107.094193] [Medline: 17938383]
33. McDuffie RH, Struck L, Burshell A. Empowerment for diabetes management: integrating true self-management into the medical treatment and management of diabetes mellitus. Ochsner J 2001 Jul;3(3):149-157 [FREE Full text] [Medline: 22754392]
34. Chen S, Qian D, Burström K, Burström B. Impact of an educational intervention in primary care on fasting blood glucose levels and diabetes knowledge among patients with type 2 diabetes mellitus in rural China. Patient Educ Couns 2020 Sep;103(9):1767-1773 [FREE Full text] [doi: 10.1016/j.pec.2020.03.010] [Medline: 32414563]
35. Kim HJ, Kim TE, Han M, Yi Y, Jeong JC, Chin HJ, et al. Effects of blood urea nitrogen independent of the estimated glomerular filtration rate on the development of anemia in non-dialysis chronic kidney disease: the results of the KNOW-CKD study. PLoS One 2021;16(9):e0257305 [FREE Full text] [doi: 10.1371/journal.pone.0257305] [Medline: 34506574]

Abbreviations

BP: blood pressure
BUN: blood urea nitrogen
CHW: community health worker
CKD: chronic kidney disease
CONSORT: Consolidated Standards of Reporting Trials
DSS: demographic surveillance system
FBS: fasting blood sugar
LMIC: low- and middle-income countries
mHealth: mobile health
MUAC: mid-upper arm circumference
QOL: quality of life
RCT: randomized controlled trial
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials

Edited by R Kukafka; submitted 16.02.22; peer-reviewed by N Koizumi, YC Chen, Y Jahan; comments to author 10.04.22; revised version received 05.06.22; accepted 13.06.22; published 11.08.22.

Please cite as:
Sarker MHR, Moriyama M, Rashid HU, Rahman MM, Chisti MJ, Das SK, Saha SK, Arifeen SE, Ahmed T, Faruque ASG
Chronic Kidney Disease Awareness Campaign and Mobile Health Education to Improve Knowledge, Quality of Life, and Motivation for a Healthy Lifestyle Among Patients With Chronic Kidney Disease in Bangladesh: Randomized Controlled Trial
J Med Internet Res 2022;24(8):e37314
URL: https://www.jmir.org/2022/8/e37314
doi:10.2196/37314
PMID:35969429

Original Paper

Smoking Cessation Smartphone App Use Over Time: Predicting 12-Month Cessation Outcomes in a 2-Arm Randomized Trial

Jonathan B Bricker[1,2], PhD; Kristin E Mull[1], MSc; Margarita Santiago-Torres[1], PhD; Zhen Miao[3], MSc; Olga Perski[4], PhD; Chongzhi Di[1], PhD

[1]Division of Public Health Sciences, Fred Hutch Cancer Center, Seattle, WA, United States

[2]Department of Psychology, University of Washington, Seattle, WA, United States

[3]Department of Statistics, University of Washington, Seattle, WA, United States

[4]Department of Behavioural Science and Health, University College London, London, United Kingdom

Corresponding Author:
Jonathan B Bricker, PhD
Division of Public Health Sciences
Fred Hutch Cancer Center
1100 Fairview Avenue North
M3-B232
Seattle, WA, 98109
United States
Phone: 1 2066675074
Email: jbricker@fredhutch.org

Abstract

Background: Little is known about how individuals engage over time with smartphone app interventions and whether this engagement predicts health outcomes.

Objective: In the context of a randomized trial comparing 2 smartphone apps for smoking cessation, this study aimed to determine distinct groups of smartphone app log-in trajectories over a 6-month period, their association with smoking cessation outcomes at 12 months, and baseline user characteristics that predict data-driven trajectory group membership.

Methods: Functional clustering of 182 consecutive days of smoothed log-in data from both arms of a large (N=2415) randomized trial of 2 smartphone apps for smoking cessation (iCanQuit and QuitGuide) was used to identify distinct trajectory groups. Logistic regression was used to determine the association of group membership with the primary outcome of 30-day point prevalence of smoking abstinence at 12 months. Finally, the baseline characteristics associated with group membership were examined using logistic and multinomial logistic regression. The analyses were conducted separately for each app.

Results: For iCanQuit, participants were clustered into 3 groups: “1-week users” (610/1069, 57.06%), “4-week users” (303/1069, 28.34%), and “26-week users” (156/1069, 14.59%). For smoking cessation rates at the 12-month follow-up, compared with 1-week users, 4-week users had 50% higher odds of cessation (30% vs 23%; odds ratio [OR] 1.50, 95% CI 1.05-2.14; *P*=.03), whereas 26-week users had 397% higher odds (56% vs 23%; OR 4.97, 95% CI 3.31-7.52; *P*<.001). For QuitGuide, participants were clustered into 2 groups: “1-week users” (695/1064, 65.32%) and “3-week users” (369/1064, 34.68%). The difference in the odds of being abstinent at 12 months for 3-week users versus 1-week users was minimal (23% vs 21%; OR 1.16, 95% CI 0.84-1.62; *P*=.37). Different baseline characteristics predicted the trajectory group membership for each app.

Conclusions: Patterns of 1-, 3-, and 4-week smartphone app use for smoking cessation may be common in how people engage in digital health interventions. There were significantly higher odds of quitting smoking among 4-week users and especially among 26-week users of the iCanQuit app. To improve study outcomes, strategies for detecting users who disengage early from these interventions (1-week users) and proactively offering them a more intensive intervention could be fruitful.

(J Med Internet Res 2022;24(8):e39208) doi:10.2196/39208

KEYWORDS

acceptance and commitment therapy; ACT; digital interventions; eHealth; engagement; iCanQuit; QuitGuide; mobile health; mHealth; smartphone apps; trajectories; tobacco; smoking; mobile phone

Introduction

Background

User engagement in digital behavior change interventions has predicted improved treatment outcomes across a wide variety of domains, including mental health, physical activity, dietary change, weight loss, alcohol use, and smoking cessation [1-9]. A central challenge in creating effective digital health behavior change interventions is that a large proportion of users disengage early from these interventions, thereby contributing to low treatment success rates [2,10,11]. Given the importance of user engagement, designing strategies to increase engagement has been a priority of digital behavioral health interventions [2,7,8,12,13].

Within the domain of cigarette smoking, smartphone apps for smoking cessation have become a ubiquitous intervention approach for which user engagement can be readily measured. Nearly 500 English language smartphone apps for smoking cessation have been downloaded more than 33 million times since 2012 (R Nelson, SensorTower.com, personal communication, April 15, 2020). Higher user engagement in smartphone interventions for smoking cessation is predictive of cessation outcomes [11,14-16]. Although this positive association may partly be driven by self-selection bias (or reverse causation), as users have not been randomized to different levels of engagement, a certain level of exposure to the intervention's active ingredients is logically necessary for successful behavior change [2]. This logic has led to a body of work focusing on identifying design strategies that promote user engagement. For example, tailoring content to individuals' characteristics or unique situations, interactivity (eg, through conversational agents) [17], and credibility have been found across a range of studies using mixed methods to be important for engagement with smoking cessation interventions.

Although smartphone intervention engagement is usually measured by the number of log-ins, little is known about how users engage with smartphone interventions *over time* and whether those temporal patterns predict higher odds of smoking cessation. In the educational literature, a well-documented finding is that learning new material becomes more effective when it occurs over a longer period [18]. This process, called spaced practice, increases the variability in learning and remembering new information [19]. The purpose of this study was to determine, in the domain of smartphone apps for smoking cessation, whether user engagement over time leads to improved cessation outcomes.

Smartphone apps for smoking cessation are commonly available for participants to use at will, resulting in high variations in use trajectories over time. For example, some users may follow a trajectory of logging in several times within the first few days of starting an intervention and then never return. Others may follow a trajectory in which they log in consistently and then gradually taper off. Other users may follow a trajectory in which they consistently log in over the course of several months. Conceivably, some groups of individuals might follow unique use trajectories over time that are associated with differential health outcomes. For example, people who log in consistently over the course of many months might have higher cessation rates, because they have consistently benefited from the information and skills presented in the app.

There is a dearth of studies analyzing use trajectories for digital smoking cessation interventions. We are aware of 3 publications for SMS text messaging interventions [20-22], 4 publications for website interventions [5,23-25], and none for smartphone interventions. Regarding SMS text messaging interventions, a study identified 5 distinct use trajectories of an SMS text messaging–based smoking cessation program over the 6 weeks after quit date, namely high engagement, increasing engagement, rapid decrease, delayed decrease, and low engagement [20]. The study found that the high engagement and increasing engagement groups were more likely than the other groups to quit smoking over the course of 6 weeks. Within the context of smoking cessation websites, our group conducted a functional clustering analysis of log-in data from both arms of a large (N=2637) randomized trial of 2 website interventions for smoking cessation (WebQuit and Smokefree), with a primary outcome of 30-day point prevalence smoking abstinence at 12 months [24]. Compared with 1-week WebQuit users, 5- and 52-week users had 57% higher odds (odds ratio [OR] 1.57, 95% CI 1.13-2.17; *P*=.007) and 124% higher odds (OR 2.24, 95% CI 1.45-3.43; *P*<.001), respectively, of being abstinent at 12 months. The 5-week use of either website predicted higher odds of quitting smoking, with the highest odds for 52-week WebQuit users. These results suggest that experimental testing strategies to increase digital intervention engagement for 4 more weeks (ie, from 1 week to 5 weeks) would be valuable. Studying distinct groups of use trajectories can help identify which use patterns are beneficial and thereby make recommendations for future program use. These results will help inform the design of more engaging digital interventions for smoking cessation, with the ultimate goal of a higher likelihood of cessation.

If we can identify smartphone intervention use trajectories that predict cessation, understanding the sociodemographic characteristics of individuals who tend to follow more or less successful trajectories is important. Knowing the characteristics of individuals who are likely to have certain engagement patterns might allow researchers and intervention designers to tailor smartphone interventions according to users' unique baseline characteristics. There is an emerging literature of randomized trial designs that algorithmically use baseline characteristics predictive of treatment outcomes in the design of tailored interventions [26,27]. Although studies have found that being female, being older, and having a higher education are generally consistent predictors of greater digital intervention use [28-31], very little is known about the user characteristics that are associated with different patterns of use over time [32,33]. For example, a study found that being female and having higher baseline motivation were associated with more consistent log-in trajectories [34]. For the WebQuit website intervention, we found that smoking for at least the past 10 years and screening negative for anxiety predicted a 90% higher odds (OR 1.90, 95% CI 1.14-3.14) and a 56% higher odds (OR 1.56, 95% CI 1.06-2.33), respectively, of being a 52-week user (compared with being a 1-week user) [24]. Regarding smoking cessation smartphone apps, we are aware of no literature on baseline predictors of their use patterns over time.

We recently developed and tested iCanQuit, a smartphone app for smoking cessation based on acceptance and commitment therapy (ACT), a behavioral approach that teaches skills for allowing cravings to smoke to pass without smoking, which is conceptually distinct from the US Clinical Practice Guidelines (USCPG)–based approaches that teach avoidance of urges [35]. In a large 2-arm randomized trial, iCanQuit was compared with QuitGuide, a USCPG-based smartphone app. At the 12-month follow-up, iCanQuit was 1.5 times more efficacious than QuitGuide for smoking cessation among 2415 adults who smoked (36% racial or ethnic minority groups) from all 50 US states [35]. The iCanQuit study was the first full-scale randomized controlled trial with long-term follow-up to show that a smartphone app was efficacious for smoking cessation.

Objectives

Using data from the iCanQuit parent randomized trial, this study identified the following: (1) distinct groups of smartphone app log-in trajectories, (2) their association with the 12-month smoking cessation outcome, and (3) baseline sociodemographic user characteristics that are associated with different use trajectory groups. Log-in trajectories (ie, log-ins over time) are a generalizable metric that can be useful for other digital intervention researchers—agnostic of the intervention-specific content contained in any one app. The overall goal of this study is to inform the design of future smartphone health interventions that could be more efficacious by identifying trajectory groups in need of further intervention. To accomplish these aims, in this study we analyzed 182 consecutive days of log-in data from both arms of the large (N=2415), 2-arm randomized trial of iCanQuit versus QuitGuide smartphone app interventions for smoking cessation (NCT02724462).

Methods

Design

A total of 2415 individuals were enrolled in the 2-arm iCanQuit randomized controlled trial for smoking cessation, with full protocol details previously described [12]. In brief, a racially and ethnically diverse sample of 2415 adult daily smokers from all 50 US states was randomized 1:1 to receive access to an ACT-based smartphone app (iCanQuit) or a USCPG-based smartphone app (QuitGuide) for smoking cessation. Data for this analysis were from the 2133 individuals who logged into their assigned app at least once and had a complete 182 days of engagement data available.

Eligibility Criteria

Eligibility criteria included individuals who (1) were aged ≥18 years; (2) smoked 5 or more cigarettes per day for the past year; (3) wanted to quit smoking within the next 30 days; (4) if concurrently using any other tobacco products, wanted to quit all tobacco products within 30 days; (5) were interested in learning skills to quit smoking and willing to be randomized to either treatment condition; (6) had daily access to their own smartphone; (7) knew how to download smartphone apps; (8) were willing and able to read in English; (9) had never used QuitGuide and not currently using another smoking cessation treatment; (10) had never participated in our prior studies; (11) had no household members already enrolled; (12) were willing to complete outcome surveys, and (13) could provide contact information for themselves and 2 relatives.

Recruitment, Enrollment, and Follow-up

Adults were recruited nationally via Facebook advertisements, a survey sampling company, search engine results, and friend or family referrals. Participants completed an encrypted web-based screening survey and were notified of their eligibility via email. They then clicked on their secured emailed link to the study website where they provided consent and completed the baseline survey. At each enrollment step, the study was presented as a comparison of 2 smartphone apps for smoking cessation.

Participants were randomized (1:1) to iCanQuit or QuitGuide using randomly permuted blocks of sizes of 2, 4, and 6, stratified by smoking frequency (≤20 vs ≥21 cigarettes per day), education (≤high school vs ≥some college), race or ethnicity (minority race or ethnicity vs non-Hispanic White), and depression screening (Center for Epidemiologic Studies Depression score ≤15 vs ≥16) [31]. Random assignments were concealed from participants throughout the trial. The random allocation sequence was generated by a database manager and implemented automatically by the study website. Neither research staff nor study participants had access to the upcoming randomized study arm assignments. In both arms, participants could access their interventions from the moment of randomization and beyond (ie, after the end of the 12-month follow-up period). All participants provided their consent on the web and were compensated with up to US $105 for completing study data collection. The data retention rate was 88% (1886/2133) and differed slightly between arms (90% in QuitGuide vs 87% in iCanQuit; P=.01).

Interventions

iCanQuit

Participants randomized to the iCanQuit arm received access to download the iCanQuit smartphone app (version 1.2.1). iCanQuit intervenes on the ACT-focused processes of acceptance of internal cues to smoke and enacting one’s values that guide smoking cessation [12]. The acceptance component of the app teaches skills to accept physical sensations, emotions, and thoughts that trigger smoking by distancing oneself from thoughts about smoking (“cognitive defusion”), mindfulness skills, and flexible perspective taking. The values component of the app teaches skills for determining the core life domains that motivate quitting smoking (eg, family, health, and spirituality) and taking repeated small actions within these domains (eg, playing with grandchildren) to develop a smoke-free life. The program is self-paced, and the content is sequentially unlocked across 8 levels. Each of the first 4 levels is made accessible immediately after the prior level is completed, whereas each of the last 4 levels is only unlocked upon recording 7 consecutive days without smoking. If a participant lapses, the program encourages (but does not require) them to set a new quit date and return to the first 4 levels for preparation. The program also includes on-demand tools to help in coping with smoking urges, tracking the daily number of cigarettes smoked, and urges passed without smoking. Content was presented in a sequenced format with short paragraphs of text and some audio or visual content for experiencing ACT concepts.

QuitGuide

Participants randomized to the QuitGuide arm received access to download the QuitGuide smartphone app (version 1.2.2). QuitGuide content is delivered in four main sections: (1) "Thinking about quitting," which focuses on motivations to quit by using reason and logic such as identifying reasons to quit and providing information on the health consequences of smoking and quitting; (2) "Preparing to Quit," which helps users develop a quit plan, identify smoking behaviors, triggers, and reasons for being smoke free, and social support for quitting; (3) "Quitting," which teaches skills for avoiding cravings to smoke; and (4) "Staying Quit," which presents tips, motivations, and actions to stay smoke free and skills for coping with slips. No smoking cessation medications or coaching was provided in either intervention arm [12]. Content was presented in a sequenced format with short paragraphs of text.

Study Measures

Baseline Characteristics and Covariates

Data collected at baseline included age, gender, race, ethnicity, education, employment, income, marital status, and sexual orientation. Study participants completed validated positive screening tools to assess mental health, including depression [31], panic [32], and posttraumatic stress disorder [33]. Alcohol consumption and heavy drinking were assessed using the Quick Drinking Screen [34]. Smoking behavior variables included nicotine dependence (measured using the Fagerström Test for Nicotine Dependence) [35], number of cigarettes smoked per day, years of smoking, use of e-cigarettes, quit attempts, and relationships with other smokers. Acceptance of internal cues to smoke was measured via the Avoidance and Inflexibility Scale (adapted from the study by Gifford et al [36]), using means of the three 9-item subscales that assess one's willingness to experience physical sensations, emotions, and thoughts that cue smoking. The items are rated on a 5-point scale from 1="not at all" to 5="very willing" and averaged, with higher scores indicating greater acceptance. A sample physical sensation item was "How willing are you to notice these bodily sensations without smoking?" and items from the emotions and thoughts subscales were similar, substituting "feelings" or "thoughts" for "bodily sensations." Valued living was measured using the 10-item Valuing Questionnaire [37], designed to assess the extent of personal values enactment. Each item is rated on a 7-point scale ranging from 0="not at all true" to 6="completely true." Scores were averaged, and 2 distinct factors were derived: progress and obstruction, with higher scores indicating either greater progress or greater obstruction toward valued living, respectively. A sample progress item was "I worked toward my goals even if I didn't feel motivated to" and a sample obstruction item was "I was basically on auto-pilot most of the time."

Engagement: Baseline to Day 182 Log-ins

Engagement with the assigned app was measured objectively using Google Analytics. The measure of engagement was the number of days each application was opened, which was consistent with other digital interventions' measures of engagement [7,24,25]. For each participant, time- and date-stamped log file records of each page opening were recorded. For this analysis, we used a binary measure indicating whether each participant logged in at least once each day (ie, had at least one log-in recorded in the log file data). Using this method, each participant had a 0/1 code for each day for 182 days from the date of randomization. Owing to a technical error in the Google Analytics system, only the first 182 days of engagement data were available for the study sample.

Smoking Cessation Outcome: 12 Months

The parent trial's primary smoking cessation outcome was specified a priori as self-reported complete case 30-day point prevalence abstinence (PPA) at the 12-month follow-up. As reported in the parent trial, for the primary outcome of 30-day PPA at the 12-month follow-up, iCanQuit participants had 1.49 times higher odds of quitting smoking as compared with QuitGuide participants (293/1040, 28.17% abstinent vs 225/1067, 21.08% abstinent; OR 1.49, 95% CI 1.22-1.83; P<.001). Note that when missing data were coded as smokers, 12-month 30-day PPA results were very similar (293/1214, 24.13% abstinent for iCanQuit vs 225/1201, 18.73% abstinent for QuitGuide; OR 1.40, 95% CI 1.14-1.71; P<.001).

Statistical Analyses

The analyses were conducted separately for each app. As mentioned in the engagement measurement, log-in data were summarized as a binary time series indicating log-in occurrence on each day of the first 6 months (ie, 182 days) of using the application, from the date of randomization of each participant. Next, log-in time series were presmoothed as the average number of days logged in over a window of 7 previous days [24,36]. This type of dense trajectories data is known as functional data. We applied functional clustering based on functional principal component (FPC) analysis. Specifically, we conducted an FPC analysis [37] by smoothed covariance to summarize each participant's log-in trajectory using a set of low-dimensional FPC scores. We chose to retain the first 3 and 4 FPC scores for clustering in the iCanQuit and QuitGuide arms, respectively, based on a minimum threshold of 90% for the percentage of variance explained. Trajectories were clustered using the Clustering for Large Applications algorithm [38] into k=2 and 3 groups in each arm, which met a minimum prediction strength [39] threshold of 0.6. The Clustering for Large Applications procedure does not rely on parametric assumptions on the shapes of trajectories and is capable of handling densely recorded longitudinal data and complex missing data patterns. We then examined the cluster solutions in each arm for a minimum group size ≥5% of the sample and reasonable separation of the mean log-in trajectories among groups to determine the optimal number of groups for each.

After determining distinct trajectory groups, smoking cessation rates were compared among the groups using logistic regression, with the lowest use group as the reference group. Baseline characteristics with significant univariate associations with cessation were considered as covariates, to control for characteristics that may confound the association between trajectory group and cessation [40]. Thus, the aim of the analysis, as guided by the study's scientific questions, was to understand the unique prediction of the 12-month cessation outcome by trajectory group membership. A shared set of covariates (ie, all baseline characteristics) was considered for each treatment arm. We conducted stepwise logistic regression using Akaike Information Criterion to determine the subset of covariates to be included in the final adjusted model [41]. Finally, baseline characteristics were compared among the groups. Those with significant univariate association with trajectory group membership were considered in stepwise selection, using Akaike Information Criterion, of an adjusted multinomial logistic regression model (iCanQuit arm) or logistic regression model (QuitGuide arm) to determine the best baseline predictors of group membership. All statistical tests were 2-sided, with α=.05, and analyses were conducted in R (version 4.0.3; R Foundation for Statistical Computing [42]), using the R packages "refund" [43] for FPC analysis, "fpc" [44] for prediction strength, and "nnet" [41] for multinomial logistic regression.

Ethics Approval

All study activities were approved by the institutional review board at the Fred Hutchinson Cancer Research Center (approval number IR-8317).

Description of Sample

Multimedia Appendix 1 shows the baseline demographics and participant characteristics in both the iCanQuit and QuitGuide arms. The mean (SD) age at enrollment was 37.8 (10.8) years. Furthermore, 70.28% (1499/2133) of the participants were female and 35.72% (762/2133) of the participants reported racial and ethnic minority backgrounds. There were 40.46% (863/2133) of the participants with a high school or less education. Regarding smoking, 74.54% (1590/2133) of the participants smoked more than half a pack (at least 11 cigarettes) per day. Less than half (785/2030, 38.67%) of the participants had made a quit attempt in the last year, and 82.47% (1759/2133) of the sample had been smoking for >10 years, with an average Fagerström Test for Nicotine Dependence score of 5.86 (moderate nicotine dependence; SD 2.04). There were no statistically significant differences between the 2 arms for any baseline variable (all *P*>.05).

Description of Distinct Groups of Trajectories

The functional clustering analysis of 26 weeks of log-ins revealed 3 distinct groups of trajectories for iCanQuit versus 2 distinct groups for QuitGuide. Log-in patterns are shown in Figure 1 (for iCanQuit) and Figure 2 (for QuitGuide). For iCanQuit (Figure 1), the first trajectory group (610/1069, 57.06%) logged in a mean of 2.0 days in the first week and then had <1 mean log-in day in weeks 2 and beyond. They were termed "1-week users." The second trajectory group (303/1069, 28.34%) logged in a mean of 4.6 days in week 1, a mean of 3.1 days in week 2, a mean of 2.0 days in week 3, a mean of 1.2 days in week 4, and then had <1 mean log-in day in weeks 5 and beyond. They were termed "4-week users." The third trajectory group (156/1069, 14.59%) logged in a mean of 5.0 to 5.4 days per week in weeks 1 through 5, a mean of 3.1 to 4.7 days per week in weeks 6 through 10, tapering to a mean of twice every week starting week 17, and continuing in this pattern until week 26. They were termed "26-week users."

For QuitGuide (Figure 2), the first trajectory group (695/1064, 65.32%) logged in a mean of 1.4 days in the first week and then had <1 mean log-in day in weeks 2 and beyond. As with iCanQuit, they were termed "1-week users." The second trajectory group (369/1064, 34.68%) logged in a mean of 2.8 times in week 1, a mean of 1.7 times in week 2, a mean of 1.1 times in week 3, and then had few log-ins after that. They were termed "3-week users."

Figure 1. Mean weekly log-ins for each trajectory group from the iCanQuit arm. Error bars represent IQRs.

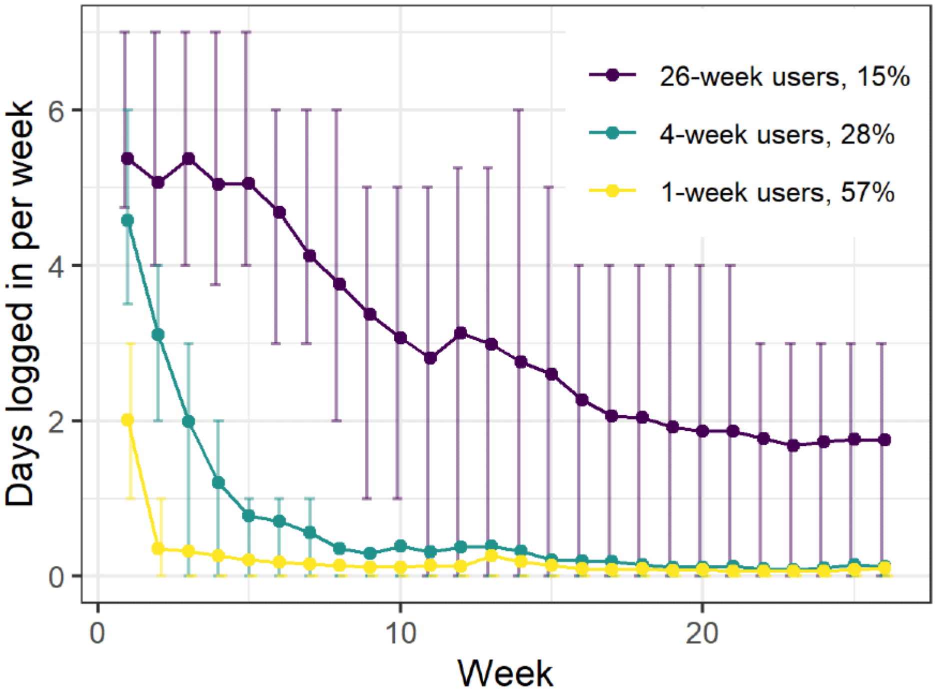
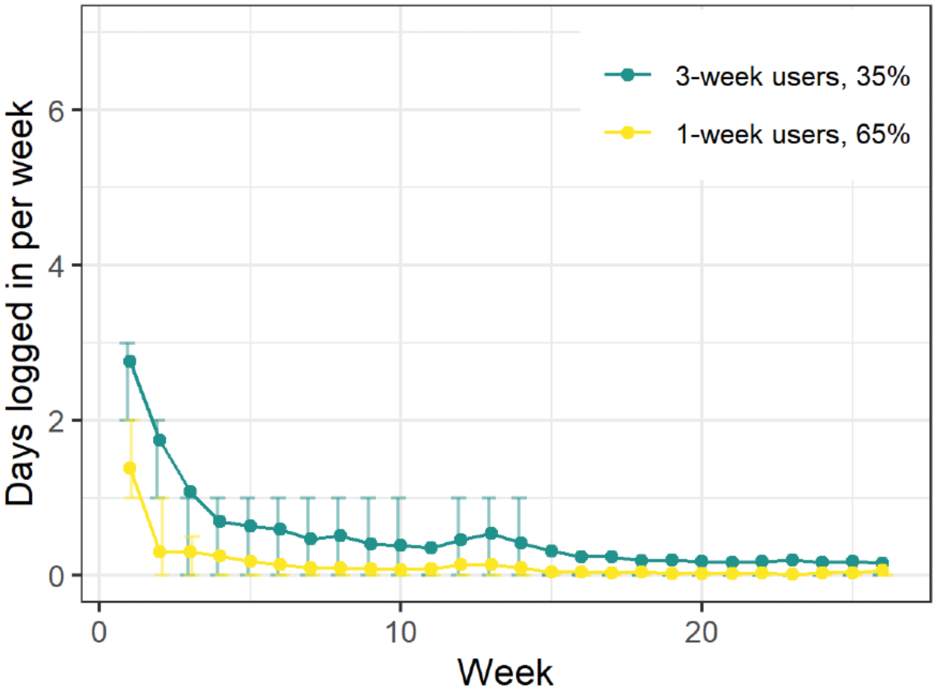


Figure 2. Mean weakly log-ins for each trajectory group from the QuitGuide arm. Error bars indicate IQRs.



Trajectory Membership Prediction of Smoking Cessation Outcome

Table 1 shows each intervention arm's trajectory group membership as a predictor of 30-day PPA at the 12-month follow-up, after controlling for all baseline covariates included in the statistical model. For iCanQuit, abstinence rates for the 3 trajectory groups were 23% for 1-week users, 30% for 4-week users, and 56% for 26-week users. Compared with the 1-week users, the 4-week users had 50% higher odds (OR 1.50, 95% CI 1.05-2.14; *P*=.03), whereas 26-week users had 397% higher odds (OR 4.97, 95% CI 3.31-7.52; *P*<.001), respectively, of being abstinent at 12 months. Descriptively, for QuitGuide, abstinence rates for the 2 trajectory groups were 21% for 1-week users and 23% for 3-week users. There was no significant difference in the odds of being abstinent at 12 months for 3-week versus 1-week users (OR 1.16, 95% CI 0.84-1.62; *P*=.37). The above models adjusted for the baseline covariates selected, as outlined in the statistical methods, and are shown in Table 1.

Table 1. Logistic regression models predicting 12-month smoking cessation outcome by log-in trajectory group, adjusted for Akaike Information Criterion model–selected covariates[a].

Treatment arm and covariate	Odds ratio (95% CI)	*P* value
iCanQuit		
4-week users	1.50 (1.05-2.14)	.03
26-week users	4.97 (3.31-7.52)	<.001
Gender (male)	1.87 (1.33-2.62)	<.001
High school or lower education	1.42 (1.03-1.95)	.03
Depression screen positive	0.69 (0.50-0.97)	.03
Panic disorder screen positive	0.74 (0.50-1.09)	.14
Used e-cigarettes at least once in past month	0.66 (0.45-0.96)	.03
Confidence in being smoke free	1.01 (1.01-1.02)	<.001
Drinks per day on a typical drinking day	0.95 (0.91-1.00)	.048
QuitGuide		
3-week users	1.16 (0.84-1.62)	.37
Used e-cigarettes at least once in past month	1.44 (0.99-2.08)	.05
Confidence in being smoke free	1.01 (1.00-1.02)	.005
Close friends who smoke	0.89 (0.81-0.97)	.01
Heavy drinker[b]	0.64 (0.37-1.07)	.10
Valuing questionnaire—progress	1.03 (1.01-1.06)	.01

[a] The reference group is 1-week users for both arms.

[b] Heavy drinkers are defined as women who had 4 or more drinks and men who had 5 or more drinks on a typical drinking day.

Baseline Characteristics Predicting Trajectory Membership

Because the trajectory groups were different across the 2 arms, Table 2 presents the results for baseline characteristics predicting membership in the groups for the 2 arms separately. For iCanQuit, the baseline characteristics significantly associated with more engaged group membership, as compared with 1-week user group membership, were age in years (OR 1.05, 95% CI 1.03-1.06 for 26-week users), smoking up to one-half pack per day (OR 1.90, 95% CI 1.25-2.87 for 26-week users), smoking first cigarette >5 minutes after waking (OR 1.42, 95% CI 1.05-1.92 for 4-week users), and higher mean acceptance of internal physical sensations (OR 1.82, 95% CI 1.41-2.35 for each 1-point increase for 4-week users).

For QuitGuide, the baseline characteristics significantly associated with 3-week user group membership, as compared with 1-week user group membership, were female gender (OR 1.46, 95% CI 1.10-1.95), minority race (people of color) or ethnicity (Hispanic; OR 1.40, 95% CI 1.08-1.83), and smoked for 10 or more years (OR 1.56, 95% CI 1.04-2.35).

Table 2. Multinomial logistic regression (iCanQuit arm) and logistic regression (QuitGuide arm) results predicting log-in trajectory group membership from Akaike Information Criterion model–selected baseline characteristics[a].

Arm, trajectory group, and characteristic	Odds ratio (95% CI)
iCanQuit	
4-week users	
Age (years)	1.01 (0.99-1.02)
Smokes ≤10 cigarettes per day	1.25 (0.89-1.76)
First cigarette >5 minutes after waking	1.42 (1.05-1.92)
Number of quit attempts in previous year	1.03 (0.99-1.08)
Each point increase in acceptance of physical sensations	1.82 (1.41-2.35)
26-week users	
Age (years)	1.05 (1.03-1.06)
Smokes ≤10 cigarettes per day	1.90 (1.25-2.87)
First cigarette >5 minutes after waking	1.42 (0.96-2.08)
Number of quit attempts in previous year	0.92 (0.83-1.02)
Each point increase in acceptance of physical sensations	1.23 (0.88-1.71)
QuitGuide: 3-week users	
Gender (female)	1.46 (1.10-1.95)
Minority race or ethnicity	1.40 (1.08-1.83)
Anxiety screen positive	0.75 (0.55-1.02)
Smoked for ≥10 years	1.56 (1.04-2.35)

[a]The reference group is 1-week users for both treatment arms.

Discussion

Principal Findings

This study contributes to the nascent literature on longitudinal use trajectories of digital health interventions and their prediction of health outcomes [20-25]. The study found (1) 1-, 4-, and 26-week trajectories for iCanQuit versus 1- and 3-week trajectories for the QuitGuide smoking cessation apps; (2) that these trajectory groups differentially predicted smoking outcomes at 12 months for iCanQuit but not for QuitGuide; and (3) that certain user characteristics were associated with membership in certain trajectory groups. Notably, compared with the 1-week iCanQuit users, the 4-week users had 50% higher odds, whereas 26-week users had 397% higher odds, respectively, of being abstinent at 12 months. The results are discussed in the following sections.

Use Trajectories and Health Outcomes

At least half of the participants in both arms were 1-week users. Similarly, our 2018 study examining log-in trajectories of 2 smoking cessation websites found that half of the participants were 1-week users, and similar to this study, that study showed that these participants were the least likely to have quit smoking at the 12-month follow-up [24]. Thus, there are now 4 separate digital interventions (2 in this study and 2 in the 2018 study [24]) showing large proportions of users (645/1309, 49.27% to 610/1069, 57.06%) who are 1-week users, suggesting overall that 1-week use may be a common engagement pattern of digital health interventions. Thus, it is imperative to identify early who would likely become a 1-week user. For example, the baseline characteristics results of this study suggest that a younger age, smoking at least one-half pack per day, smoking the first cigarette within 5 minutes of waking (a marker of nicotine dependence), and scoring lower on acceptance of internal physical sensations that trigger smoking (a marker of avoidance of cigarette cravings) predicted membership in the iCanQuit 1-week trajectory group. Measuring these factors at baseline might allow for the early identification of individuals who would be more likely to disengage from iCanQuit in the first week. Another approach that might be worth testing in future research is investigating use patterns within the first week (eg, number of log-ins per day and time spent on the app per day) to predict whether a participant would become a 1-week user. Once identified, more intensive intervention strategies could be used with this group, which might range from push notification communications or proactive intervention. Beyond this study, it would be worthwhile to determine whether 1-week use is a common pattern across multiple digital platforms (eg, websites and smartphone apps) and health domains (eg, tobacco, exercise, and diet) and to what extent this use pattern affects health outcomes.

The second trajectory group for each arm was the 4-week users for iCanQuit (303/1069, 28.34%) and 3-week users for QuitGuide (369/1064, 34.68%). Although the length of each group was similar (3-4 weeks) and the proportion of each group was somewhat smaller in iCanQuit, only the iCanQuit 4-week users had significantly higher quit rates than their 1-week comparators (ie, 50% higher odds of quitting). Two potential reasons why iCanQuit's, but not QuitGuide's, second trajectory group had higher quit rates are most likely due to the content and structure of the iCanQuit app. Regarding content, we have published multiple studies showing that the effect of iCanQuit (but not QuitGuide) on smoking cessation was mediated by ACT-based processes of acceptance of internal cues to smoke (ie, sensations, thoughts, and emotions) [35,45-48]. The differences in content, with iCanQuit focused on ACT versus QuitGuide focused on standard USCPG content [49], suggest that 4 weeks of engaging with ACT content that targets acceptance of internal cues is effective at improving quit rates. Regarding structure, the iCanQuit app presented content in a sequenced interactive format (eg, content is unlocked in a sequential manner) with short paragraphs of text and some audio or visual for experiencing ACT concepts, whereas the QuitGuide app presented content in a sequenced format with short paragraphs of text [35]. Thus, the extent to which intervention engagement predicts behavior change might depend on the content and structure of the intervention, which is a valuable topic for future research. For iCanQuit, these results suggest that strategies to increase engagement 3 more weeks (ie, from 1 to 4) could be an effective approach to improving quit rates. Example strategies worth testing include (1) proactive check-ins (via SMS text messages or phone calls) from staff about progress with the iCanQuit app, (2) rewarding each day's use of iCanQuit, and (3) a "four-week challenge," which shows other users' daily log-in progress toward the goal of 4 weeks of use.

Only iCanQuit had a third trajectory group, namely 26-week users (156/1069, 14.59% of the iCanQuit arm sample). The 26-week users' group had nearly 400% higher odds of quitting smoking (as compared with 1-week users). The 12-month 56% quit rates observed in this group are the highest we have ever observed in a digital smoking cessation intervention and suggest that iCanQuit could be a highly effective and scalable intervention for this group of users. Our 2018 paper found a similar group of long-term users on the WebQuit website (159/1240, 12.82% of WebQuit arm sample) who had high 12-month quit rates (34.2% [24]) but not as high as those found here for iCanQuit. The iCanQuit ACT-based content and structure may have encouraged long-term, spaced skills practice [6]. Taken together, the findings for both the iCanQuit and WebQuit third trajectory groups suggest that consistent use of each program over time is prognostic of a better health outcome, which is contrary to the notion that consistent log-ins may be a marker of ongoing challenges and struggles to change a health behavior. Instead, consistent log-ins over time may be a marker of a participant's commitment to changing a health behavior. Digital intervention designs could focus on methods to encourage commitment and prevent lapses over time, which may include strategies similar to those suggested above, in addition to just-in-time adaptive interventions that aim to provide the right type of lapse-prevention support to smokers at the right time [50].

Personal Characteristics and Use Trajectories

The impact of personal characteristics on use trajectories appears to vary according to the intervention. For example, among iCanQuit participants, smoking the first cigarette >5 minutes after waking and higher levels of acceptance of physical cues to smoke predicted being a 4-week user, whereas older age and smoking ≤10 cigarettes per day predicted being a 26-week user. The findings generally suggest that less dependence and greater acceptance of cravings predict long-term engagement with iCanQuit. The results on increasing age predicting iCanQuit's 26-week use trajectory membership are consistent with past research showing that older age is a predictor of higher digital health intervention use [28-31], including our 2018 WebQuit trajectories paper [24]. In contrast to the view that as people age, their willingness to use technology decreases [51], this study suggests that increasing age may actually indicate which one is more likely to remain long-term users of iCanQuit, and in turn, have very high quit rates. In contrast, although there were baseline factors that predicted QuitGuide's use trajectory membership, neither of the 2 trajectories predicted smoking cessation, so the value of these baseline prediction results is unclear. Nonetheless, we recommend that future research explore a variety of baseline subgroup differences (eg, sex, race, and age) in digital intervention trajectories to better understand who is most or least likely to engage over time. Overall, these analyses suggest a need for further research on which baseline factors might predict different use trajectories and therefore inform the development of tailored interventions that facilitate long-term, consistent engagement based on an individual's specific baseline characteristics.

Limitations

This study has several key limitations. First, only 2 smartphone apps were tested, and both were focused on smoking cessation; thus, future research should examine the extent to which the results generalize to other behaviors and to other types of digital interventions. Cessation outcome data were self-reported for the reasons stated in the Methods section. Remote biochemical validation of smoking cessation would have introduced biases including low response rates, prohibitive cost, challenges with confirming the identity of the person providing the sample, and inability to confirm abstinence beyond 24 hours [52-59]. Owing to a technical error, log-ins were recorded for the first 6 months of the trial. Finally, as users self-select to different app use patterns (rather than being randomized), the associations observed in this study may not be causal, and care should be taken in their interpretation.

Conclusions

Patterns of 1-, 3-, and 4-week use of smartphone apps may be common for how people engage in digital health interventions. In addition, 4-week users, and especially 26-week users of iCanQuit, have higher odds of quitting smoking. Strategies to detect potential 1-week iCanQuit users and proactively offer them more intensive intervention could be fruitful.

Acknowledgments

The funders, National Cancer Institute (R01 CA192849) and Cancer Research UK (C1417/A22962), had no role in the trial conduct or interpretation of results.

Conflicts of Interest

OP is an unpaid member of the scientific advisory board for the Smoke Free app.

Multimedia Appendix 1

Summary of baseline characteristics of participants from the iCanQuit and QuitGuide arms by log-in trajectories.

[DOCX File , 25 KB - jmir_v24i8e39208_app1.docx]

References

1. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
2. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
3. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
4. Graham AL, Papandonatos GD, Cha S, Erar B, Amato MS. Improving adherence to smoking cessation treatment: smoking outcomes in a Web-based randomized trial. Ann Behav Med 2018 Mar 15;52(4):331-341 [FREE Full text] [doi: 10.1093/abm/kax023] [Medline: 29878062]
5. Graham AL, Zhao K, Papandonatos GD, Erar B, Wang X, Amato MS, et al. A prospective examination of online social network dynamics and smoking cessation. PLoS One 2017 Aug 23;12(8):e0183655 [FREE Full text] [doi: 10.1371/journal.pone.0183655] [Medline: 28832621]
6. Danaher BG, Smolkowski K, Seeley JR, Severson HH. Mediators of a successful Web-based smokeless tobacco cessation program. Addiction 2008 Oct;103(10):1706-1712 [FREE Full text] [doi: 10.1111/j.1360-0443.2008.02295.x] [Medline: 18715238]
7. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
8. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to Internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
9. Spaulding EM, Marvel FA, Piasecki RJ, Martin SS, Allen JK. User engagement with smartphone apps and cardiovascular disease risk factor outcomes: systematic review. JMIR Cardio 2021 Feb 03;5(1):e18834 [FREE Full text] [doi: 10.2196/18834] [Medline: 33533730]
10. Mclaughlin M, Delaney T, Hall A, Byaruhanga J, Mackie P, Grady A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180 [FREE Full text] [doi: 10.2196/23180] [Medline: 33605897]
11. Browne J, Halverson TF, Vilardaga R. Engagement with a digital therapeutic for smoking cessation designed for persons with psychiatric illness fully mediates smoking outcomes in a pilot randomized controlled trial. Transl Behav Med 2021 Sep 15;11(9):1717-1725. [doi: 10.1093/tbm/ibab100] [Medline: 34347865]
12. Ubhi HK, Michie S, Kotz D, van Schayck OC, Selladurai A, West R. Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features. Transl Behav Med 2016 Sep;6(3):410-417 [FREE Full text] [doi: 10.1007/s13142-015-0352-x] [Medline: 27528530]
13. Bendotti H, Lawler S, Ireland D, Gartner C, Hides L, Marshall H. What do people want in a smoking cessation app? An analysis of user reviews and app quality. Nicotine Tob Res 2022 Feb 01;24(2):169-177. [doi: 10.1093/ntr/ntab174] [Medline: 34460922]
14. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
15. Sridharan V, Shoda Y, Heffner J, Bricker J. A pilot randomized controlled trial of a Web-based growth mindset intervention to enhance the effectiveness of a smartphone app for smoking cessation. JMIR Mhealth Uhealth 2019 Jul 09;7(7):e14602 [FREE Full text] [doi: 10.2196/14602] [Medline: 31290404]

16. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
17. 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]
18. Shaughnessy JJ. Long-term retention and the spacing effect in free-recall and frequency judgments. Am J Psychol 1977 Dec;90(4):587-598. [doi: 10.2307/1421733]
19. Pyc MA, Rawson KA. Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? J Mem Lang 2009 May;60(4):437-447. [doi: 10.1016/j.jml.2009.01.004]
20. Christofferson DE, Hertzberg JS, Beckham JC, Dennis PA, Hamlett-Berry K. Engagement and abstinence among users of a smoking cessation text message program for veterans. Addict Behav 2016 Nov;62:47-53 [FREE Full text] [doi: 10.1016/j.addbeh.2016.06.016] [Medline: 27318948]
21. Heminger CL, Boal AL, Zumer M, Abroms LC. Text2Quit: an analysis of participant engagement in the mobile smoking cessation program. Am J Drug Alcohol Abuse 2016 Jul;42(4):450-458. [doi: 10.3109/00952990.2016.1149591] [Medline: 27120396]
22. Coa KI, Wiseman KP, Higgins B, Augustson E. Associations between engagement and outcomes in the SmokefreeTXT program: a growth mixture modeling analysis. Nicotine Tob Res 2019 Apr 17;21(5):663-669 [FREE Full text] [doi: 10.1093/ntr/nty073] [Medline: 29668984]
23. Zhao K, Wang X, Cha S, Cohn AM, Papandonatos GD, Amato MS, et al. A multirelational social network analysis of an online health community for smoking cessation. J Med Internet Res 2016 Aug 25;18(8):e233 [FREE Full text] [doi: 10.2196/jmir.5985] [Medline: 27562640]
24. Bricker JB, Sridharan V, Zhu Y, Mull KE, Heffner JL, Watson NL, et al. Trajectories of 12-month usage patterns for two smoking cessation websites: exploring how users engage over time. J Med Internet Res 2018 Apr 20;20(4):e10143 [FREE Full text] [doi: 10.2196/10143] [Medline: 29678799]
25. Perski O, Watson NL, Mull KE, Bricker JB. Identifying content-based engagement patterns in a smoking cessation website and associations with user characteristics and cessation outcomes: a sequence and cluster analysis. Nicotine Tob Res 2021 Jun 08;23(7):1103-1112 [FREE Full text] [doi: 10.1093/ntr/ntab008] [Medline: 33433609]
26. Golas SB, Nikolova-Simons M, Palacholla R, Op den Buijs J, Garberg G, Orenstein A, et al. Predictive analytics and tailored interventions improve clinical outcomes in older adults: a randomized controlled trial. NPJ Digit Med 2021 Jun 10;4(1):97 [FREE Full text] [doi: 10.1038/s41746-021-00463-y] [Medline: 34112921]
27. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Q-learning: a data analysis method for constructing adaptive interventions. Psychol Methods 2012 Dec;17(4):478-494 [FREE Full text] [doi: 10.1037/a0029373] [Medline: 23025434]
28. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in Internet-based interventions. Patient Prefer Adherence 2008 Feb 02;2:57-65 [FREE Full text] [Medline: 19920945]
29. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored Web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 04;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
30. Balmford J, Borland R, Benda P. Patterns of use of an automated interactive personalized coaching program for smoking cessation. J Med Internet Res 2008 Dec 17;10(5):e54 [FREE Full text] [doi: 10.2196/jmir.1016] [Medline: 19097975]
31. Zbikowski SM, Hapgood J, Smucker Barnwell S, McAfee T. Phone and Web-based tobacco cessation treatment: real-world utilization patterns and outcomes for 11,000 tobacco users. J Med Internet Res 2008 Nov 14;10(5):e41 [FREE Full text] [doi: 10.2196/jmir.999] [Medline: 19017583]
32. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, manage my pain: results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
33. Cillessen L, van de Ven MO, Compen FR, Bisseling EM, van der Lee ML, Speckens AE. Predictors and effects of usage of an online mindfulness intervention for distressed cancer patients: usability study. J Med Internet Res 2020 Oct 02;22(10):e17526 [FREE Full text] [doi: 10.2196/17526] [Medline: 33006567]
34. Goh G, Tan NC, Malhotra R, Padmanabhan U, Barbier S, Allen JC, et al. Short-term trajectories of use of a caloric-monitoring mobile phone app among patients with type 2 diabetes mellitus in a primary care setting. J Med Internet Res 2015 Feb 03;17(2):e33 [FREE Full text] [doi: 10.2196/jmir.3938] [Medline: 25648130]
35. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
36. Gifford EV, Kohlenberg BS, Hayes SC, Antonuccio DO, Piasecki MM, Rasmussen-Hall ML, et al. Acceptance-based treatment for smoking cessation. Behav Ther 2004;35(4):689-705 [FREE Full text] [doi: 10.1016/S0005-7894(04)80015-7]

37. Ramsay JO, Silverman BW. Applied Functional Data Analysis: Methods and Case Studies. New York, NY, USA: Springer; 2002.
38. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, NJ, USA: Wiley Online Library; 1990.
39. Tibshirani R, Walther G. Cluster validation by prediction strength. J Comput Graph Stat 2005;14(3):511-528. [doi: 10.1198/106186005X59243]
40. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002 Oct 15;21(19):2917-2930. [doi: 10.1002/sim.1296] [Medline: 12325108]
41. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th edition. New York, NY, USA: Springer; 2002.
42. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2020. URL: https://www.R-project.org/ [accessed 2022-03-28]
43. Goldsmith J, Scheipl F, Huang L, Wrobel J, Di C, Gellar J, et al. refund: Regression with Functional Data. The Comprehensive R Archive Network. 2022 Apr 16. URL: https://CRAN.R-project.org/package=refund [accessed 2022-03-28]
44. Hennig C. fpc: Flexible Procedures for Clustering. The Comprehensive R Archive Network. 2020. URL: https://CRAN.R-project.org/package=fpc [accessed 2022-03-28]
45. Bricker JB, Levin M, Lappalainen R, Mull K, Sullivan B, Santiago-Torres M. Mechanisms of smartphone apps for cigarette smoking cessation: results of a serial mediation model from the iCanQuit randomized trial. JMIR Mhealth Uhealth 2021 Nov 09;9(11):e32847 [FREE Full text] [doi: 10.2196/32847] [Medline: 34751662]
46. Santiago-Torres M, Mull KE, Sullivan BM, Kendzor DE, Bricker JB. Efficacy and utilization of smartphone applications for smoking cessation among low-income adults: secondary analysis of the iCanQuit randomized trial. Drug Alcohol Depend 2022 Feb 01;231:109258. [doi: 10.1016/j.drugalcdep.2021.109258] [Medline: 35026491]
47. Santiago-Torres M, Mull KE, Sullivan BM, Kwon D, Nollen NL, Zvolensky MJ, et al. Efficacy and utilization of an acceptance and commitment therapy-based smartphone application for smoking cessation among Black adults: secondary analysis of the iCanQuit randomized trial. Addiction 2022 Mar;117(3):760-771. [doi: 10.1111/add.15721] [Medline: 34890104]
48. Santiago-Torres M, Mull K, Sullivan B, Kwon DM, Nez Henderson P, Nelson LA, et al. Efficacy and utilization of smartphone applications for smoking cessation among American Indians and Alaska Natives: results from the iCanQuit trial. Nicotine Tob Res 2022 Mar 01;24(4):544-554. [doi: 10.1093/ntr/ntab213] [Medline: 34644389]
49. Fiore MC. Treating tobacco use and dependence: an introduction to the US Public Health Service Clinical Practice Guideline. Respir Care 2000 Oct;45(10):1196-1199. [Medline: 11203101]
50. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
51. Wandke H, Sengpiel M, Sönksen M. Myths about older people's use of information and communication technology. Gerontology 2012;58(6):564-570 [FREE Full text] [doi: 10.1159/000339104] [Medline: 22739502]
52. Cha S, Ganz O, Cohn AM, Ehlke SJ, Graham AL. Feasibility of biochemical verification in a Web-based smoking cessation study. Addict Behav 2017 Oct;73:204-208 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.020] [Medline: 28551588]
53. Herbec A, Brown J, Shahab L, West R. Lessons learned from unsuccessful use of personal carbon monoxide monitors to remotely assess abstinence in a pragmatic trial of a smartphone stop smoking app - a secondary analysis. Addict Behav Rep 2019 Jun;9:100122 [FREE Full text] [doi: 10.1016/j.abrep.2018.07.003] [Medline: 31193683]
54. Thrul J, Meacham MC, Ramo DE. A novel and remote biochemical verification method of smoking abstinence: predictors of participant compliance. Tob Prev Cessat 2018 May;4:20 [FREE Full text] [doi: 10.18332/tpc/90649] [Medline: 30417159]
55. Garrison KA, Pal P, O'Malley SS, Pittman BP, Gueorguieva R, Rojiani R, et al. Craving to quit: a randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2020 Mar 16;22(3):324-331 [FREE Full text] [doi: 10.1093/ntr/nty126] [Medline: 29917096]
56. van der Aalst CM, de Koning HJ. Biochemical verification of the self-reported smoking status of screened male smokers of the Dutch-Belgian randomized controlled lung cancer screening trial. Lung Cancer 2016 Apr;94:96-101. [doi: 10.1016/j.lungcan.2016.02.001] [Medline: 26973213]
57. Wong SL, Shields M, Leatherdale S, Malaison E, Hammond D. Assessment of validity of self-reported smoking status. Health Rep 2012 Mar;23(1):47-53 [FREE Full text] [Medline: 22590805]
58. Benowitz NL, Bernert JT, Foulds J, Hecht SS, Jacob P, Jarvis MJ, et al. Biochemical verification of tobacco use and abstinence: 2019 update. Nicotine Tob Res 2020 Jun 12;22(7):1086-1097 [FREE Full text] [doi: 10.1093/ntr/ntz132] [Medline: 31570931]
59. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]

Abbreviations

ACT: acceptance and commitment therapy
FPC: functional principal component
OR: odds ratio
PPA: point prevalence abstinence
USCPG: US Clinical Practice Guidelines

Edited by G Eysenbach; submitted 02.05.22; peer-reviewed by N Maglaveras, K Wiseman; comments to author 24.05.22; revised version received 03.06.22; accepted 13.07.22; published 18.08.22.

Please cite as:
Bricker JB, Mull KE, Santiago-Torres M, Miao Z, Perski O, Di C
Smoking Cessation Smartphone App Use Over Time: Predicting 12-Month Cessation Outcomes in a 2-Arm Randomized Trial
J Med Internet Res 2022;24(8):e39208
URL: https://www.jmir.org/2022/8/e39208
doi:10.2196/39208
PMID:35831180

Original Paper

Understanding the Interactions Between Driving Behavior and Well-being in Daily Driving: Causal Analysis of a Field Study

Paul Stephan[1], BA; Felix Wortmann[1], Prof Dr; Kevin Koch[1], MSc

Bosch IoT Lab, Institute of Technology Management, University of St Gallen, St Gallen, Switzerland

Corresponding Author:
Paul Stephan, BA
Bosch IoT Lab
Institute of Technology Management
University of St Gallen
Dufourstrasse 40a
St Gallen, 9000
Switzerland
Phone: 43 66303102233
Email: paul.stephan@student.unisg.ch

Abstract

Background: Investigating ways to improve well-being in everyday situations as a means of fostering mental health has gained substantial interest in recent years. For many people, the daily commute by car is a particularly straining situation of the day, and thus researchers have already designed various in-vehicle well-being interventions for a better commuting experience. Current research has validated such interventions but is limited to isolating effects in controlled experiments that are generally not representative of real-world driving conditions.

Objective: The aim of the study is to identify cause–effect relationships between driving behavior and well-being in a real-world setting. This knowledge should contribute to a better understanding of when to trigger interventions.

Methods: We conducted a field study in which we provided a demographically diverse sample of 10 commuters with a car for daily driving over a period of 4 months. Before and after each trip, the drivers had to fill out a questionnaire about their state of well-being, which was operationalized as arousal and valence. We equipped the cars with sensors that recorded driving behavior, such as sudden braking. We also captured trip-dependent factors, such as the length of the drive, and predetermined factors, such as the weather. We conducted a causal analysis based on a causal directed acyclic graph (DAG) to examine cause–effect relationships from the observational data and to isolate the causal chains between the examined variables. We did so by applying the backdoor criterion to the data-based graphical model. The hereby compiled adjustment set was used in a multiple regression to estimate the causal effects between the variables.

Results: The causal analysis showed that a higher level of arousal before driving influences driving behavior. Higher arousal reduced the frequency of sudden events (P=.04) as well as the average speed (P=.001), while fostering active steering (P<.001). In turn, more frequent braking (P<.001) increased arousal after the drive, while a longer trip (P<.001) with a higher average speed (P<.001) reduced arousal. The prevalence of sunshine (P<.001) increased arousal and of occupants (P<.001) increased valence (P<.001) before and after driving.

Conclusions: The examination of cause–effect relationships unveiled significant interactions between well-being and driving. A low level of predriving arousal impairs driving behavior, which manifests itself in more frequent sudden events and less anticipatory driving. Driving has a stronger effect on arousal than on valence. In particular, monotonous driving situations at high speeds with low cognitive demand increase the risk of the driver becoming tired (low arousal), thus impairing driving behavior. By combining the identified causal chains, states of vulnerability can be inferred that may form the basis for timely delivered interventions to improve well-being while driving.

(J Med Internet Res 2022;24(8):e36314) doi:10.2196/36314

KEYWORDS

well-being; daily driving; causal inference; commute; field study; directed acyclic graph; just-in-time interventions; mental well-being; stress; mental health

Introduction

With rising numbers of mental disorders worldwide, maintaining well-being has become an important public health issue [1], with a special focus on untreated cases [2]. In recent years, there has been an increased interest in investigating ways to improve well-being in everyday situations in order to prevent mental disorders [3,4]. Especially interventions aimed at improving well-being in moments when a person is susceptible to a deteriorating mental state, a so-called *state of vulnerability*, have shown great promise [5,6]. In this regard, the just-in-time adaptive intervention (JITAI) framework has recently gained major attraction as it guides researchers in their way to develop effective interventions that are delivered at the right time and in the right situation. To implement such a JITAI effectively, a profound understanding of the contextual factors leading to a state of vulnerability is necessary [7].

A suitable everyday situation in which an intervention may promise beneficial effects on well-being is daily driving [8]. Daily car commuters spend a considerable amount of time on the street every day, often associated with events causing frustrations and loss of time, such as caused by traffic jams, congestion, and unpredictability [9,10]. Kahneman et al [11] found the daily commute to be one of the least pleasant activities of the day. Accordingly, states of deteriorating well-being are likely to occur in daily driving. Simultaneously, the car is a suitable place for JITAIs as there are multiple sensors to detect the current driving conditions (eg, lane or traffic object detection [12,13]) and driver states (eg, arousal states [14] or emotions [15,16]) as well as to deliver interventions using advanced multimedia systems.

Recent work investigated when drivers are interruptible by [17,18] or even responsive to interventions [19] while driving. Moreover, researchers designed and validated the effect of well-being interventions that can be conducted while driving, for example, breathing exercises [20] and music or mindfulness experiences [21]. According to the JITAI framework, interventions are most effective when triggered in a state of reduced well-being [7]. To identify such states of vulnerability and thereby improve road safety [22], the factors influencing well-being while driving must be determined. Since well-being likely influences driving behavior and vice versa, a thorough understanding of the causal relationships is crucial for evaluating the driver's mental state. Therefore, this study examines the interactions between driving behavior and well-being during daily driving.

Because drivers are exposed to a variety of contextual factors, it is difficult to establish robust causal relationships based on existing statistical analysis [23]. Previous studies on driving and well-being have been primarily limited to isolating specific relationships in simulation experiments [24-28]. However, this controlled environment limits the results as stimuli are artificially induced, and thus effects do not generalize well to the wide range of situations encountered in everyday road traffic [29-31]. To thoroughly understand the relationship between driving and well-being, it seems necessary to study real-world data using novel methods for causal analysis.

The aim of this analysis is to unveil a robust causal architecture, that is, the underlying network of causes and effects [32], between driving behavior and well-being. We applied novel causal inference algorithms to derive these relationships from complex observational data collected in a real-world driving study, in which we investigated drivers' well-being over a period of 4 months. The derived causal architecture forms a basis for inferring states of vulnerability that can be targeted by digital interventions.

Methods

Field Study Setting and Variables

The data were gathered in a field study in which we handed over to 10 participants between the ages of 26 and 55 years a car each for daily driving. For maximizing external validity, we selected a broad spectrum of typical daily commuters with different demographics, life and family situations, and driving habits (purposive sampling). Detailed information about the participants is documented in Multimedia Appendix 1. Participants completed most of their driving in their residence area, which for all of them was the region around Stuttgart (Germany). The field study lasted for a period of 4 months, from July to November 2019.

To measure a wide variety of factors that could impact the well-being of the driver, we retrofitted the study cars for data collection. The participants self-assessed their current emotional state before and after driving based on questionnaires by a smartphone mounted next to the multimedia system. Moreover, we installed in every car a data collection system that recorded various variables from the vehicle in high frequency (eg, the steering wheel speed, brake pedal and gas pedal positions, or the Global Positioning System [GPS] location) to measure the driving behavior as well as the vehicle state. Our final data set comprised 13 variables that were classified into 4 categories: emotions, driving behavior, trip-dependent factors, and predetermined factors. We chose these 4 categories based on related work examining driving behavior and well-being [19,21]. A detailed list of the included variables can be found in Multimedia Appendix 2. This set of variables provides a comprehensive exploratory basis for understanding how driving behavior and well-being relate to each other. We explain these categories in the following paragraphs.

The *emotions* of the driver were assessed according to the circumplex model of affect, which is composed of the 2 dimensions *arousal* and *valence* [33,34]. The arousal dimension describes the drivers' feeling of being awake, and the valence dimension indicates the corresponding level of happiness. Before and after each drive, the driver indicated their state of arousal and valence on a scale from 0 (very low) to 100 (very high) using the *Affective Slider* [35], which is depicted in Multimedia Appendix 3.

We measured the *driving behavior* using the sensors that each car was equipped with. To quantify the driving behavior, we analyzed the steering and braking behavior of the driver. The *steeringbehavior* was quantified as the proportion of the trip in which the driver was turning the steering wheel. Analogously, the *brakingbehavior* reflected the ratio of the seconds in which the brake pedal was engaged to the total duration of the trip. To account for the risk that a driver takes, we included the frequency of sudden accelerations, sudden braking, or sudden steering as the variable *sudden events*. An event was classified as sudden when the acceleration or steering angle exceeded a prespecified threshold. To identify these events, we used the same approach as in prior work [19] based on the peak detection algorithm from the Python package SciPy [36].

Since the field study was conducted in an uncontrolled setting, we needed to account for contextual and environmental factors that drivers experience. We distinguished between *trip-dependent factors*, which are related to the drive itself, and *predetermined factors*, which are explicitly known to drivers before starting. The *trip-dependent factors* comprise information about the *length* of the trip in seconds and the average *speed* in kilometers per hour. In addition, the *flow* of the trip quantifies the ratio of the actual speed to the potential maximum speed throughout the trip. By combining these 3 factors, we aimed at representing the built environment in which the trip takes place [37]. For example, urban driving will likely result in short trips at low speeds with low flow.

The included *predeterminedfactors* in the causal model were the weather, quantified by the minutes of *sunlight* in the hour that the drive started and whether the trip was or was not performed on the *weekend*. Furthermore, we recorded whether another *occupant* was present and whether the trip was a *commute* between home and work, as derived from the GPS location at the beginning and end of each trip. To reduce skewness and to establish a common scale across all continuous variables, we standardized the data to a mean of 0 and an SD of 1.

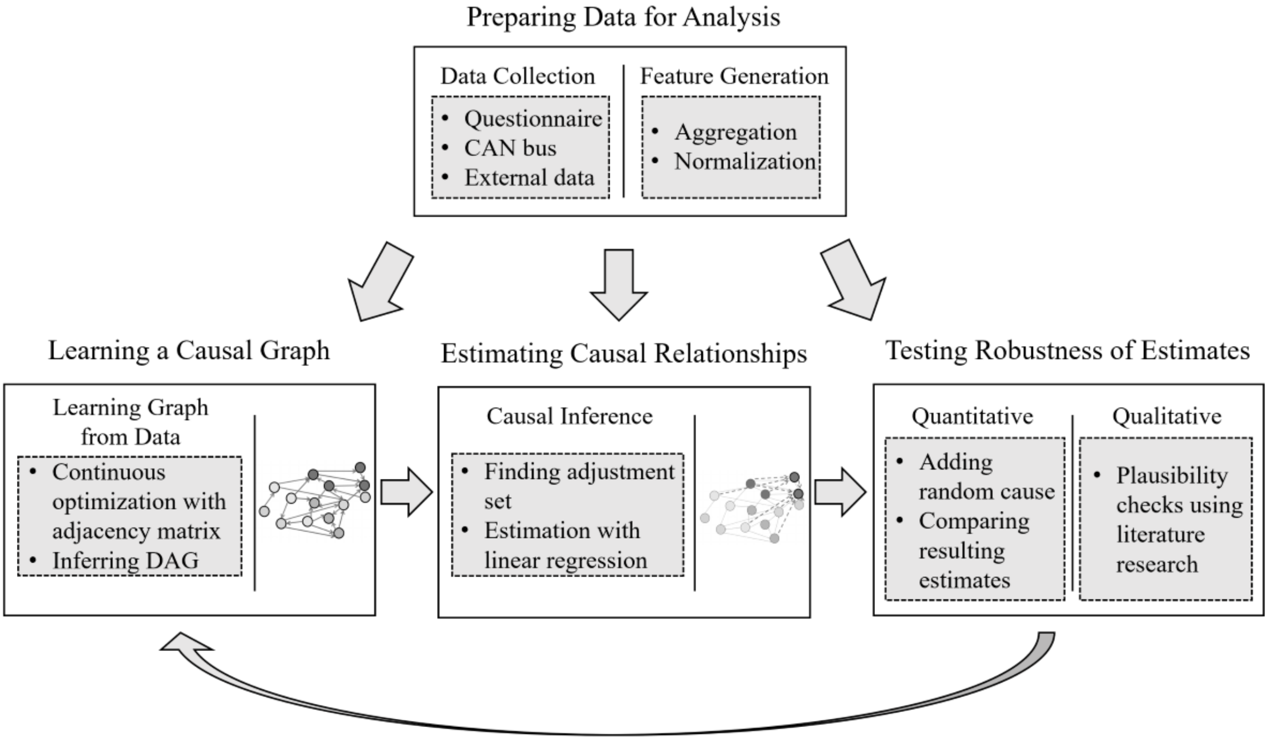
Ethical Considerations

This field study was reviewed by the Institutional Review Board of the University of Bern, Switzerland (approval #2019.04-00003).

Establishing a Causal Relationship

We conducted a field study to examine well-being in real-world driving situations, while maximizing the generalizability of our results. However, this purely observational study design comes at the cost of controllability. Therefore, the interactions between factors of well-being and driving behavior cannot be directly estimated as in a controlled experiment. Instead, we propose a framework for causal inference based on a causal directed acyclic graph (DAG), which visually represents the causal architecture formed by all recorded variables [38-68]. Our workflow for causal inference was designed as a 3-step process, as illustrated in Figure 1. Detailed theoretical information about the workflow and the causal methodology can be found in Multimedia Appendix 4.

Figure 1. Our workflow for causal analysis. CAN: controller area network (car sensor data); DAG: directed acyclic graph.



The causal DAG was constructed using the *DAG with NOTEARS* algorithm [69] on the data of the field study. This algorithm performs continuous optimization on a matrix representation of the graph rather than using constraint-based or local methods for inferring a graphical model. Thereby, a single graph maximizing the score function of the algorithm is found. We used the resulting graph to determine paths, which depict relationships between variables. To isolate the effect of one variable on another, the paths carrying spurious associations must be eliminated, while preserving the paths that transmit the causal effect. We isolated the relevant paths by applying the *backdoor criterion* [47], which identifies the set of variables that need to be controlled. This so-called *adjustment set* was subsequently used in a multiple regression to identify the causal effect between the variables of interest.

For determining the robustness of the resulting estimates, the effect was recalculated in a DAG with an added random confounder [65]. More specifically, a difference between the original and the new estimate close to 0 indicates that an effect is robust to unobserved confounders. Moreover, this test indicates the robustness of the estimate against a potential violation of the linear regression assumptions. Additionally, trivially impossible effects, such as an effect from arousal after to before driving, were a priori excluded from the causal DAG. A full list of excluded effects can be found in Multimedia Appendix 5.

Results

Descriptive Results

The 10 participants completed on average 163.8 trips (SD 89.28) during the 4 months of the study. The mean duration of a trip was 29 minutes (SD 20), and an average participant drove 19.5 km (SD 28.15). Our data set appears to cover typical daily driving, as the drivers followed a large variety of routes both in urban and in rural areas. Of the 1638 trips, 1343 (82%) took place during the week. In addition, 393 (24%) of the trips were labeled as a commute, and 1245 (76%) were drives to frequently visited destinations. On all these trips, the participants completed the affective slider. The affective slider results before driving were on average 73.66 (SD 18.24) for arousal and 69.76 (SD 16.74) for valence and after driving were on average 71.53 (SD 19.58) for arousal and 69.26 (SD 16.91) for valence.

Causal Analysis

To understand which factors impact well-being, we conducted a causal analysis based on the causal DAG learned from data, which can be found in Multimedia Appendix 6. The causal effect sizes, hereinafter abbreviated as *CE*, describe the impact of a 1-SD change in the source variable on the target variable. As all variables were standardized and to facilitate comparisons, the resulting effect size is also given in SDs. The statistically significant (α=.05) causal effects grouped by origin and target nodes are listed in Table 1. The robustness test reports the difference between the causal estimate from the analysis and the causal estimate when adding a random confounder to the model. A value close to 0 indicates robust causal estimates.

Table 1. Results of our causal analysis.

Source	Target	CE[a]	95% CI	*P* value	Robustness test
Emotions on driving behavior					
Before arousal	Steering	0.13	0.09- 0.16	<.001	−0.00003
Before arousal	Sudden events	−0.06	–0.11 to 0.02	.04	0.00031
Before arousal	Speed	−0.11	–0.13 to –0.07	.001	−0.00104
Driving behavior on emotions					
Braking	After arousal	0.10	0.06-0.14	<.001	−0.00089
Speed	After arousal	−0.17	–0.20 to –0.12	<.001	0.00001
Predetermined factors on emotions					
Sun	Before arousal	0.12	0.08-0.18	<.001	−0.00048
Sun	After arousal	0.14	0.11-0.18	<.001	0.00016
Occupants	Before valence	0.38	0.26-0.49	<.001	−0.00008
Occupants	After valence	0.37	0.24-0.51	<.001	0.00021
Trip-dependent factors on emotions					
Length	After arousal	−0.13	–0.45 to –0.06	.001	−0.00073
Emotions on emotions					
Before arousal	Before valence	0.18	0.14-0.22	.002	−0.00084
Before arousal	After arousal	0.74	0.70-0.76	<.001	0.00000
Before arousal	After valence	0.19	0.15-0.24	.01	0.00009
Before valence	After valence	0.77	0.74-0.80	<.001	−0.00161
After valence	After arousal	0.13	0.08-0.18	<.001	0.00013

[a] CE: causal effect size.

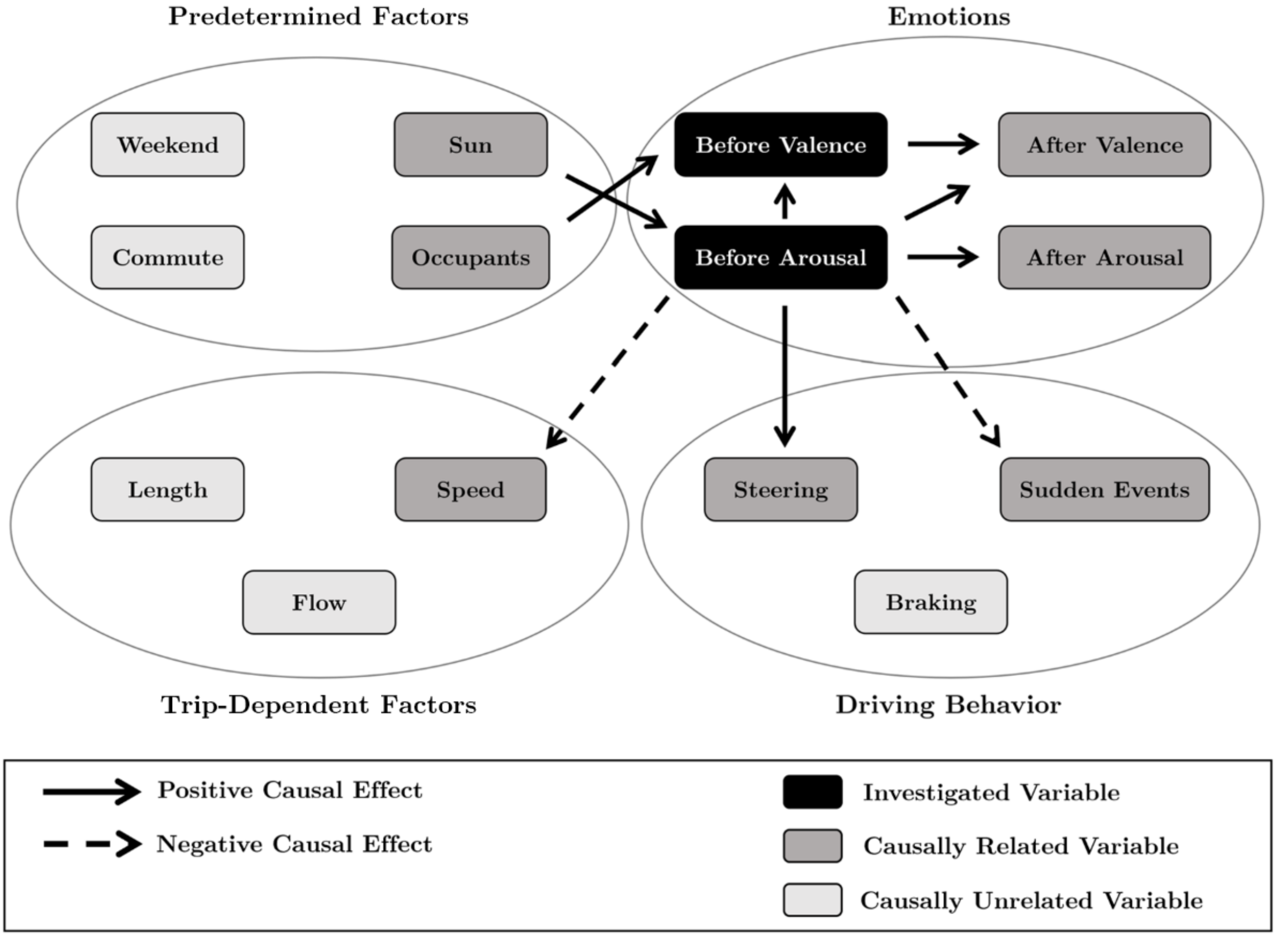
For developing a better understanding of the interaction between driving and well-being, we investigated the effects in both directions (ie, well-being on driving and driving on well-being). In the following paragraphs, we report on the significant results (α=.05). All effects that we discuss are highly robust with respect to omitted variables and to violations of the linear regression assumptions as the change in the causal estimate is smaller than 0.001 after adding a random confounder.

Effects Related to Well-being Before Driving

Figure 2 shows all causal effects related to well-being before driving. The analysis showed that before-driving emotions cause changes in the driving behavior as well as in trip-dependent factors. Regarding behavioral variables, a higher level of *before-driving arousal* significantly increased the frequency of *steering* (CE=0.13, *P*<.001) and decreased the occurrence of *sudden events* (CE=−0.06, *P*=.04). More specifically, these effects mean that an increase of 1 SD of arousal prevented 8 sudden events per hour. Moreover, higher *before-driving arousal* decreased the *speed* of trips (CE=−0.11, *P*=.001), which amounted to 2 km/hour per 1 SD of arousal. Moreover, a significant interaction between emotions existed. *Before-driving arousal* positively influenced the level of *before-driving valence* (CE=0.18, *P*=.002). *Before-driving valence* had no statistically significant effects on driving behavior or on trip-dependent factors.

Before-driving emotions were also influenced by predetermined factors. The presence of *occupants* caused an increase in *before-driving valence* (CE=0.38, *P*<.001), and more *sunlight* caused higher levels of *before-driving arousal* (CE=0.12, *P*<.001). The variable *weekend* had no causal impact on the *before-driving valence* or the *arousal* of the participants.

Figure 2. Causal effects regarding well-being before driving.



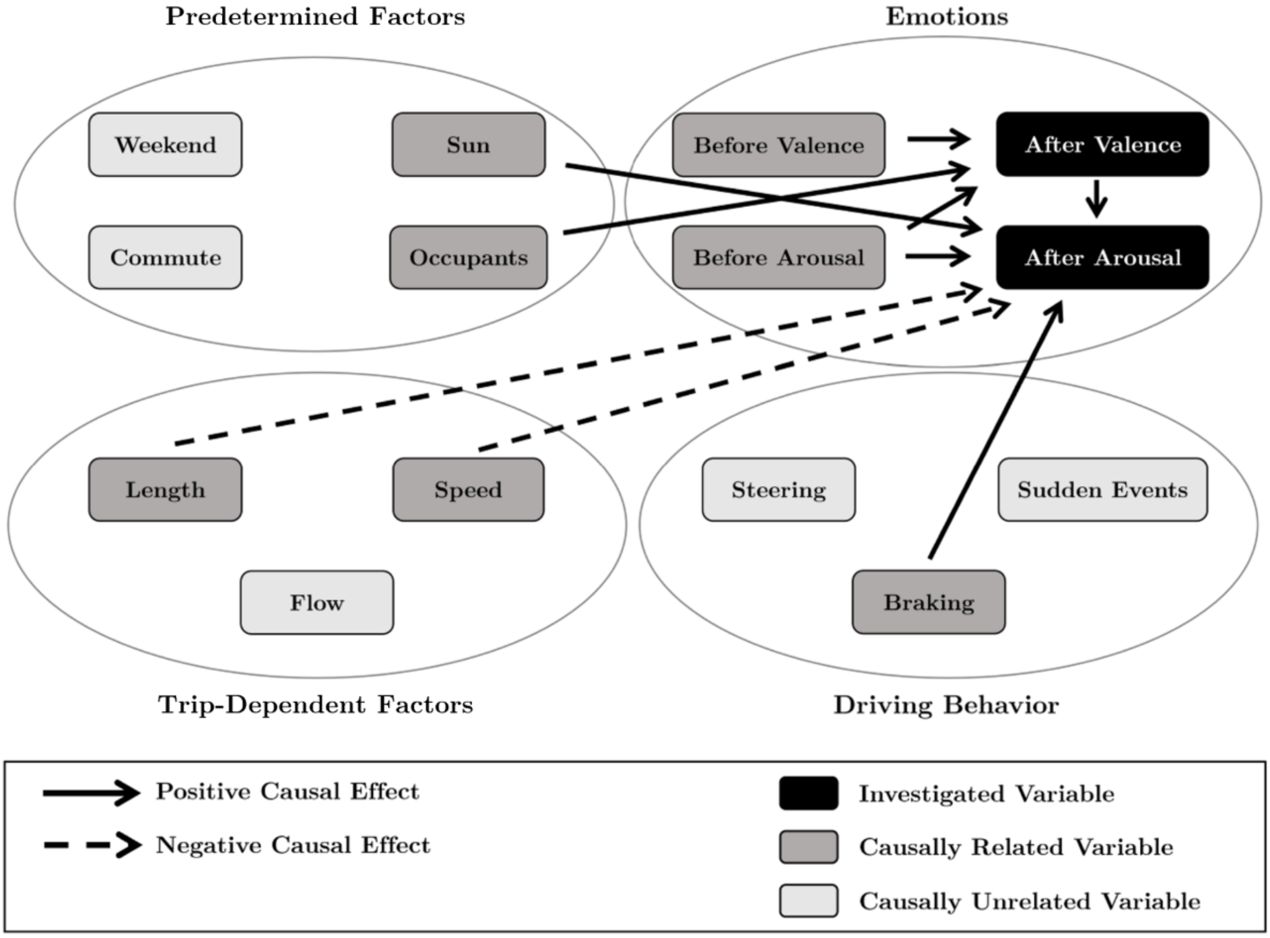
Effects Related to Well-being After Driving

Figure 3 shows the causal effects related to well-being after driving. The emotions after driving a car are influenced by the driving behavior as well as by trip-dependent and predetermined factors. Variables related to actual driving showed that both higher average *speeds* (CE=−0.17, *P*<.001) as well as longer *trips* (CE=−0.13, *P*=.001) caused lower levels of *arousal*. Moreover, driving behavior had an influence, with more frequent *braking* increasing the *after-driving arousal* (CE=0.1, *P*<.001).

Analogously to the effects of predetermined factors before the trip, *sunlight* increased *after-driving arousal* (CE=0.14, *P*<.001) and the presence of *occupants* increased *after-driving valence* (CE=0.37, *P*<.001). Thus, the effect of *sunlight* on *arousal* was stronger after than before driving, whereas the effect from *occupants* on *valence* was smaller after driving.

The emotions before starting the trip strongly influenced the emotions after having completed the trip. This relationship was especially evident when examining the causal effects from *before-driving* to *after-driving arousal* (CE=0.74, *P*<.001) and *valence* (CE=0.77, *P*<.001). Further, a significant interaction existed between *emotions*, with the *before-driving arousal* influencing the level of *after-driving valence* (CE=0.19, *P*=.005). In addition, higher *after-driving arousal* states causally increased *after-driving valence* (CE=0.13, *P*<.001).

Figure 3. Causal effects regarding well-being after driving.



Discussion

Principal Findings

The results of our field study indicate that well-being significantly influences driving behavior and vice versa. Moreover, we found effects from predetermined and trip-dependent factors on valence. In the following paragraphs, we highlight our findings and contextualize them with potential explanations.

We found a significant impact of arousal on several driving behavior variables. With higher levels of arousal, drivers had fewer sudden maneuvers, steered more, and drove at lower speeds. We explain these effects with improved alertness due to high arousal [70]. More alert drivers react faster and in a more controlled manner to unexpected events. Therefore, they can proactively avoid sudden driving maneuvers, which reduces the risk of accidents [24]. Moreover, this anticipatory driving behavior with higher arousal leads to more steering and is potentially a sign of active control of the vehicle. Due to alertness and anticipatory driving, drivers may also proactively adapt the speed of the vehicle earlier to changing driving situations, which results in lower speeds.

For the inverse relationship, we found significant effects showing that driving-related factors impact the arousal of drivers. First, the higher the average speed was, the lower the after-driving arousal state was. Second, we found that the length of the trip negatively influences arousal. Since this deterioration of arousal is counteracted by frequent braking, we assume that monotonous driving situations (ie, long trips at high speeds with no need to brake frequently) cause a decrease in arousal. Cognitive tasks, such as braking, seemed to interrupt the perceived monotony and, thus, reduced the negative effect on arousal throughout the trip.

In contrast, we could not identify any statistically significant effect between valence and driving. The missing impact of low flow or sudden events on valence may be explained by the high driving experience of the participants, who may have grown accustomed to these conditions (eg, daily experience of traffic jams on commutes). However, the lack of effects may also be explained by a possible transient impact of adverse events, such as a traffic jam or consecutive red lights. After having reached the destination, these occurrences may have been forgotten and other thoughts may determine the disclosed end-of-trip valence. Further studies should evaluate the immediate impact of adverse conditions on valence.

Besides the actual driving, we identified predetermined factors that influenced well-being. First, more sunlight (ie, better weather) increased before- and after-driving arousal. Sunlight is known to impact daily mood in general and to reduce tiredness [71]. Second, the presence of occupants increased before- and after-driving valence. The explanation of this effect may be that social interaction is associated with a better sense of well-being [72]. In contrast, occupants had no influence on the arousal of drivers. Although occupants may reduce the monotony of a drive, the social interactions may also lead to social fatigue [73] and thus limit a potential arousal improvement.

Furthermore, we found significant interactions between the dimensions of well-being. The levels of arousal and valence before driving were highly correlated with the respective levels after driving. Most likely, carry-over effects occur, for example, awake drivers are still more awake at the end of the trip than drivers who already started while feeling tired. Moreover, both well-being dimensions are positively associated with each other. Building upon our prior reasoning, we propose 2 potential explanations. First, alert drivers experience fewer adverse events and therefore feel more positive by the end of the trip. Second, higher valence can make drivers more resistant to boredom, which reduces the feeling of monotony.

Comparison With Prior Work

This study aimed at investigating the complex causal architecture of well-being in daily driving using a real-world, uncontrolled field study. In contrast, previous studies have mainly focused on isolating specific effects in controlled experiments (eg, simulator studies). In the following paragraphs, we compare the significant effects between daily driving and well-being of our exploratory study to prior driving studies. These studies serve hereby as a plausibility check of our findings.

Our explanation of the positive impact of increased arousal on driving behavior is in line with the prior literature. Corfitsen [74] found in a survey combined with a reaction time test that low arousal states (ie, fatigue) are a major cause of longer reaction times while driving at night. Moreover, McGehee et al [24] showed in an experimental study on a test track that these longer reaction times are a major risk factor for accidents. However, our findings concerning valence differ from the previous literature. Prior simulator studies have revealed a significant negative effect of extreme valence states (very happy and very unhappy) on driving behavior [25]. The lack of effects of valence in our study could be explained by the setting of the field study. Whereas in the simulator study [25], strong valence-changing stimuli were induced, our study aimed to collect data on everyday driving situations with less strong valence changes.

Furthermore, we find confirmation that monotonous driving reduces the arousal of drivers. Thiffault and Bergeron [26] observed in a simulator study that continuous driving without any external stimulus induces fatigue and tiredness, which increases with time. Moreover, our conclusion that arousal levels are reduced by driving at high speeds due to the monotonous setting is supported by a simulator study by Ting et al [27]. Their study showed that highway driving leads to fatigue, which negatively affects driving performance and increases the risk of accidents, as priorly discussed [24,71]. We can further confirm that cognitive tasks, such as frequent braking, improve after-driving arousal. The results of a simulator experiment by Dunn and Williamson [28] showed that cognitive demand mitigates monotony.

Implications for Intervention Research and Practice

Our findings can be used to allow for more effective JITAIs by providing an estimate for when drivers are likely at risk of feeling tired or unhappy, that is, when they are in a state of vulnerability. By improving the well-being of the driver, such interventions have the potential to increase road safety and reduce the frequency of accidents.

Interventions for increasing well-being while driving can be conceptualized in 2 ways. First, the causes for states of vulnerability can be directly targeted according to our causal architecture. For instance, the findings indicate that a long drive with little braking and steering generates a monotonous driving situation, which sets the driver at risk of a state of low arousal. Second, interventions can react to detected states of vulnerability. For instance, if an increased number of sudden events, less steering, or increased speeds are recognized, it is an indication that the arousal of the driver has decreased. Thus, an intervention could be triggered that acts as a mental stimulus to increase arousal and thereby prevent drowsy driving. Past research developed and evaluated such interventions, for example, using highly personalized music playlists [21] or gamified driving challenges [75]. Our findings could pave the way for the ideation of new interventions. For instance, valence can be raised with an intervention that leads to social interaction, for example, by recommending calling someone during a break. As another example, a driver’s arousal deterioration due to high-speed driving could be addressed by reminding drivers about the speed limit.

Strengths and Limitations

We identified relationships between well-being and driving from a 4-month longitudinal field study on real roads with a sample representative of a wide range of commuters. To derive robust relationships, we applied causal inference methods in our analysis. This overall approach has multiple benefits. First, we observed in our study setup the true emotions participants experienced while they were driving. Contrary to laboratory experiments, which often induce or inspect single isolated effects, our results reflect realistic driving situations and therefore generalize better to the real world. Second, our findings can serve as a basis for delivering interventions to react to well-being changes impacting driving behavior. Finally, all identified effects are explainable by the prior literature and robust to violations of assumptions. Therefore, our study serves as a practical example that inferring causal architectures from observational field studies is feasible and may provide insights that go beyond the capabilities of controlled experiments.

The exploratory design of the study comes with some limitations. The analysis was conducted on the aggregated data set of trips during a longitudinal field study of naturalistic driving on public roads. Hence, it does not regard variation between drivers on a personal level. However, by combining experiences from 10 drivers, we could examine interactions that are present across multiple individuals. Further research could include a psychological analysis on the personal level and could combine valence and arousal to construct more complex emotions. Regarding the causal methodology, the *DAG with NOTEARS* algorithm does not definitely guarantee a precise causal DAG and is especially sensitive with respect to the scale of the variables [63]. We mitigated this issue by standardizing all continuous variables and by introducing a random confounder for testing the robustness of the estimates. Further research should establish a framework for assessing the robustness of the causal DAG itself.

Conclusion

In this paper, we unveiled the complex causal architecture of well-being in daily driving in a real-world field study. Daily driving is a complex setting in which many contextual and personal factors interact. In a real-world field study, this complexity can be replicated more adequately than in a controlled experiment. However, in observational studies, an elaborate causal methodology is necessary for identifying causal effects. Our study identified that arousal is more susceptible to changes while driving than valence. Especially monotonous driving situations, such as long drives on a highway without the need to decelerate or steer frequently, set the driver at risk of becoming more tired. This tiredness impairs driving behavior and can be seen as a state of vulnerability that can be utilized as a trigger for interventions. The knowledge about robust causal effects between well-being and driving behavior can therefore be applied as a basis for deciding when to initiate an intervention to improve the well-being of the driver.

Acknowledgments

This research was funded and supported by the Bosch IoT Lab of the University of St. Gallen and the Swiss Federal Institute of Technology (ETH), Zürich.

Data Availability

The raw data of the field study are available upon request, which should be sent to the corresponding author (PS). Any requests will be reviewed by the scientific study board leading the involved research group at the University of St. Gallen. Applications should outline the intended purpose of the data transfer. Only applications for noncommercial use will be considered. All data shared will be anonymized.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of participants.
[DOC File , 28 KB - jmir_v24i8e36314_app1.doc]

Multimedia Appendix 2
Further information about variables.
[DOC File , 40 KB - jmir_v24i8e36314_app2.doc]

Multimedia Appendix 3
Well-being questionnaire.
[DOC File , 171 KB - jmir_v24i8e36314_app3.doc]

Multimedia Appendix 4
Detailed methodology.
[DOC File , 233 KB - jmir_v24i8e36314_app4.doc]

Multimedia Appendix 5
A priori excluded direct effects.
[DOC File , 49 KB - jmir_v24i8e36314_app5.doc]

Multimedia Appendix 6
Data-based causal directed acyclic graph.
[DOC File , 271 KB - jmir_v24i8e36314_app6.doc]

References

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the

Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
2. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
3. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
4. Hao T, Rogers J, Chang H, Ball M, Walter K, Sun S, et al. Towards precision stress management: design and evaluation of a practical wearable sensing system for monitoring everyday stress. iproc 2017 Sep 22;3(1):e15 [FREE Full text] [doi: 10.2196/iproc.8441]
5. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
6. Smyth J, Heron K. Is providing mobile interventions 'just-in-time' helpful? An experimental proof of concept study of just-in-time intervention for stress management. IEEE Wireless Health (WH) 2016:1-7. [doi: 10.1109/wh.2016.7764561]
7. Sarker H, Hovsepian K, Chatterjee S, Nahum-Shani I, Murphy S, Spring B, et al. From markers to interventions: the case of just-in-time stress intervention. In: Rehg J, Murphy S, Kumar S, editors. Mobile Health. Cham: Springer; 2017:411-433.
8. Paredes PE, Hamdan NA, Clark D, Cai C, Ju W, Landay JA. Evaluating in-car movements in the design of mindful commute interventions: exploratory study. J Med Internet Res 2017 Dec 04;19(12):e372 [FREE Full text] [doi: 10.2196/jmir.6983] [Medline: 29203458]
9. Legrain A, Eluru N, El-Geneidy AM. Am stressed, must travel: the relationship between mode choice and commuting stress. Transp Res Part F Traffic Psychol Behav 2015 Oct;34:141-151. [doi: 10.1016/j.trf.2015.08.001]
10. Chatterjee K, Chng S, Clark B, Davis A, De Vos J, Ettema D, et al. Commuting and wellbeing: a critical overview of the literature with implications for policy and future research. Transp Rev 2019 Aug 01;40(1):5-34. [doi: 10.1080/01441647.2019.1649317]
11. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: the day reconstruction method. Science 2004 Dec 03;306(5702):1776-1780. [doi: 10.1126/science.1103572] [Medline: 15576620]
12. Zou Q, Jiang H, Dai Q, Yue Y, Chen L, Wang Q. Robust lane detection from continuous driving scenes using deep neural networks. IEEE Trans Veh Technol 2020 Jan;69(1):41-54. [doi: 10.1109/tvt.2019.2949603]
13. Fisher Y, Chen H, Wang X, Xian W, Chen Y, Liu F, et al. Bdd100k: a diverse driving dataset for heterogeneous multitask learning. 2020 Presented at: Conference on Computer Vision and Pattern Recognition; June 13-19, 2020; Seattle, WA p. 2636-2645. [doi: 10.1109/cvpr42600.2020.00271]
14. Sikander G, Anwar S. Driver fatigue detection systems: a review. IEEE Trans Intell Transp Syst 2019 Jun;20(6):2339-2352. [doi: 10.1109/tits.2018.2868499]
15. Humanizing Technology. URL: http://www.affectiva.com/ [accessed 2021-10-22]
16. Liu S, Koch K, Zhou Z, Föll S, He X, Menke T, et al. The Empathetic Car. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2021 Sep 09;5(3):1-34. [doi: 10.1145/3478078]
17. Kim A, Choi W, Park J, Kim K, Lee U. Interrupting drivers for interactionsi. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec 27;2(4):1-28. [doi: 10.1145/3287053]
18. Semmens R, Martelaro N, Kaveti P, Stent S, Ju W. Is now a good time? An empirical study of vehicle-driver communication timing. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-12. [doi: 10.1145/3290605.3300867]
19. Koch K, Mishra V, Liu S, Berger T, Fleisch E, Kotz D, et al. When do drivers interact with in-vehicle well-being interventions? An exploratory analysis of a longitudinal study on public roads. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Mar 19;5(1):1-30 [FREE Full text] [doi: 10.1145/3448116] [Medline: 35178497]
20. Balters S, Mauriello ML, Park SY, Landay JA, Paredes PE. Calm commute. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar 18;4(1):1-19. [doi: 10.1145/3380998]
21. Koch K, Tiefenbeck V, Liu S, Berger T, Fleisch E, Wortmann F. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3446865]
22. Hu T, Xie X, Li J. Negative or positive? The effect of emotion and mood on risky driving. Transp Res Part F Traffic Psychol Behav 2013 Jan;16:29-40. [doi: 10.1016/j.trf.2012.08.009]
23. Wagner A. Causality in complex systems. Biol Philos 1999 Jan;14(1):83-101. [doi: 10.1023/a:1006580900476]
24. McGehee DV, Mazzae EN, Baldwin GS. Driver reaction time in crash avoidance research: validation of a driving simulator study on a test track. 2016 Nov 05 Presented at: XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennium'; July 29 to August 4, 2000; San Diego, CA p. 3-320-3-323. [doi: 10.1177/154193120004402026]

25. Du N, Zhou F, Pulver EM, Tilbury DM, Robert LP, Pradhan AK, et al. Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. Transp Res Part C Emerg Technol 2020 Mar;112:78-87. [doi: 10.1016/j.trc.2020.01.006]
26. Thiffault P, Bergeron J. Fatigue and individual differences in monotonous simulated driving. Pers Individ Differ 2003 Jan 27;34(1):159-176 [FREE Full text] [doi: 10.1016/s0191-8869(02)00119-8]
27. Ting P, Hwang J, Doong J, Jeng M. Driver fatigue and highway driving: a simulator study. Physiol Behav 2008 Jun 09;94(3):448-453. [doi: 10.1016/j.physbeh.2008.02.015] [Medline: 18402991]
28. Dunn N, Williamson A. Driving monotonous routes in a train simulator: the effect of task demand on driving performance and subjective experience. Ergonomics 2012 Jul 17;55(9):997-1008. [doi: 10.1080/00140139.2012.691994] [Medline: 22803577]
29. Cook T. Quasi-experimental design. In: Wiley Encyclopedia of Management. Chichester, West Sussex, UK: John Wiley & Sons; 2015.
30. Flannelly KJ, Flannelly LT, Jankowski KRB. Threats to the internal validity of experimental and quasi-experimental research in healthcare. J Health Care Chaplain 2018 Jan 24;24(3):107-130. [doi: 10.1080/08854726.2017.1421019] [Medline: 29364793]
31. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006 Jan 01;13(1):16-23. [doi: 10.1197/jamia.m1749]
32. Keyes KM, Galea S. Commentary: the limits of risk factors revisited: is it time for a causal architecture approach? Epidemiology 2017;28(1):1-5. [doi: 10.1097/ede.0000000000000578]
33. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
34. Banholzer N, Feuerriegel S, Fleisch E, Bauer GF, Kowatsch T. Computer mouse movements as an indicator of work stress: longitudinal observational field study. J Med Internet Res 2021 Apr 02;23(4):e27121 [FREE Full text] [doi: 10.2196/27121] [Medline: 33632675]
35. Betella A, Verschure PFMJ. The affective slider: a digital self-assessment scale for the measurement of human emotions. PLoS One 2016 Feb 5;11(2):e0148037 [FREE Full text] [doi: 10.1371/journal.pone.0148037] [Medline: 26849361]
36. Jones E, Oliphant T, Peterson P. SciPy: Open Source Scientific Tools for Python. 2001. URL: http://www.scipy.org/ [accessed 2021-03-18]
37. Reyer M, Fina S, Siedentop S, Schlicht W. Walkability is only part of the story: walking for transportation in Stuttgart, Germany. Int J Environ Res Public Health 2014 May 30;11(6):5849-5865 [FREE Full text] [doi: 10.3390/ijerph110605849] [Medline: 24886755]
38. Verma T, Pearl J. Causal networks: semantics and expressiveness. Mach Intell Pattern Recognit 1990;9:69-76. [doi: 10.1016/B978-0-444-88650-7.50011-1]
39. Steel D. Social mechanisms and causal inference. Philos Soc Sci 2016 Aug 18;34(1):55-78. [doi: 10.1177/0048393103260775]
40. Peters J, Bühlmann P, Meinshausen N. Causal inference by using invariant prediction: identification and confidence intervals. J R Stat Soc B 2016 Oct 11;78(5):947-1012. [doi: 10.1111/rssb.12167]
41. Reichenbach H. The Direction of Time. 65th Volume. Berkeley: University of California Press; 1991.
42. Heckerman D. A bayesian approach to learning causal networks. In: Advances in Decision Analysis: From Foundations to Applications. Cambridge: Cambridge University Press; 2007:220.
43. Peters J, Janzing D, Schölkopf B. Elements of Causal Inference: Foundations and Learning Algorithms. Cambridge: MIT Press; 2017:9780262037310.
44. Ben-Gal I. Bayesian networks. In: Encyclopedia of Statistics in Quality and Reliability. Hoboken, NJ: John Wiley & Sons; 2007.
45. Pearl J, Mackenzie D. The Book of Why: The New Science of Cause and Effect. New York, NY: Basic Books; 2018.
46. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004 Sep;15(5):615-625. [doi: 10.1097/01.ede.0000135174.63482.43] [Medline: 15308962]
47. Pearl J, Glymour M, Jewell N. Causal Inference in Statistics: A Primer. Hoboken, NJ: John Wiley & Sons; 2016.
48. Pearl J. From Bayesian networks to causal networks. In: Mathematical Models for Handling Partial Knowledge in Artificial Intelligence. Boston, MA: Springer; 1995.
49. Bareinboim E, Brito C, Pearl J. Local Characterizations of Causal Bayesian Networks BT - Graph Structures for Knowledge Representation and Reasoning. Berlin, Heidelberg: Springer; 2012.
50. Hauser A, Bühlmann P. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. J Mach Learn Res 2012;13(1):2409 [FREE Full text]
51. Odgaard-Jensen J, Vist G, Timmer A, Kunz R, Akl EA, Schünemann H, et al. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2011 Apr 13(4):MR000012 [FREE Full text] [doi: 10.1002/14651858.MR000012.pub3] [Medline: 21491415]
52. Coyle P. How to Build a Bayesian Model in 30 Minutes or Less. Towards Data Science. URL: https://towardsdatascience.com/how-to-build-a-bayesian-model-in-30-minutes-or-less-fd7a23ca1ecf [accessed 2021-03-18]

53. Cano A, Masegosa AR, Moral S. A method for integrating expert knowledge when learning Bayesian networks from data. IEEE Trans. Syst., Man, Cybern. B 2011 Oct;41(5):1382-1394. [doi: 10.1109/tsmcb.2011.2148197]
54. Huang L, Cai G, Yuan H, Chen J. A hybrid approach for identifying the structure of a Bayesian network model. Expert Syst Appl 2019 Oct;131:308-320. [doi: 10.1016/j.eswa.2019.04.060]
55. Scutari M, Graafland C, Gutíerrez J. Who learns better Bayesian network structures: constraint-based, score- based or hybrid algorithms? 2018 Presented at: PGM 2018: International Conference on Probabilistic Graphical Models; September 11-14, 2018; Prague, Czech Republic.
56. Heinze-Deml C, Maathuis MH, Meinshausen N. Causal structure learning. Annu. Rev. Stat. Appl 2018 Mar 07;5(1):371-391. [doi: 10.1146/annurev-statistics-031017-100630]
57. Liu Z, Malone B, Yuan C. Empirical evaluation of scoring functions for Bayesian network model selection. BMC Bioinformatics 2012 Sep 11;13(S15):S14. [doi: 10.1186/1471-2105-13-s15-s14]
58. Yu Y, Chen J, Gao T, Yu M. DAG structure learning with graph neural networks. 2019 Presented at: ICML 2019: 36th International Conference on Machine Learning; June 10-15, 2019; Long Beach, CA.
59. Glover F. Tabu search—part I. ORSA J Comput 1989 Aug;1(3):190-206. [doi: 10.1287/ijoc.1.3.190]
60. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Mach Learn 2006 Mar 28;65(1):31-78. [doi: 10.1007/s10994-006-6889-7]
61. Eaton D, Murphy K. Bayesian structure learning using dynamic programming and MCMC. 2007 Presented at: 23rd Conference on Uncertainty in Artificial Intelligence; July 19-22, 2007; Vancouver, Canada p. 101-108.
62. Zheng B, Agresti A. Summarizing the predictive power of a generalized linear model. Stat Med 2000 Jul 15;19(13):1771-1781. [doi: 10.1002/1097-0258(20000715)19:13<1771::aid-sim485>3.0.co;2-p]
63. Kaiser M, Sipos M. Unsuitability of NOTEARS for causal graph discovery when dealing with dimensional quantities. Neural Process Lett 2022;54:1587-1595. [doi: 10.1007/s11063-021-10694-5]
64. QuantumBlack. Welcome to CausalNex’s API Docs and Tutorials!. URL: https://causalnex.readthedocs.io [accessed 2021-03-27]
65. Sharma A, Kiciman E. DoWhy: an end-to-end library for causal inference. DoWhy. 2022. URL: https://py-why.github.io/dowhy/v0.8/ [accessed 2021-03-27]
66. Pearl J. Causality. Cambridge: Cambridge University Press; 2009.
67. Park J, Liang M, Alpert JM, Brown RF, Zhong X. The causal relationship between portal usage and self-efficacious health information–seeking behaviors: secondary analysis of the health information national trends survey data. J Med Internet Res 2021 Jan 27;23(1):e17782 [FREE Full text] [doi: 10.2196/17782] [Medline: 33502334]
68. Faruqui SHA, Alaeddini A, Chang MC, Shirinkam S, Jaramillo C, NajafiRad P, et al. Summarizing complex graphical models of multiple chronic conditions using the second eigenvalue of graph Laplacian: algorithm development and validation. JMIR Med Inform 2020 Jun 17;8(6):e16372 [FREE Full text] [doi: 10.2196/16372] [Medline: 32554376]
69. Zheng X, Aragam B, Ravikumar P, Xing E. DAGs with NOTEARS: continuous optimization for structure learning. 2022 Presented at: Advances in Neural Information Processing Systems 31; December 4-9, 2017; Long Beach, USA.
70. Lindsley DB. Activation, arousal, alertness, and attention. In: States of Brain and Mind. Readings from the Encyclopedia of Neuroscience. Boston, MA: Birkhäuser; 1988:1-3.
71. Denissen JJA, Butalid L, Penke L, van Aken MAG. The effects of weather on daily mood: a multilevel approach. Emotion 2008 Oct;8(5):662-667. [doi: 10.1037/a0013497] [Medline: 18837616]
72. Sun J, Harris K, Vazire S. Is well-being associated with the quantity and quality of social interactions? J Pers Soc Psychol 2020 Dec;119(6):1478-1496. [doi: 10.1037/pspp0000272] [Medline: 31647273]
73. Zhang S, Zhao L, Lu Y, Yang J. Do you get tired of socializing? An empirical explanation of discontinuous usage behaviour in social network services. Info Manage 2016 Nov;53(7):904-914. [doi: 10.1016/j.im.2016.03.006]
74. Corfitsen M. Tiredness and visual reaction time among young male nighttime drivers: a roadside survey. Accident Anal Prev 1994 Oct;26(5):617-624. [doi: 10.1016/0001-4575(94)90023-x]
75. Bier L, Emele M, Gut K, Kulenovic J, Rzany D, Peter M, et al. Preventing the risks of monotony related fatigue while driving through gamification. Eur Transp Res Rev 2019 Oct 29;11(44):1-19 [FREE Full text] [doi: 10.1186/s12544-019-0382-4]

Abbreviations

CE: causal effect size
DAG: directed acyclic graph
GPS: Global Positioning System
JITAI: just-in-time adaptive intervention

Edited by T Leung; submitted 10.01.22; peer-reviewed by J Delgado-Ron, Z Dai; comments to author 02.06.22; revised version received 04.07.22; accepted 01.08.22; published 30.08.22.

Please cite as:
Stephan P, Wortmann F, Koch K
Understanding the Interactions Between Driving Behavior and Well-being in Daily Driving: Causal Analysis of a Field Study
J Med Internet Res 2022;24(8):e36314
URL: https://www.jmir.org/2022/8/e36314
doi:10.2196/36314
PMID:36040791

Original Paper

Identifying Patients With Inflammatory Bowel Disease on Twitter and Learning From Their Personal Experience: Retrospective Cohort Study

Maya Stemmer[1*], MSc; Yisrael Parmet[1*], PhD; Gilad Ravid[1*], PhD

[1] Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[*] all authors contributed equally

Corresponding Author:
Maya Stemmer, MSc
Department of Industrial Engineering and Management
Ben-Gurion University of the Negev
POB 653
Beer-Sheva, 84105
Israel
Phone: 972 8 6461434
Email: mayast@post.bgu.ac.il

Abstract

Background: Patients use social media as an alternative information source, where they share information and provide social support. Although large amounts of health-related data are posted on Twitter and other social networking platforms each day, research using social media data to understand chronic conditions and patients' lifestyles is limited.

Objective: In this study, we contributed to closing this gap by providing a framework for identifying patients with inflammatory bowel disease (IBD) on Twitter and learning from their personal experiences. We enabled the analysis of patients' tweets by building a classifier of Twitter users that distinguishes patients from other entities. This study aimed to uncover the potential of using Twitter data to promote the well-being of patients with IBD by relying on the wisdom of the crowd to identify healthy lifestyles. We sought to leverage posts describing patients' daily activities and their influence on their well-being to characterize lifestyle-related treatments.

Methods: In the first stage of the study, a machine learning method combining social network analysis and natural language processing was used to automatically classify users as patients or not. We considered 3 types of features: the user's behavior on Twitter, the content of the user's tweets, and the social structure of the user's network. We compared the performances of several classification algorithms within 2 classification approaches. One classified each tweet and deduced the user's class from their tweet-level classification. The other aggregated tweet-level features to user-level features and classified the users themselves. Different classification algorithms were examined and compared using 4 measures: precision, recall, F1 score, and the area under the receiver operating characteristic curve. In the second stage, a classifier from the first stage was used to collect patients' tweets describing the different lifestyles patients adopt to deal with their disease. Using IBM Watson Service for entity sentiment analysis, we calculated the average sentiment of 420 lifestyle-related words that patients with IBD use when describing their daily routine.

Results: Both classification approaches showed promising results. Although the precision rates were slightly higher for the tweet-level approach, the recall and area under the receiver operating characteristic curve of the user-level approach were significantly better. Sentiment analysis of tweets written by patients with IBD identified frequently mentioned lifestyles and their influence on patients' well-being. The findings reinforced what is known about suitable nutrition for IBD as several foods known to cause inflammation were pointed out in negative sentiment, whereas relaxing activities and anti-inflammatory foods surfaced in a positive context.

Conclusions: This study suggests a pipeline for identifying patients with IBD on Twitter and collecting their tweets to analyze the experimental knowledge they share. These methods can be adapted to other diseases and enhance medical research on chronic conditions.

(J Med Internet Res 2022;24(8):e29186) doi:10.2196/29186

KEYWORDS

patient identification; inflammatory bowel disease; IBD; user classification; Twitter; natural language processing; NLP; sentiment analysis

Introduction

Background

Social networking sites and web-based communities have served as alternative information sources for patients in recent years. Patients everywhere use social media to share health and treatment information, learn from each other's experiences, and provide social support. Mining these informative conversations may shed some light on patients' ways of life and support research on chronic conditions.

In recent years, text mining and social network analysis have been used to detect mentions of health on Twitter [1,2] or to track the spread of the COVID-19 pandemic and symptoms [3-5]. Regarding chronic conditions, previous research has focused on analyzing patients' tweets and uncovering their Twitter community [6-10]. Although a relatively large amount of research has been dedicated to diabetes or cancer, research on inflammatory bowel disease (IBD) is only just starting to consolidate.

IBD is a chronic inflammatory condition of the digestive system characterized by flares and remission states. The 2 primary diseases identified with IBD, Crohn disease and ulcerative colitis, are usually diagnosed in young patients (in the age range of 15-30 years). The incidence of IBD is rapidly increasing, and it has evolved into a global disease [11-14].

There are no medications or surgical procedures that can cure IBD. Treatment options can only help with symptoms, and they affect each patient differently. They involve prescription drugs and lifestyle-related solutions such as diets and therapies. Symptoms include abdominal pain, diarrhea, and fatigue, and severe cases may result in hospitalization or surgical interventions [15,16]. As chronic bowel diseases, both Crohn disease and ulcerative colitis require day-to-day care for drug consumption and special nutrition.

Patients describe IBD as an embarrassing disease that causes the immediate disruption of daily activities. They experience difficulties in adjusting to the changes it entails and consider themselves different from their peers. As IBD is characterized by frequent bowel movements, people do not hasten to share their disease with others [15,17-19]. According to patients with IBD, part of the embarrassment can be attributed to a lack of public awareness. Outsiders cannot see that a person's stomach hurts or that their bowels are scarred. The disease is invisible, and others might doubt that it exists [20,21].

The embarrassment caused by IBD and the need to confide in people with similar experiences help explain the creation of IBD-related communities on Twitter. By overcoming space and distance, Twitter users form a community that disregards physical boundaries or immobility. A sense of common ground can help break down barriers and enable conversation, increasing a person's willingness to share [22,23]. It may be easier to consult with other patients who can relate and better understand the situation based on personal experience. One can identify more closely with user stories similar to one's own and embrace their advice more easily [24]. When people disclose health information on Twitter, they expose themselves to a large variety of opinions and reduce the uncertainty about their disease [25].

Owing to the nature of IBD and its influence on the digestive system, patients with IBD are forced to deal with their disease daily, adhere to strict dietary regimens, and maintain a calm routine. Changes in nutrition or physical activity, which are currently tested by trial and error, result in a long and excruciating process for the patients. We can learn from their personal experiences and provide an additional foundation for existing medical knowledge of the disease by collecting and analyzing patients' social media data. Complementary recommendations based on the wisdom of the crowd can ease patients' lives and shorten the process of finding the right lifestyle for them.

Objective and Contribution

This study aimed to uncover the potential of using Twitter data to promote the well-being of patients with IBD by collecting and analyzing the personal experiences they share about their disease. We suggested a framework for identifying patients with IBD on Twitter and examining the content they share regarding their disease. We started by building a user classifier that distinguishes patients from other entities who talk about IBD on Twitter and then used the classifier to collect patients' tweets and explore the lifestyle-related treatments they undergo to cope with their disease.

This study focused on creating a pipeline for using Twitter data for identifying patients with IBD and exploring the information they share. Although each part of this study can be extended by trying other classification methods or enriching the analysis of the patients' tweets, this study shows the potential of using Twitter data to enhance medical knowledge of IBD. We showed that patients can be identified on Twitter based on their communication even using classic, simple classification algorithms. We compared the performances of 2 different approaches for user classification—a single instance (SI) learning approach and a multiple instance (MI) learning approach—and showed the benefits of using the latter. The preliminary analysis in the second part of this study showed that it is possible to derive health-related insights from self-reported tweets by patients.

Using the suggested framework to identify more patients and collect more of their data could uncover their sentiments toward the treatments they try or explore other aspects of the disease, such as its influence on patients' quality of life. The framework is also feasibly extended to other chronic conditions. It can be used to compare discussion patterns of patients with IBD with those of the general population or of patients with other chronic conditions.

Related Work

Twitter and Health

The study of social media in the context of health and well-being continues to position Twitter as a new medium for disseminating health-related information. Health-related tweets range from a simple toothache to more severe and chronic diseases such as diabetes, asthma, or cancer [9,10,26,27]. Patients with amyotrophic lateral sclerosis use Twitter as a means of communication, and local health departments in the United States use Twitter to educate and disseminate information related to diabetes [28,29]. Even a sensitive disease such as HIV is discussed on Twitter [30-32]. Communication patterns regarding who tweets about what and why vary by disease [26].

Twitter is a powerful tool for disseminating health information and an accessible platform for patients needing immediate social support or relief. It provides a collaborative environment for health-related conversations where patients with chronic illnesses share their health status daily. They use Twitter to exchange knowledge about lifestyle implications or better understand a medical procedure. Through Twitter, they can easily and conveniently reach a large audience and various opinions [33].

In total, 2 previous studies have presented models for detecting personal health mentions on Twitter and shown promising, scalable results [1,2]. However, their goal differs from ours as they considered all tweets that discussed a specific person's health condition as positive. In our study, we sought to identify patients with a specific disease. We not only classified tweets written by patients but also classified the users themselves.

Communication Patterns on Twitter

Different types of users communicate differently on Twitter. They connect differently with others, have different tweeting habits, and differ in style and linguistic content. Studying the conversational connections between Twitter users and text mining their tweets can help classify users based on their characteristics and identify different types of users [34-38].

Private individuals reflect mainly on their personal experiences or sentiments and tend to engage with others. They are both frequently mentioned and frequently mentioning other users. By contrast, organizations often point to external information sources via URLs and are not that active at connecting with others. They are frequently mentioned in tweets, perhaps as sources of information, but are much less inclined to mention other users [39,40].

By analyzing a user's screen name (ie, the username of their Twitter account) or their biography (ie, their Twitter user description), one can determine whether the user is an ordinary individual or an organization and reveal latent user properties [41,42].

Our study relies on those previous findings and constructs classification features that help differentiate patients with IBD from other users who tweet about the disease. We adapted and extended previous methods to cope with the different task of identifying patients with IBD on Twitter.

Twitter and IBD

Exploring the entities that engage in IBD-related discussions on Twitter reveals that patients with IBD are the most common type of users who talk about IBD on Twitter [43,44]. Patients with IBD use Twitter for sharing personal experiences and seeking social support. They exchange thoughts about symptoms and medications and recommend treatments to one another [45,46]. By sharing their life experiences with the disease on Twitter, patients fight disease invisibility and raise public awareness of IBD [47].

Perez et al [48] explored the IBD community on Twitter and identified the types of users who talk about the disease and the key topics they discuss. They categorized users based on their Twitter profiles by analyzing their screen names and biographies. In our study, we investigated a large set of classification features and suggested a model to detect patients with IBD on Twitter based on the way they communicate and the content they share.

Patients with IBD tend to be more emotional and negative than patients with other chronic conditions [49]. They usually express a negative sentiment when they talk about the disease and its symptoms but positively address the diets and drugs that help manage them [48]. Patients who engage in tweets offering social support are more likely to post positive tweets [50].

Unlike previous research related to patients' sentiments on Twitter [48-50], we focused our research on entity sentiment rather than the sentiment of the entire tweet. By analyzing patients' sentiments toward specific keywords related to nutrition and fitness, we uncovered the sentiments of certain lifestyles that influence the disease.

Methods

Overview

This study was conducted in 2 main stages. In the Patient Identification section, we built a user classifier that distinguishes patients from other entities who talk about IBD on Twitter. We considered three types of classification features: (1) features extracted from the user's activity on Twitter, (2) the content of the user's tweets, and (3) the social structure of the user's network. We compared the performances of several classification algorithms within 2 classification approaches: one that starts by classifying tweets separately and then deduces the user's class from their tweet-level classification and one that starts by aggregating tweet-level features to user-level features and then classifies the users themselves.

In the Analyzing Patients' Tweets section covering the second stage of the study, we derived insights regarding IBD from the personal experiences that patients share on Twitter. We collected lifestyle-related tweets by querying the Twitter application programming interface (API) for special keywords related to nutrition or fitness. We then filtered their authors using a classifier from the first stage of the study to obtain a collection of tweets where patients with IBD describe the different diets and physical activities they adopt to deal with their disease. We identified frequently mentioned lifestyles and used IBM Watson Service for entity sentiment analysis to assess their effectiveness.

Figures 1 and 2 describe the general flow of the 2 main stages of the study. Figure 1 describes how we used Twitter data to classify users and identify patients with IBD. Figure 2 demonstrates how we used the classification to analyze patients' tweets.

Figure 1. The general workflow of the first stage of the study: building a classifier of Twitter users for identifying patients with inflammatory bowel disease (IBD).

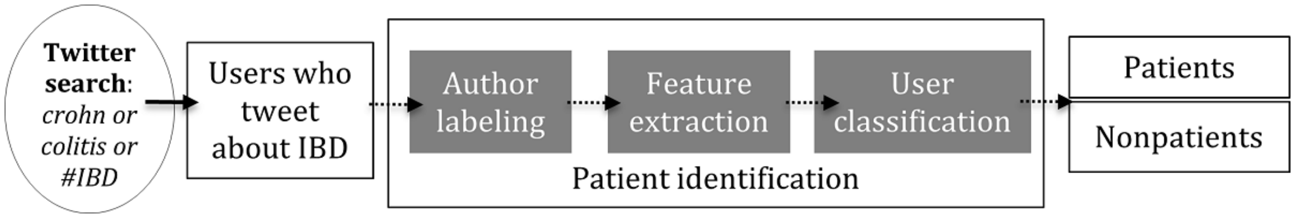
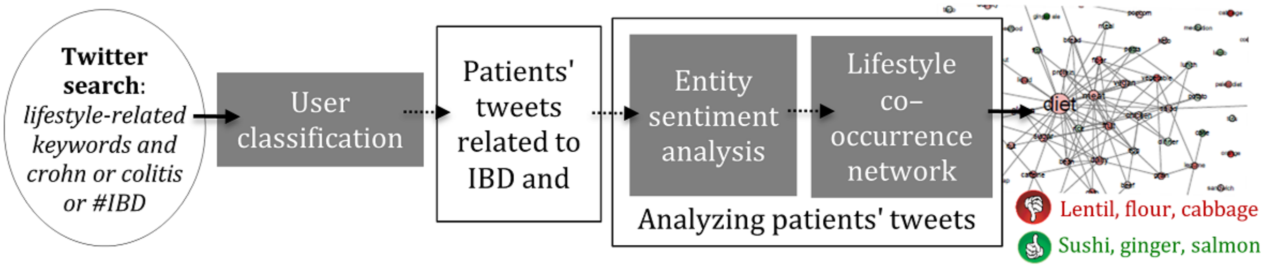


Figure 2. The general workflow of the second stage of the study: using the classification from the first stage for analyzing patients' tweets. IBD: inflammatory bowel disease.



Patient Identification

Data Collection and Preparation

We used the Twitter Search API to collect 10 days of IBD-related tweets (from February 11, 2018, at noon to February 21, 2018, at noon). We used the OR operator to search for at least one of 3 keywords: *crohn*, *colitis*, and *#IBD*. The abbreviation IBD was searched as a hashtag to avoid news-related tweets by the Investor's Business Daily Editorials account, which is usually marked with *IBD*. We limited the search to tweets written in English and collected 2045 tweets.

The 722 authors of the collected tweets were then manually classified as patients (1) or not (0). In total, 3 different annotators, the authors of this paper (MS, YP, and GR), did the labeling process and labeled the users based on their tweets. Each user received a tag of 1 if they had at least one tweet revealing their illness and a tag of 0 otherwise (ie, if none of their tweets suggested that they were patients with IBD).

Regarding 655 users (n=181, 27.6% patients and n=474, 72.4% other users), the annotators were in complete agreement, and their labels were set. To settle the dispute regarding the other 9.3% (67/722) of the users, the annotators challenged their tweet-based decisions by considering the users' screen names and biographies and reviewing their timelines if necessary. Considering the new data, of the 67 remaining users, 45 (67%) were classified as patients after explicitly mentioning their illness in their biographies or timelines. A total of 12% (8/67) talked about others who were sick, and the annotators unanimously agreed that they were not patients with IBD themselves. Regarding the remaining 21% (14/67) users, the annotators did not reach a consensus and, therefore, the users were omitted from the data set. The labeling process ended with a collection of 708 tagged users: 226 (31.9%) patients and 482 (68.1%) nonpatients.

To train the tweet-level classifiers, we had to annotate the tweets manually as well. We addressed the tweets collected in the original search query (in February 2018) and excluded retweets (RTs) from the collection. As the purpose was to identify patients, we were not interested in reshared content and only considered the user's tendency to RT as a behavioral classification feature. After excluding RTs and the 14 users for whom we did not reach an annotation consensus, we were left with 1687 tweets. To consider the users' biographies as we did when annotating users, we added each biography as another *tweet* by its author. A total of 83.5% (591/708) of the users had nonempty biographies, and the process resulted in a collection of 2278 *tweets*.

During the annotation process, we wanted to determine whether a certain tweet revealed that the user was a patient with IBD. Tweets that unambiguously implied that their authors were patients with IBD received a tag of 1, and all others received a tag of 0. As we had already annotated the users, all 1638 tweets written by nonpatient users automatically received a tag of 0. The 3 annotators (MS, YP, and GR) then manually classified all the tweets written by patients. A total of 346 tweets were unanimously classified as 1, and 288 tweets were unanimously classified as 0. The annotators did not reach a consensus on 6 tweets (written by 6 different users), and they were excluded from the collection. All 6 users had at least one more tweet and, therefore, none of them were excluded entirely from our data set. Finally, we reached a collection of 2272 tweets, of which 346 (15.23%) explicitly revealed their authors' illness.

To enrich our data, we collected another week of tweets (from June 10, 2018, at noon to June 17, 2018, at noon) for each tagged user, this time without additional filtering. In the months that had passed, 6.6% (47/708) of the users had been either suspended by Twitter or changed their accounts to private, and their data were no longer available for collection. The additional week was collected for the other 93.4% (661/708) of the users, and the process resulted in a data set of 82,884 tweets overall written by 194 patients and 467 nonpatients. We excluded the same 47 users from the tweet data set as well, and the final data set contained 2204 tweets, with 325 (14.75%) positive tweets.

MI Learning Approach

Traditional classification problems are supervised learning problems in which one receives a collection of individually labeled instances and tries to predict the class label for new instances. MI learning, by contrast, is a supervised learning approach in which each learning example is a *bag of instances* associated with 1 label, and the task is to predict the labels for unseen bags [51].

Previous research related to identifying health mentions on Twitter has relied on traditional supervised learning to determine whether a *tweet* discusses a health condition [1,2]. However, we wished to determine whether *patients* can be identified on Twitter and not examine the tweets separately. Our unique task and the unbalanced structure of our data were compatible with an MI learning approach—we had 661 users and a different number of tweets posted by each of them. Positive tags (patients) were determined collectively by finding at least one piece of evidence that the user had IBD; negative tags (nonpatients) meant that all the user's evidence suggested otherwise or, rather, was not sufficient for a positive tag.

We used the metadata-based MI approach and extracted a vector of metadata for each bag (user) that was not related to any specific instance (tweet) [52]. The Classification Features section explains how we applied feature engineering techniques to generate features that characterize the users themselves and not just their tweets.

To assess the effectiveness of using this collective approach, we compared the results of 5 standard classification algorithms in both user- and tweet-level classification, as explained in detail in the Classification Models section.

Classification Features

Overview

Rao et al [38] and Pennacchiotti and Popescu [36,37] showed that Twitter users' demographics and political views could be distinguished by considering 3 types of user classification features: behavioral features (features extracted from the user's activity on Twitter), linguistic features (features extracted from the content of the user's tweets), and social structure features (features describing the user's social network). We followed their work and adapted these types to our different domains of distinguishing patients with IBD from others who talk about the disease. We also integrated MI learning into our classification setting, which was not part of their research. We constructed a set of classification features for each feature type, as explained in detail in the following sections and summarized in Table 1.

Table 1. Summary of classification features and their types.

User classification feature, feature level, and features	Type
Behavioral features	
Tweet-level features	
Tweet counter	Integer
Retweet counter	Integer
Retweet to tweet ratio	Float (0 to 1)
IBD[a] flag	Binary
User-level IBD ratio	Float (0 to 1)
Crohn flag	Binary
User-level Crohn ratio	Float (0 to 1)
Colitis flag	Binary
User-level colitis ratio	Float (0 to 1)
User-level features	
Tweet counter	Integer
Retweet counter	Integer
Retweet to tweet ratio	Float (0 to 1)
IBD counter	Integer
Bio-IBD flag	Binary
IBD ratio	Float (0 to 1)
Crohn counter	Integer
Bio-Crohn flag	Binary
Crohn ratio	Float (0 to 1)
Colitis counter	Integer
Bio-colitis flag	Binary
Colitis ratio	Float (0 to 1)
Linguistic features	
Tweet-level features	
Emoji counter	Integer
Interjection counter	Integer
Profanity counter	Integer
Mention counter	Integer
Hashtag counter	Integer
URL flag	Binary
First-person flag	Binary
Number of words	Integer
Number of characters	Integer
Polarity	Float (−1 to 1)
Positive polarity flag (1 if polarity >0, else 0)	Binary
Negative polarity flag (1 if polarity <0, else 0)	Binary
Subjectivity	Float (0 to 1)
LDA[b] topic distribution (document=tweet)	20×float (0 to 1)
User-level features	

User classification feature, feature level, and features	Type
Emoji sum	Integer
Emoji average	Float
Bio-emoji counter	Integer
Interjection sum	Integer
Interjection average	Float
Bio-interjection counter	Integer
Profanity sum	Integer
Profanity average	Float
Bio-profanity counter	Integer
Mention sum	Integer
Mention average	Float
Bio-mention counter	Integer
Hashtag sum	Integer
Hashtag average	Float
Bio-hashtag counter	Integer
URL sum	Integer
URL average	Float (0 to 1)
Bio-URL flag	Binary
First-person sum	Integer
First-person average	Float (0 to 1)
Bio–first-person flag	Binary
Word average	Float
Bio-number of words	Integer
Character average	Float
Bio-number of characters	Integer
Bio-polarity	Float (−1 to 1)
Positive polarity sum	Integer
Positive polarity average	Float (0 to 1)
Negative polarity sum	Integer
Negative polarity average	Float (0 to 1)
Subjectivity average	Float (0 to 1)
Bio-subjectivity	Float (0 to 1)
LDA topic distribution (document=all the user’s tweets)	20×float (0 to 1)
Social structure features	
Tweet-level features	
User-level log in-degree	Float
User-level log out-degree	Float
User-level closeness	Float (0 to 1)
User-level features	
Log in-degree	Float
Log out-degree	Float

User classification feature, feature level, and features	Type
Closeness	Float (0 to 1)

[a] IBD: inflammatory bowel disease.

[b] LDA: latent Dirichlet allocation.

Behavioral Features

Features of this type were designed to capture users' activity on Twitter: How often do they tweet? Do they write new content or mainly RT others? Furthermore, how often do they refer to IBD? We counted the number of tweets and RTs in our data set and calculated the RT ratio for each user. We counted the number of times they used one of our keywords in their tweets to account for the frequency with which they addressed IBD. Aggregated features for user-level classification were also copied to all the users' tweets to enrich the tweet-level classification.

Linguistic Features

The second class of features is derived from the users' linguistic style on Twitter: Do they write in first-person voice? Do they tend to use emoticons or add a reference to an external source via URL? We used 2 types of linguistic features. On the basis of previous research [36-38] and our data's nature, we extracted several features from the text that we believed would help the classification.

Acknowledging that individuals and organizations communicate differently on Twitter [35,39], we searched for specific characteristics that could distinguish private persons from businesses and help identify patients. We checked specific characteristics for each tweet in our data: Was there use of emojis, interjections, or profanities? Was it written in the first person? Did it point to an external source via URL? Did it contain Twitter special characters indicating mentions (@) or hashtags (#)? We used a Python (Python Software Foundation) library called *TextBlob* to add sentiment-related features such as the text's polarity and subjectivity. The length of the tweets and the number of words they contained were also considered. The Python library *emoji* was used to detect emojis within the text. A part-of-speech identifier from the library *nltk* was used to indicate the use of first person and identify interjections. On the basis of the Python library *profanity*, we established a list of swear words that we searched for in the text. We had to adjust the list to the special domain of IBD as words related to metabolism were not necessarily swear words.

We started with tweet-level features, which were later grouped by user to represent personal writing style. To reflect the way a user expresses themselves on Twitter, we excluded RTs from the aggregation. The number of tweets in which the URL was used, for example, was counted on the original tweets only. As the users' biographies were considered as tweets in the tweet-level classifiers, we added the linguistic features that were extracted from the biographies as bio-features in the user-level classifiers.

In natural language processing, there are several methods to obtain a vector representation of text. One of the more well-known and well-researched techniques is the Bayesian probabilistic model of text documents called latent Dirichlet allocation (LDA). LDA is a topic modeling technique used for discovering the abstract *topics* that occur in a collection of documents [53].

We used LDA to represent text in both tweet- and user-level classification features. In tweet-level features, each tweet was considered a document, and the representations were obtained per tweet. For user-level features, all tweets by the same author were consolidated into 1 document to obtain representations per user. All the features used unigram and bigram representations of the text after data cleaning. The text cleaning process included converting to lower case, removing punctuation and stop words, and normalizing links and other special signs to standard representations.

Social Structure Features

The last type of feature we addressed represented the users' social connections on Twitter. We used the Twitter API to collect each user's followers and followees. For each user, we kept the number of followers they had (out-degree in the sense of influence) and the number of followees they had (in-degree) and scaled the results using a logarithmic scale. We also computed the closeness centrality measure for each user. Aggregated features for user-level classification were also copied to all the users' tweets to enrich the tweet-level classification.

Classification Models

Aiming to distinguish between patients with IBD and other users who tweet about IBD, we compared the performances of several classification algorithms within 2 classification approaches: the SI learning approach, which starts by classifying tweets separately and then deduces the user's class from their tweet-level classification, and the metadata-based MI learning approach, which starts by aggregating tweet-level features to user-level features and then classifies the users themselves.

The metadata-based MI approach starts by transforming the data from MI to SI, and then a standard SI algorithm can be applied to the transformed problem [54,55]. To achieve the users' characterization for the MI approach, we applied arithmetic sum and average to the tweet-level features and obtained aggregated features per user (refer to the Classification Features section for more details). Note that this process may cause some information loss [56].

For both approaches, we tested 5 standard and well-known algorithms for binary classification tasks such as ours: AdaBoost, gradient boosting classifier, linear support vector machine, logistic regression, and random forest. All the algorithms were applied from the scikit-learn (sklearn) package in Python [57].

Experiment

We split our data set by users into training and test sets (approximately 80%-20%). The training set had 155 patients and 377 nonpatients, and the test set had 39 patients and 90 nonpatients; thus, the sets maintained the ratio between the groups.

In the tweet-level classification, the split into training and test sets was performed based on the split of the users—tweets by users belonging to the training set were ascribed to the tweet training set, whereas tweets by users belonging to the test set were ascribed to the tweet test set. As a result, the tweet training set contained 263 positive tweets and 1586 negative tweets, whereas the test set contained 62 positive tweets and 293 negative tweets.

We started with a hyperparameter optimization for all algorithms using a 5-fold cross-validation over the training data in both approaches. The values tested for each algorithm and parameter can be found in Multimedia Appendix 1.

In total, 4 common metrics were used to evaluate the models: precision, recall, F_1 score, and the area under the receiver operating characteristic curve (ROC AUC). All 4 metrics were calculated over the positive class that was of interest to us. In our setting, precision depicts the probability that a positive prediction is indeed a patient, recall depicts the classifier's ability to retrieve patients, and the F_1 score combines the 2. ROC AUC considers the recall of both classes and measures the ability of the model to retrieve patients without collecting a lot of unwanted other users.

To select the best algorithm variant, we used a 10-fold cross-validation technique for a reliable evaluation of the prediction power. In this process, we randomly divided the training set into 10 equal-sized parts; then, we iteratively performed the training on 9 parts and evaluated the model on the part that was left out. We repeated this iteration 10 times, leaving out a different part each time. In addition, we repeated the 10-fold cross-validation process 10 times with different seed initializations to vary the random split. The performance metrics were computed each time, and the results presented in the Results section show the average across these 100 iterations.

In the user-level classification, we obtained all 4 metrics during the classification process using the sklearn package in Python. However, in the tweet-level classification, another aggregation stage was needed before obtaining the metrics directly from the sklearn package—the process returned the predictions for each tweet (whether it was written by a patient), and we had to infer the users' predictions by aggregating the predictions given to their tweets. As in the manual annotation process, if all the user's tweets received a prediction of 0, the user was considered a nonpatient and received a negative prediction. Alternatively, if the user had at least one positive prediction, they were considered a patient and received a positive prediction. We then used the sklearn package to compute the user-level metrics based on the users' predictions that we obtained and their true labels.

Finally, we trained the models from each approach (MI and SI) on the entire training set and evaluated their predictions on the test set. We used built-in sklearn methods for feature importance to investigate the contribution of each feature to both logistic regression and random forest algorithms. The absolute value of the coefficient represents the feature importance for logistic regression.

Analyzing Patients' Tweets

A Corpus of Lifestyle-Related Tweets

The next aim of this study was to obtain a collection of tweets in which patients describe the lifestyle-related treatments they have tried and their symptoms. By filtering and merging different web-based databases [58,59], we established a list of 420 words that are types of food or physical activities (ie, lifestyle-related words; the full list can be found in Multimedia Appendix 2). The Twitter Premium API was used to search for all tweets that mentioned IBD (containing at least one of the 3 keywords described in the Data Collection and Preparation section: *crohn*, *colitis*, and *#IBD*) and at least one of the 420 lifestyle-related words. To build the search query, we used the OR operator within the IBD keywords and the lifestyle-related words and then connected the 2 groups using the AND operator.

We searched for relevant tweets from January 1, 2019, to September 30, 2019. We excluded RTs and duplicated tweets from the search and limited the search to tweets written in English. The search resulted in 20,136 unique tweets containing new content written by 8519 different users.

We used the classifier from the first part of the study on the new data we gathered to classify the tweets as patients' tweets and user tweets. We needed to recreate the classification features for the new set of 8519 users. As we did in the first stage, we collected another week of tweets for all the users from October 1, 2019, to October 7, 2019, without keyword filtering and including RTs. A total of 39.52% (3367/8519) of the users were private, suspended, or otherwise unavailable. The process resulted in a data set of 5152 users who authored 402,843 tweets overall.

We constructed all the classification features described in the Classification Features section on the new data except for the closeness centrality. Obtaining this feature was costly and time-consuming as it was the only feature that required collecting all followers and followees for each user and building their Twitter network. As it was not one of the 10 most helpful classification features, we decided to omit it.

We then used the MI random forest model we trained in the first stage (refer to the Classification Models section for more details) to classify the users and identify patients. A total of 45.79% (2359/5152) of the users were classified as patients, and they authored 4160 of the original tweets containing our keywords. We performed a simple text cleaning of those tweets by removing all screen names (identified by the @ character) and URLs and continued our analysis with the 4160 clean tweets.

Sentiment Analysis of Lifestyle-Related Words

The Natural Language Understanding (NLU) module by IBM Cloud [60] was used to apply category classification and keyword extraction to each of our tweets. The category classification feature aims to identify the theme of the text. Given a text, the NLU module provides a list of possible categories and subcategories and their corresponding likelihoods. The keyword extraction feature recognizes words and phrases of high importance within the text and calculates their sentiments. Given a text, the NLU module returns a list of keywords and their corresponding sentiments represented as scores on the closed interval of −1 to 1: −1 for extremely negative sentiment and 1 for extremely positive sentiment. A score of 0 means that the keyword was mentioned in a neutral context. The *TextBlob* library used for sentiment analysis in the Linguistic Features section only enables full-text sentiment analysis and does not support entity-level sentiment analysis. Although it was free and easy to use, it did not suit our new task and, therefore, we chose to replace it with the NLU module.

The goal was to identify the lifestyle-related treatments that patients undergo to manage their disease and determine their sentiments toward them. Hence, we focused our analysis on keywords related to health and nutrition. We grouped all tweets that were categorized by the NLU module as related to *health and fitness* (2080 tweets), *food and drink* (1568 tweets), or *religion and spirituality* (15 tweets). Overall, 3663 tweets were selected for keyword sentiment analysis. We gathered all the keywords that appeared in our predefined list of lifestyle-related words and their corresponding sentiments within each tweet. In total, 3 examples of this process are presented in Table 2. Notice how, in the second example, the first word of the original tweet (marked with the @ symbol) is a screen name and was therefore removed in the cleaning process.

Table 2. Three examples of category classification and keyword sentiment extraction after text cleaning.

Number	Original text	Text after cleaning	Category classification	Keyword sentiment
1	Spinach is an inflammatory food with a lot of sulfur. Ban that too. (I noticed my Crohn's tended to flare around spinach season.)	Spinach is an inflammatory food with a lot of sulfur. Ban that too. (I noticed my Crohn's tended to flare around spinach season.)	Food and drink	Spinach: −0.63
2	@bottomline_ibd great poll. I do have the odd binge, but IBD has changed what I can drink. No more red wine or ale	great poll. I do have the odd binge, but IBD has changed what I can drink. No more red wine or ale	Food and drink	Red wine: −0.83; ale: −0.83
3	I am living proof that yoga can help #uchicagoibd #studiothree #yoga #ibd	I am living proof that yoga can help #uchicagoibd #studiothree #yoga #ibd	Religion and spirituality	Yoga: 0.69

To examine the effectiveness of each lifestyle-related phrase (lifestyle, in short) and to assess its overall sentiment, we aggregated the results by lifestyle and calculated the following statistics: the total number of times the lifestyle appeared in all tweets, the number of times it appeared in a positive (or negative) context, the positive to negative ratio of the number of appearances (odds), and the mean sentiment of the lifestyle.

We used the statistics to build a co-occurrence network that visualized the connections between lifestyles and their mean sentiments. The different lifestyles were the nodes, and an arc connected 2 lifestyles if they appeared in the same tweet. The more times they appeared together, the stronger the connection between the lifestyles was. Therefore, the resulting network was undirected and weighted by the number of times the lifestyles co-occurred. The purpose was to identify helpful lifestyles (frequently mentioned in a positive context) and lifestyles that it is better to avoid (frequently mentioned in a negative context) and examine whether certain lifestyles tend to be implemented together.

The network was obtained using *Gephi* software (GNU General Public License) for network analysis and visualization. Each node was colored on a scale from green to red based on the mean sentiment of the lifestyle it represented, with green being very positive and red being very negative. The sizing of the nodes reflected the number of times the lifestyles were mentioned in the tweet database: the more times they appeared, the larger their nodes were. The thickness of each arc represented the number of times the 2 lifestyles it connected co-occurred: the thicker the arc, the more times the 2 lifestyles appeared together. To avoid obtaining an overdense network, we only considered the nodes of lifestyles mentioned at least five times in our database. We included arcs between lifestyles that co-occurred at least four times. The process resulted in 144 lifestyles presented in the network and sorted in a table by mean sentiment.

Ethical Note

The collection and analysis of Twitter data may entail ethical challenges that should be addressed and handled properly. Twitter data are public and available for research via Twitter APIs. By accepting Twitter's Terms of Service and Privacy Policy, Twitter users acknowledge that their tweets can be viewed instantly worldwide and that their information may be collected by third parties [61]. Nonetheless, social media studies have revealed that users on Twitter feel as if they are engaged in a private conversation with their followees and followers [62,63]. Although they are generally not concerned with their posts being used for research purposes, they expect anonymity in publication and to be asked for their consent before publication.

Obtaining informed consent from all the users who *participate* in research on Twitter data may be unfeasible. Data sets are likely to be large and involve many authors [61-63]. Individually seeking consent from all 722 users in our study would be labor-intensive or impossible as some might be unreachable.

Moreover, providing total anonymity to users while directly quoting their content is not practical; tweets are easily searchable, leaving their authors vulnerable to identification.

To adhere to ethical norms and maintain user privacy, we only published aggregated results that do not reveal the specific users. The 3 examples containing direct quotes from tweets (in Table 2) are presented in this study after obtaining informed consent from their authors.

Results

Patient Identification

Table 3 shows the 10-fold cross-validation and test results for the 2 classification approaches: SI classifying tweets and MI classifying users. The table shows the results of the 4 metrics for all 5 classification algorithms.

Table 3. The 10-fold cross-validation and test results for the single instance (SI) and multiple instance (MI) classifications.

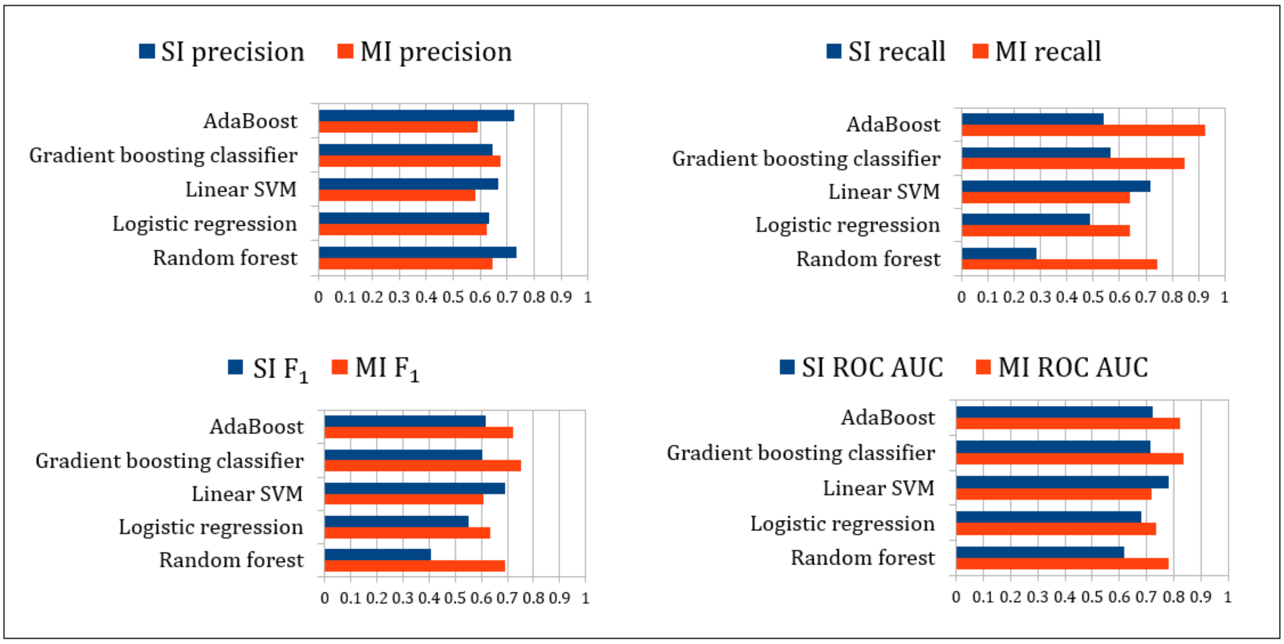
Algorithm and metric	SI tweet-level classification		MI user-level classification	
	10-fold	Test	10-fold	Test
AdaBoost				
Precision	0.6775	0.7241	0.6151	0.5902
Recall	0.6297	0.5385	0.7284	0.9231
F_1 score	0.6525	0.6176	0.6542	0.7200
ROC AUC[a]	0.7532	0.7248	0.8469	0.8226
Gradient boosting classifier				
Precision	0.7416	0.6471	0.6668	0.6735
Recall	0.6465	0.5641	0.6778	0.8462
F_1 score	0.6906	0.6027	0.6711	0.7500
ROC AUC	0.7768	0.7154	0.8658	0.8342
Linear SVM[b]				
Precision	0.7249	0.6667	0.6648	0.5814
Recall	0.6832	0.7179	0.6398	0.6410
F_1 score	0.7034	0.6914	0.6472	0.6098
ROC AUC	0.7883	0.7812	0.8463	0.7205
Logistic regression				
Precision	0.7405	0.6333	0.6594	0.6250
Recall	0.6335	0.4872	0.6358	0.6410
F_1 score	0.6829	0.5507	0.6423	0.6329
ROC AUC	0.7712	0.6825	0.8473	0.7372
Random forest				
Precision	0.7676	0.7333	0.6721	0.6444
Recall	0.4355	0.2821	0.6646	0.7436
F_1 score	0.5555	0.4074	0.6595	0.6905
ROC AUC	0.6906	0.6188	0.8722	0.7829

[a] ROC AUC: area under the receiver operating characteristic curve.

[b] SVM: support vector machine.

Both approaches showed satisfactory classification results for the patient classes. Although the precision rates were slightly higher for the SI approach, the recall index of the MI approach was better, and the results for the ROC AUC measure were consistently higher in the MI approach. Figure 3 shows the differences among the 4 measures within the test set results.

Figure 3. Test result comparison between the 2 classification approaches. MI: multiple instance; ROC AUC: area under the receiver operating characteristic curve; SI: single instance; SVM: support vector machine.



Investigating the contribution of each feature to both the logistic regression and random forest algorithms showed the importance of the use of first-person voice. In both classification approaches and algorithms, the most important feature was the use of the first person, which had a significant advantage over the other features. The first-person flag was the best feature of the SI approach, and its average was the best feature of the MI approach. Another dominant feature was the use of profanities as it was one of the most significant features in both approaches and algorithms.

The analysis also highlighted the importance of the LDA features derived from the text. The second-best feature of the SI approach was LDA topic 11 for both the logistic regression and random forest algorithms. This was the only topic that did not contain IBD-related words. The fourth and fifth most important topics of the MI approach were identical for both algorithms—LDA topics 17 and 9, respectively. The LDA topics that were created over the training data for each approach can be found in Multimedia Appendix 3.

Analyzing Patients' Tweets

In the second stage of the study, a network of connections between lifestyles was built and visualized. The obtained network describing the relationships between the different lifestyles can be found in Multimedia Appendix 4.

The most frequent word in our database was *diet*, encapsulating all the nutritional adjustments that patients undergo to manage their disease. Specific diets such as *paleo*, *vegetarian*, or *liquid* diets also surfaced and in a negative context.

It is interesting to note that the negative and positive lifestyles revealed by the analysis were in line with what is known about suitable nutrition for IBD. Among the most negative lifestyles (by mean sentiment), we found *alcohol*, *milk*, *spicy*, *cabbage*, *flour*, *lentil*, and *orange juice*, all known to cause inflammation and irritate the stomach. Among the most positive lifestyles (by mean sentiment), we found activity-related lifestyles such as *fitness* or *yoga* and healing foods such as *salmon*, *ginger*, and *garlic*. The most positive lifestyle turned out to be *sushi*, which usually contains anti-inflammatory ingredients such as *salmon* or *tuna*, *seaweed*, and *rice*. Table 4 presents the 20 most positive and 20 most negative lifestyle-related words sorted by mean sentiment.

Table 4. The 20 most positive and 20 most negative lifestyles sorted by mean sentiment.

Rank	Keyword	Count	Sentiment, mean (SD)	Count of positive	Count of negative	Odds
1	Sushi	9	0.466 (0.814)	7	2	3.500
2	Ginger ale	5	0.407 (0.597)	3	1	3.000
3	Salmon	7	0.344 (0.691)	4	3	1.333
4	Cherry	10	0.33 (0.696)	6	2	3.000
5	Breakfast	29	0.28 (0.75)	19	9	2.111
6	Garlic	8	0.244 (0.671)	4	2	2.000
7	Bagel	5	0.224 (0.633)	3	1	3.000
8	Almond	9	0.193 (0.668)	6	3	2.000
9	Yogurt	14	0.189 (0.688)	7	3	2.333
10	Yoga	15	0.186 (0.693)	7	5	1.400
11	Ham	5	0.184 (0.535)	2	1	2.000
12	Biscuit	13	0.172 (0.75)	8	5	1.600
13	Spinach	6	0.171 (0.76)	4	2	2.000
14	Vegan cheese	5	0.164 (0.92)	3	2	1.500
15	Lamb	5	0.14 (0.861)	3	2	1.500
16	Cake	26	0.13 (0.752)	16	9	1.778
17	Fitness	19	0.114 (0.728)	9	6	1.500
18	Ginger	17	0.112 (0.724)	8	7	1.143
19	Tomato	10	0.089 (0.608)	5	3	1.667
20	Cafe	7	0.081 (0.783)	3	3	1.000
125	Fodmap	12	−0.501 (0.573)	2	9	0.222
126	Cocktail	5	−0.51 (0.769)	1	4	0.250
127	Fiber	63	−0.512 (0.547)	7	47	0.149
128	Spicy	37	−0.514 (0.572)	7	28	0.250
129	Vegetable	49	−0.533 (0.529)	6	39	0.154
130	Corn	28	−0.534 (0.487)	2	22	0.091
131	Alcohol	64	−0.545 (0.545)	9	51	0.176
132	Milkshake	5	−0.556 (0.811)	1	4	0.250
133	Milk	44	−0.565 (0.5)	4	35	0.114
134	Vegetarian diet	10	−0.567 (0.409)	1	8	0.125
135	Snack	10	−0.573 (0.568)	2	8	0.250
136	Fig	5	−0.578 (0.621)	1	4	0.250
137	Turkey	10	−0.608 (0.626)	2	8	0.250
138	Yeast	16	−0.624 (0.391)	1	13	0.077
139	Orange	7	−0.638 (0.449)	0	5	0.000
140	Beverage	7	−0.661 (0.616)	1	6	0.167
141	Cabbage	8	−0.675 (0.19)	0	8	0.000
142	Orange juice	5	−0.682 (0.385)	0	4	0.000
143	Flour	6	−0.785 (0.211)	0	6	0.000
144	Lentil	6	−0.785 (0.188)	0	6	0.000

Discussion

Principal Findings

This study presents a workflow for identifying patients with IBD on Twitter and exploring their tweets. The aim was to identify patients with IBD based on the way they communicate on Twitter and to learn from the personal experiences they share.

In the first stage of the study, a classifier of Twitter users designed to distinguish patients with IBD from other users was constructed and evaluated. Classification features combining social data and text analysis were extracted from the users' activity on Twitter, their social connections, and the content of their tweets. Various classification algorithms were considered, and 4 evaluation measures were calculated for each of them. The encouraging results shown in the previous section helped convince us that patients with IBD can be identified on Twitter based on such features.

Classification results from both the SI and MI approaches show that patients with IBD differ in the way they communicate on Twitter from other users who tweet about the disease. They talk in the first person more often and use more profanities in their tweets. These gaps, which can be explained by the fact that patients are private individuals whereas nonpatients also include organizations and voluntary associations that communicate in a much more formal manner, helped distinguish patients from other entities in the different classification models we tried in this study.

Our analysis differs from previous research regarding user classification on Twitter [36-38] in 2 aspects. Conceptually, we investigate a different domain and try to identify patients on Twitter. Practically, we compare the results from the user-level classification with a tweet-level classification.

In the second stage of the study, tweets of patients with IBD were collected to investigate the different lifestyles they implemented to deal with their disease and assess these lifestyles' effectiveness. Unlike previous research on patients' sentiments on Twitter [48-50], we focused our research on entity sentiment for specific words rather than the entire tweet's sentiment. We suggested a novel approach by considering entity sentiment analysis to obtain patients' sentiments toward the different nutrition and fitness-based solutions they try. These findings were in line with what is known about IBD as several foods known to cause inflammation were pointed out in a negative sentiment, whereas relaxing activities and anti-inflammatory foods surfaced in a positive context.

This study suggests that there is room for collaboration between physicians and engineers regarding understanding chronic diseases. Owing to the chronic nature of the disease and the fact that it involves bowel movements, patients with IBD are compelled to follow special nutrition and maintain a calm routine. By collecting and analyzing patients' personal experiences on social media, we can monitor patients' lifestyles and support medical knowledge of IBD. We can identify and assess complementary treatments to diets and physical activity and maybe ease patients' processes of finding the right treatments for them. Although such analysis should not strive to replace physicians or draw conclusions of a clinical nature, it may provide complementary recommendations for healthy lifestyles based on the wisdom of the crowd.

Limitations and Future Work

Overview

The focus of this study was on showing the potential of identifying patients with IBD on Twitter and learning from their tweets. This study emphasized the entire process, and we did not perfect each part separately. As this section explains, each part can be improved by trying different methods and enriching the analysis.

Patient Identification

The classifier developed in the first stage of this study uses 1-level, binary classification to separate patients with IBD from other users who tweet about the disease. Some of its features distinguish organizations from individuals in general and do not necessarily detect patients, such as the use of the first person in the tweet. Therefore, our nonpatient class is heterogeneous and somewhat ambiguous, containing both organizations that significantly differ from patients in their communication patterns and healthy individuals who differ from patients in a more refined manner. Even during the manual labeling process, all 14 users excluded from the data set owing to classification disagreements were individuals talking in the first-person voice.

A possible direction for future work would be to try a 2-step classification: separating persons from organizations and continuing by searching for patients among these individuals. It can improve the robustness of some of the features by overcoming the heterogeneity of the nonpatient class in our model. Alternatively, we could try replacing the binary classification with a multinomial one that will capture not only organizations and patients but also individuals who talk about the disease and maybe mention other patients but are not sick themselves.

During the construction of the network-based features, we only collected immediate connections on Twitter (ie, the followers and followees of each patient). The sampling method resulted in basic network features that mainly included degree measures. We encourage future research to consider more interesting network features such as other centrality measures or structures. Such enhancement will require collecting at least one more level of connections (eg, followees of followees) to understand network patterns better.

Finally, the classifier uses standard classification algorithms and did not try current state-of-the-art learning techniques based on neural networks. Text representation using word embeddings, where words are mapped to vectors of real numbers in a predefined vector space [64,65], is also worth examining.

Analyzing Patients' Tweets

The NLU module by IBM Cloud was used in this study for entity sentiment analysis as a proof of concept. We did not evaluate its results or compare them with similar tools available in the market, such as the Natural Language API by Google Cloud. Future research should consider performing similar analyses with different natural language processing tools and comparing their results. Even training designated algorithms on data from lifestyle-related tweets such as those used in this study can benefit the analysis.

Overall, the results for the second part are preliminary, and much more can be done to understand what patients with IBD are talking about on Twitter. For example, by characterizing treatment options and patients' sentiments toward them, one can derive recommendations for a healthy lifestyle based on the wisdom of the crowd. Thoroughly exploring outliers, such as the 4 positive mentions of milk as opposed to the 35 negative ones, can reveal new information regarding the disease that has not yet been covered in the literature.

Conclusions

In the era of personalized medicine and patient-centered care, it is important to derive insights that reflect the patients' perspectives as manifested in social media. Although the time between physician appointments can be lengthy, messages on social media are being posted each day, and patients constantly use them to exchange inputs and recommendations.

This study provides a potential pipeline for identifying patients with chronic illnesses on Twitter and collecting their tweets to analyze the experimental knowledge they share on the web. The method presented in this study was applied to IBD and can also help explore other medical conditions. The classifier for IBD-related entities can be adapted to identify other patients with chronic illnesses. The analysis of patients' tweets can benefit research on other chronic conditions with similar characteristics. With conditions such as celiac disease or diabetes, which involve strict dietary guidelines, one can better understand patients' difficulties with adherence to their new lifestyles. When considering diseases that cause embarrassment, such as HIV, one can learn more about the constant struggle of patients living with the disease.

Therefore, the contribution of this study is 2-fold: it provides an analytical contribution to the fields of text mining and social media and a practical contribution by better understanding chronic conditions and promoting a healthy lifestyle for patients with chronic illnesses.

Acknowledgments

This study was supported by a grant from the European Research Area Network Cofund Healthy Diet for a Healthy Life-Intestinal Microbiomics under the Joint Programming Initiative *A healthy diet for a healthy life* umbrella.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Parameter optimization for classification algorithms.
[DOCX File , 20 KB - jmir_v24i8e29186_app1.docx]

Multimedia Appendix 2
A list of 420 lifestyle-related words.
[DOCX File , 39 KB - jmir_v24i8e29186_app2.docx]

Multimedia Appendix 3
Latent Dirichlet allocation topics created over the training data for each classification approach.
[DOCX File , 20 KB - jmir_v24i8e29186_app3.docx]

Multimedia Appendix 4
A network of relationships between lifestyle-related words.
[PDF File (Adobe PDF File), 148 KB - jmir_v24i8e29186_app4.pdf]

References

1. Yin Z, Fabbri D, Rosenbloom ST, Malin B. A scalable framework to detect personal health mentions on Twitter. J Med Internet Res 2015 Jun 05;17(6):e138 [FREE Full text] [doi: 10.2196/jmir.4305] [Medline: 26048075]
2. Karisani P, Agichtein E. Did you really just have a heart attack? Towards robust detection of personal health mentions in social media. In: Proceedings of the 2018 World Wide Web Conference. 2018 Presented at: WWW '18; April 23-27, 2018; Lyon, France p. 137-146. [doi: 10.1145/3178876.3186055]
3. 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]

4. Jahanbin K, Rahmanian V. Using Twitter and Web news mining to predict COVID-19 outbreak. Asian Pac J Trop Med 2020;13(8):378-380. [doi: 10.4103/1995-7645.279651]
5. Lopreite M, Panzarasa P, Puliga M, Riccaboni M. Early warnings of COVID-19 outbreaks across Europe from social media. Sci Rep 2021 Jan 25;11(1):2147 [FREE Full text] [doi: 10.1038/s41598-021-81333-1] [Medline: 33495534]
6. Gabarron E, Dorronzoro E, Rivera-Romero O, Wynn R. Diabetes on Twitter: a sentiment analysis. J Diabetes Sci Technol 2019 May;13(3):439-444 [FREE Full text] [doi: 10.1177/1932296818811679] [Medline: 30453762]
7. Cooper A, Kar P. A new dawn: the role of social media in diabetes education. J Diabetes Nurs 2014 Jan;18(2):68-71.
8. Beguerisse-Díaz M, McLennan AK, Garduño-Hernández G, Barahona M, Ulijaszek SJ. The 'who' and 'what' of #diabetes on Twitter. Digit Health 2017 Jan 1;3:2055207616688841 [FREE Full text] [doi: 10.1177/2055207616688841] [Medline: 29942579]
9. Sugawara Y, Narimatsu H, Hozawa A, Shao L, Otani K, Fukao A. Cancer patients on Twitter: a novel patient community on social media. BMC Res Notes 2012 Dec 27;5:699 [FREE Full text] [doi: 10.1186/1756-0500-5-699] [Medline: 23270426]
10. Tsuya A, Sugawara Y, Tanaka A, Narimatsu H. Do cancer patients tweet? Examining the twitter use of cancer patients in Japan. J Med Internet Res 2014 May 27;16(5):e137 [FREE Full text] [doi: 10.2196/jmir.3298] [Medline: 24867458]
11. Kaplan GG. The global burden of IBD: from 2015 to 2025. Nat Rev Gastroenterol Hepatol 2015 Dec;12(12):720-727. [doi: 10.1038/nrgastro.2015.150] [Medline: 26323879]
12. Loftus Jr EV. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. Gastroenterology 2004 May;126(6):1504-1517. [doi: 10.1053/j.gastro.2004.01.063] [Medline: 15168363]
13. Roccetti M, Marfia G, Salomoni P, Prandi C, Zagari RM, Gningaye Kengni FL, et al. Attitudes of Crohn's disease patients: infodemiology case study and sentiment analysis of Facebook and Twitter posts. JMIR Public Health Surveill 2017 Aug 09;3(3):e51 [FREE Full text] [doi: 10.2196/publichealth.7004] [Medline: 28793981]
14. Trivedi I, Keefer L. The emerging adult with inflammatory bowel disease: challenges and recommendations for the adult gastroenterologist. Gastroenterol Res Pract 2015;2015:260807 [FREE Full text] [doi: 10.1155/2015/260807] [Medline: 26064089]
15. Norton BA, Thomas T, 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]
16. Rubin DT, Dubinsky MC, Panaccione R, Siegel CA, Binion DG, Kane SV, et al. The impact of ulcerative colitis on patients' lives compared to other chronic diseases: a patient survey. Dig Dis Sci 2010 Apr;55(4):1044-1052. [doi: 10.1007/s10620-009-0953-7] [Medline: 20155319]
17. Brydolf M, Segesten K. Living with ulcerative colitis: experiences of adolescents and young adults. J Adv Nurs 1996 Jan;23(1):39-47. [doi: 10.1111/j.1365-2648.1996.tb03133.x] [Medline: 8708222]
18. Devlen J, Beusterien K, Yen L, Ahmed A, Cheifetz AS, Moss AC. The burden of inflammatory bowel disease: a patient-reported qualitative analysis and development of a conceptual model. Inflamm Bowel Dis 2014 Mar;20(3):545-552 [FREE Full text] [doi: 10.1097/01.MIB.0000440983.86659.81] [Medline: 24407484]
19. Hall NJ, Rubin GP, Dougall A, Hungin AP, Neely J. The fight for 'health-related normality': a qualitative study of the experiences of individuals living with established inflammatory bowel disease (ibd). J Health Psychol 2005 May;10(3):443-455. [doi: 10.1177/1359105305051433] [Medline: 15857873]
20. Frohlich DO. The social construction of inflammatory bowel disease using social media technologies. Health Commun 2016 Nov;31(11):1412-1420. [doi: 10.1080/10410236.2015.1077690] [Medline: 27050670]
21. Kemp K, Griffiths J, Lovell K. Understanding the health and social care needs of people living with IBD: a meta-synthesis of the evidence. World J Gastroenterol 2012 Nov 21;18(43):6240-6249 [FREE Full text] [doi: 10.3748/wjg.v18.i43.6240] [Medline: 23180944]
22. Becker KL. Cyberhugs: creating a voice for chronic pain sufferers through technology. Cyberpsychol Behav Soc Netw 2013 Feb;16(2):123-126. [doi: 10.1089/cyber.2012.0361] [Medline: 23276258]
23. Wiese J, Kelley PG, Cranor LF, Dabbish L, Hong JI, Zimmerman J. Are you close with me? Are you nearby?: investigating social groups, closeness, and willingness to share. In: Proceedings of the 13th International Conference on Ubiquitous Computing. 2011 Presented at: UbiComp '11; September 17-21, 2011; Beijing, China p. 197-206. [doi: 10.1145/2030112.2030140]
24. Paek HJ, Hove T, Ju Jeong H, Kim M. Peer or expert? The persuasive impact of YouTube public service announcement producers. Int J Advert 2015 Jan 07;30(1):161-188. [doi: 10.2501/ija-30-1-161-188]
25. Lin WY, Zhang X, Song H, Omori K. Health information seeking in the Web 2.0 age: trust in social media, uncertainty reduction, and self-disclosure. Comput Human Behav 2016 Mar;56:289-294. [doi: 10.1016/j.chb.2015.11.055]
26. Kimberly C. Data mining Twitter for cancer, diabetes, and asthma insights. Purdue University. 2016. URL: https://docs.lib.purdue.edu/dissertations/AAI10170604/ [accessed 2018-01-24]
27. Heaivilin N, Gerbert B, Page J, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
28. Harris JK, Mueller NL, Snider D, Haire-Joshu D. Local health department use of Twitter to disseminate diabetes information, United States. Prev Chronic Dis 2013 May 02;10:E70 [FREE Full text] [doi: 10.5888/pcd10.120215] [Medline: 23639765]

29. Hemsley B, Palmer S. Two studies on Twitter networks and tweet content in relation to Amyotrophic Lateral Sclerosis (ALS): conversation, information, and 'Diary of a Daily Life'. In: Digital Health Innovation for Consumers, Clinicians, Connectivity and Community: Selected Papers from the 24th Australian National Health Informatics Conference. Amsterdam, The Netherlands: IOS Press; 2016 Presented at: HIC '16; July 25-27, 2016; Melbourne, Australia p. 41-47. [doi: 10.3233/978-1-61499-666-8-41]
30. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. JMIR Public Health Surveill 2015 Jul 27;1(2):e7 [FREE Full text] [doi: 10.2196/publichealth.4488] [Medline: 27227141]
31. Sioula-Georgoulea I. Approaching Twitter sociologically: a case study of the public humiliation of HIV-positive women. Επιθεώρηση Κοινωνικών Ερευνών 2015 Nov 25;144:103-128. [doi: 10.12681/grsr.8625]
32. Odlum M, Yoon S. HIV/AIDS and the millennium development goals: a public sentiment analysis of world AIDS day Twitter chat. Int J AIDS Res 2016 Nov 14;3(9):134-137 [FREE Full text] [doi: 10.19070/2379-1586-1600026]
33. De Choudhury M, Morris MR, White RW. Seeking and sharing health information online: comparing search engines and social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 26-May 1, 2014; Toronto, Canada p. 1365-1376. [doi: 10.1145/2556288.2557214]
34. Arakawa Y, Kameda A, Aizawa A, Suzuki T. Adding Twitter-specific features to stylistic features for classifying tweets by user type and number of retweets. J Assn Inf Sci Tec 2014 Jan 22;65(7):1416-1423. [doi: 10.1002/asi.23126]
35. Holmberg K, Bowman TD, Haustein S, Peters I. Astrophysicists' conversational connections on Twitter. PLoS One 2014 Aug 25;9(8):e106086 [FREE Full text] [doi: 10.1371/journal.pone.0106086] [Medline: 25153196]
36. Pennacchiotti M, Popescu AM. Democrats, republicans and starbucks afficionados: user classification in Twitter. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2011 Presented at: KDD '11; August 21-24, 2011; San Diego, CA, USA p. 430-438. [doi: 10.1145/2020408.2020477]
37. Pennacchiotti M, Popescu AM. A machine learning approach to Twitter user classification. Proc Int AAI Conf Web Soc Media 2011;5(1):281-288.
38. Rao D, Yarowsky D, Shreevats A, Gupta M. Classifying latent user attributes in Twitter. In: Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents. 2010 Presented at: SMUC '10; October 30, 2010; Toronto, Canada p. 37-44. [doi: 10.1145/1871985.1871993]
39. De Choudhury M, Diakopoulos N, Naaman M. Unfolding the event landscape on Twitter: classification and exploration of user categories. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12; February 11-15, 2012; Seattle, WA, USA p. 241-244. [doi: 10.1145/2145204.2145242]
40. Holmberg K, Eriksson-Backa K, Ek S. Tweeting about diabetes and diets – content and conversational connections. In: Proceedings of the 5th International Conference on Well-Being in the Information Society. 2014 Presented at: WIS '14; August 18-20, 2014; Turku, Finland p. 45-56. [doi: 10.1007/978-3-319-10211-5_5]
41. Bergsma S, Dredze M, Van Durme B, Wilson T, Yarowsky D. Broadly improving user classification via communication-based name and location clustering on Twitter. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013 Presented at: HLT-NAACL '13; June 9-14, 2013; Atlanta, GA, USA p. 1010-1019.
42. Harris JK, Mart A, Moreland-Russell S, Caburnay CA. Diabetes topics associated with engagement on Twitter. Prev Chronic Dis 2015 May 07;12:E62 [FREE Full text] [doi: 10.5888/pcd12.140402] [Medline: 25950569]
43. Khan A, Silverman A, Rowe A, Rowe S, Tick M, Testa S, et al. Who is saying what about Inflammatory Bowel Disease on Twitter? In: GW Annual Research Days 2018. 2018 Presented at: GW Research '18; April 10-11, 2018; Virtual p. 176 URL: https://hsrc.himmelfarb.gwu.edu/gw_research_days/2018/SMHS/62/
44. Rowe A, Rowe S, Silverman A, Borum ML. P024 Crohn's disease messaging on Twitter: who's talking? Gastroenterology 2018 Jan 1;154(1):S13-S14. [doi: 10.1053/j.gastro.2017.11.059]
45. Roccetti M, Casari A, Marfia G. Inside chronic autoimmune disease communities: a social networks perspective to Crohn's patient behavior and medical information. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2015 Presented at: ASONAM '15; August 25-28, 2015; Paris, France p. 1089-1096. [doi: 10.1145/2808797.2808813]
46. O'Neill P, Shandro B, Poullis A. Patient perspectives on social-media-delivered telemedicine for inflammatory bowel disease. Future Healthc J 2020 Oct;7(3):241-244 [FREE Full text] [doi: 10.7861/fhj.2020-0094] [Medline: 33094237]
47. Frohlich DO, Zmyslinski-Seelig AN. How Uncover Ostomy challenges ostomy stigma, and encourages others to do the same. New Media Soc 2014 Jul 09;18(2):220-238. [doi: 10.1177/1461444814541943]
48. Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using Twitter to understand the human bowel disease community: exploratory analysis of key topics. J Med Internet Res 2019 Aug 15;21(8):e12610 [FREE Full text] [doi: 10.2196/12610] [Medline: 31411142]
49. Margolis PA, Maddali HT, Gloor PA. Comparing online community structure of patients of chronic diseases. Int J Organ Des Eng 2016;4(1/2):113-136. [doi: 10.1504/ijode.2016.10001025]
50. Cohen ER, Spiegel BM, van Oijen MG. Tu1068 Twitter offers insight into Health Related Quality of Life (HRQoL) in ulcerative colitis. Gastroenterology 2013 May;5(144):S-751. [doi: 10.1016/s0016-5085(13)62785-7]

51. Dietterich TG, Lathrop RH, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles. Artif Intell 1997 Jan;89(1-2):31-71. [doi: 10.1016/s0004-3702(96)00034-3]
52. Xu X. Statistical learning in multiple instance problems. The University of Waikato. 2003 Jun. URL: https://www.cs.waikato.ac.nz/~ml/publications/2003/xinxu_thesis.pdf [accessed 2018-07-15]
53. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022.
54. Foulds J, Frank E. A review of multi-instance learning assumptions. Knowl Eng Rev 2010 Mar 01;25(1):1-24. [doi: 10.1017/s026988890999035x]
55. Dong L. A comparison of multi-instance learning algorithms. The University of Waikato. 2006 Feb. URL: https://researchcommons.waikato.ac.nz/bitstream/handle/10289/2453/thesis.pdf?sequence=1&isAllowed=y [accessed 2018-07-15]
56. Schulte O, Routley K. Aggregating predictions vs. aggregating features for relational classification. In: Proceedings of the 2014 IEEE Symposium on Computational Intelligence and Data Mining. 2014 Presented at: CIDM '14; December 9-12, 2014; Orlando, FL, USA p. 121-128. [doi: 10.1109/cidm.2014.7008657]
57. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830.
58. Composition of foods integrated dataset (CoFID). Public Health England. 2015 Mar 25. URL: https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid#:~:text=The%20'Composition%20of%20Foods%20Integrated,recipes%20within%20the%20pork%20section [accessed 2019-09-16]
59. FoodData Central. U.S. Department of Agriculture, Agricultural Research Service. 2019. URL: https://fdc.nal.usda.gov/ [accessed 2019-09-16]
60. Natural Language Understanding. IBM Cloud API Docs. 2020. URL: https://cloud.ibm.com/apidocs/natural-language-understanding [accessed 2021-07-15]
61. Ahmed W, Bath PA, Demartini G. Using Twitter as a data source: an overview of ethical, legal, and methodological challenges. In: Woodfield K, editor. The Ethics of Online Research (Advances in Research Ethics and Integrity, Volume 2). Bingley, UK: Emerald Group Publishing; 2017:79-107.
62. Fiesler C, Proferes N. “Participant” perceptions of Twitter research ethics. Soc Media Soc 2018 Mar 10;4(1):205630511876336. [doi: 10.1177/2056305118763366]
63. Williams ML, Burnap P, Sloan L. Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation. Sociology 2017 Dec;51(6):1149-1168 [FREE Full text] [doi: 10.1177/0038038517708140] [Medline: 29276313]
64. Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. J Mach Learn Res 2003;3:1137-1155.
65. Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning. In: Proceedings of the 25th International Conference on Machine Learning. 2008 Presented at: ICML '08; July 5-9, 2008; Helsinki, Finland p. 160-167. [doi: 10.1145/1390156.1390177]

Abbreviations

API: application programming interface
IBD: inflammatory bowel disease
LDA: latent Dirichlet allocation
MI: multiple instance
NLU: Natural Language Understanding
ROC AUC: area under the receiver operating characteristic curve
RT: retweet
SI: single instance

Edited by JMIRPE Office; submitted 29.03.21; peer-reviewed by J Chen, H Wang, S Kiritchenko; comments to author 05.06.21; revised version received 29.10.21; accepted 20.05.22; published 02.08.22.

Please cite as:
Stemmer M, Parmet Y, Ravid G
Identifying Patients With Inflammatory Bowel Disease on Twitter and Learning From Their Personal Experience: Retrospective Cohort Study
J Med Internet Res 2022;24(8):e29186
URL: https://www.jmir.org/2022/8/e29186
doi:10.2196/29186
PMID:35917151

Original Paper

Detecting Potentially Harmful and Protective Suicide-Related Content on Twitter: Machine Learning Approach

Hannah Metzler[1,2,3,4,5], Mag, PhD; Hubert Baginski[3,6], BSc, MSc; Thomas Niederkrotenthaler[2], MBA, Dr med, PhD; David Garcia[1,3,4], MSc, PhD

[1]Section for the Science of Complex Systems, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

[2]Unit Suicide Research and Mental Health Promotion, Center for Public Health, Medical University of Vienna, Vienna, Austria

[3]Complexity Science Hub Vienna, Vienna, Austria

[4]Computational Social Science Lab, Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria

[5]Institute for Globally Distributed Open Research and Education, Vienna, Austria

[6]Institute of Information Systems Engineering, Vienna University of Technology, Vienna, Austria

Corresponding Author:
Hannah Metzler, Mag, PhD
Section for the Science of Complex Systems
Center for Medical Statistics, Informatics and Intelligent Systems
Medical University of Vienna
Spitalgasse 23
Vienna, 1090
Austria
Phone: 43 159991 ext 604
Email: metzler@csh.ac.at

Abstract

Background: Research has repeatedly shown that exposure to suicide-related news media content is associated with suicide rates, with some content characteristics likely having harmful and others potentially protective effects. Although good evidence exists for a few selected characteristics, systematic and large-scale investigations are lacking. Moreover, the growing importance of social media, particularly among young adults, calls for studies on the effects of the content posted on these platforms.

Objective: This study applies natural language processing and machine learning methods to classify large quantities of social media data according to characteristics identified as potentially harmful or beneficial in media effects research on suicide and prevention.

Methods: We manually labeled 3202 English tweets using a novel annotation scheme that classifies suicide-related tweets into 12 categories. Based on these categories, we trained a benchmark of machine learning models for a multiclass and a binary classification task. As models, we included a majority classifier, an approach based on word frequency (term frequency-inverse document frequency with a linear support vector machine) and 2 state-of-the-art deep learning models (Bidirectional Encoder Representations from Transformers [BERT] and XLNet). The first task classified posts into 6 main content categories, which are particularly relevant for suicide prevention based on previous evidence. These included personal stories of either suicidal ideation and attempts or coping and recovery, calls for action intending to spread either problem awareness or prevention-related information, reporting of suicide cases, and other tweets irrelevant to these 5 categories. The second classification task was binary and separated posts in the 11 categories referring to actual suicide from posts in the off-topic category, which use suicide-related terms in another meaning or context.

Results: In both tasks, the performance of the 2 deep learning models was very similar and better than that of the majority or the word frequency classifier. BERT and XLNet reached accuracy scores above 73% on average across the 6 main categories in the test set and F_1-scores between 0.69 and 0.85 for all but the suicidal ideation and attempts category (F_1=0.55). In the binary classification task, they correctly labeled around 88% of the tweets as about suicide versus off-topic, with BERT achieving F_1-scores of 0.93 and 0.74, respectively. These classification performances were similar to human performance in most cases and were comparable with state-of-the-art models on similar tasks.

Conclusions: The achieved performance scores highlight machine learning as a useful tool for media effects research on suicide. The clear advantage of BERT and XLNet suggests that there is crucial information about meaning in the context of words beyond mere word frequencies in tweets about suicide. By making data labeling more efficient, this work has enabled large-scale investigations on harmful and protective associations of social media content with suicide rates and help-seeking behavior.

(J Med Internet Res 2022;24(8):e34705) doi:10.2196/34705

KEYWORDS

suicide prevention; Twitter; social media; machine learning; deep learning

Introduction

Background

Suicide is a major public health problem worldwide, accounting for 1.4% of all deaths, equaling almost 800,000 in 2017, with many more suicide attempts [1]. Research shows that exposure to suicide-related news media content can influence suicidal behavior in vulnerable individuals in both harmful and beneficial ways. Whether suicide cases increase or decrease after exposure to suicide-related news seems to depend on specific elements of media content and language. As a recent meta-analysis of media effects research on suicide shows, most solid evidence exists for increases in suicides after exposure to news about celebrity deaths by suicide [2]. This imitation of suicidal behavior is commonly referred to as the Werther effect [3]. In contrast, exposure to other types of content may have a protective effect, with the strongest evidence existing for stories of hope, recovery, and coping [4-6]. However, broader prevention texts (ie, texts focused on prevention that were not personal stories of recovery) have also been found to be associated with protective effects in some studies [7,8]. The association of positive messaging on suicide prevention with later decreases in suicide rates has been labeled as the Papageno effect [4].

Studies investigating how exposure to media content is associated with suicidal behavior have mainly focused on traditional news outlets, such as print and web-based newspapers, radio, and television broadcasts. Investigations of the associations between social media content and suicides remain extremely scarce [9-13]. Most of the previous research on social media has focused on detecting suicidal ideation in users' posts with the purpose to identify individuals at risk, but very little research has been conducted to analyze media effects. The applied methods for identifying such individuals include machine learning as well as word dictionaries, word frequencies, topic models, and social network analysis (eg, [14-18]; for more information, see reviews by Bernert et al [19], Castillo-Sánchez G [20], Ji et al [21], Wongkoblap et al [22], and Yin et al [23]). A small number of studies have started developing machine learning classifiers for content other than suicidal ideation, despite evidence from research on traditional media that other content types can affect suicidal behavior (eg, [2,4]).

Limitations of Previous Similar Machine Learning Studies

A machine learning study categorized tweets according to expressed emotions [9], whereas 2 further studies [24,25] classified typically occurring content types, including celebrity suicide reports, suicidal intent, awareness campaigns, prevention information, condolences, and flippant remarks. Although these 2 studies include several different prevention-relevant content types, they both use the same and relatively small data set, which is limited to tweets containing celebrity names or suicidal intent. Furthermore, all these machine learning studies have used word frequency statistics as predefined features for model training, which cannot capture differences in the meaning of words across different contexts. This study addresses several gaps in the existing literature on media effects on suicidal behavior. The first is the lack of research on suicide-related social media content other than suicidal ideation. Suicide is a leading cause of deaths among young adults [1] who predominantly receive news on such platforms [13,26]. This highlights the urgency of systematic research on social media effects. In addition, social media posts often feature other content types than traditional news outlets, on which research is required. This includes diary-like posts in which people describe their personal experiences or posts addressed to their social network with the intention to prevent suicides. In this study, we investigated Twitter data and created a detailed annotation scheme for the types of suicide-related tweets that are potentially relevant to prevention efforts.

Second, regarding prevention-related media content, there is currently a discussion in the literature on whether content that highlights prevalence data to increase problem awareness has a protective effect or may even be detrimental [27]. By highlighting the prevalence of suicide and risk factors such as mental health or abuse without mentioning solutions to the issue, attempts to spread awareness may normalize suicidal behavior and trigger harmful effects [28]. In this study, we have addressed the lack of studies differentiating between prevention messages focusing on prevalence and prevention opportunities [27]. Specifically, we distinguish between awareness and prevention-focused calls for action on Twitter.

Third, the samples used in previous studies on suicide-related social media content are limited either in size (eg, [10,11]) or by a set of search terms used to collect tweets (eg, [12,29]). Sample sizes are usually small as all content needs to be annotated manually. Search terms either narrowly focus on events such as the suicide of specific celebrities (eg, [12]) or broadly include all texts containing the word suicide. Therefore, the effects of different content types may cancel each other out [2]. Thus, to systematically investigate the potentially harmful and protective effects, a large-scale and simultaneously fine-grained approach is necessary.

Overview of This Study

We have addressed these challenges by first developing a comprehensive annotation scheme that systematically organizes tweets about suicide into categories most likely to beneficially or harmfully affect suicidal and help-seeking behavior based on available evidence (eg, [2,4,6,30]). Second, we compared different natural language processing and machine learning methods to automatically detect and classify particularly important categories in large quantities of social media data. Extending previous work on different prevention-related social media content types [9,24,25], we included not only word frequency–based models but also 2 deep learning models that can capture content-dependent meanings of words [16]. We trained all models in two tasks: a multiclass classification problem with 6 main content categories and a binary classification problem of tweets about actual suicide versus off-topic tweets, which use the word suicide in another meaning or context.

The 6 main categories assessed include 5 content types that are particularly relevant for suicide prevention based on previous research. As described earlier, the strongest evidence exists for celebrity suicide case reports having harmful effects and for personal stories of hope and coping having protective effects. A second type of personal stories in Twitter posts mentions suicidal ideation and attempts without any hint at coping or recovery. Preliminary evidence suggests that such posts may have a protective effect [11]. Some evidence also suggests protective effects for general prevention messages [7,8]. We have distinguished between general prevention messages calling for action by either spreading prevention-related information or solution-oriented attitudes from those spreading problem awareness only. Finally, we included an irrelevant category to identify tweets outside the other 5 possible categories described.

Objectives

The objective of our study was to enable large-scale studies on the association between tweet content and suicidal and help-seeking behaviors. Specifically, we aimed to provide volume estimates for the different prevention-relevant tweet categories for follow-up studies on the associations of these estimates with the number of suicide cases and helpline calls.

Methods

Data Set for Training Machine Learning Models

Given that this study is part of a project on media messaging for suicide prevention in the United States, all data sets of this study include English tweets of users located in the United States. We retrieved tweet IDs via the data reseller Crimson Hexagon (now known as Brandwatch), previously used for suicide research [9,12], and then downloaded the full text of these tweets via the Twitter application programming interface. Crimson Hexagon provides access to the entire history of Twitter data and includes reliable language and location filters. The location algorithm matches 90% of all posts in a country to a location using a combination of geocoordinates, location information from user profiles, and users' time zones and languages [1].

Using a list of keywords and exclusion terms, we created a pool of unique tweets without duplicates or retweets, based on which we prepared a labeled set of tweets for training the machine learning models. We retrieved tweets posted between January 1, 2013, and May 31, 2020 (see note on dates in Multimedia Appendix 1), which contained at least one of the suicide-related search terms taken from a previous study [11]. The search terms were *suicide*, *suicidal*, *killed himself*, *killed herself*, *kill himself*, *kill herself*, *hung himself*, *hung herself*, *took his life*, *took her life*, *take his life*, *take her life*, *end his own life*, *end her own life*, *ended his own life*, *ended her own life*, *end his life*, *end her life*, *ended his life*, *ended her life*, *ends his life*, and *ends her life*.

The exclusion terms were identified by inspecting word frequency plots for common terms that may indicate that tweets used the term suicide to describe something other than someone ending their life or terms that indicated tweets about suicide bombing. We then verified whether these terms were actual mismatches by reading examples of tweets containing these terms. Thus, tweets with the most common use of the term suicide in contexts that do not refer to actual suicide could be excluded. The final list of exclusion terms was *suicide squad (a movie)*, *suicidechrist*, *SuicideGirl* (a website featuring pin-up photographs of models)*, *SuicideBoy* (male models)*, *suicideleopard (a frequently mentioned Twitter user)*, *suicidexjockey* (a Twitter user)*, *suicidal grind (a music album)*, *Epstein (excessive speculations about whether the death of Jeffrey Epstein was or was not a suicide)*, *political suicide (tweets using suicide as a metaphor for political failure)*, *Trump*, *clinton**,*Hillary*, *Biden*, and *sanders (also mostly about political suicide)*.

To avoid overlearning from multiple identical tweets, we ensured that the labeled data used for machine learning did not include any tweet duplicates. We excluded retweets (tweets categorized as retweets by Crimson Hexagon given the metadata of tweets as well as tweets containing the manual labels RT for retweets or MT for slightly modified tweets). We assembled a labeled data set of 3202 tweets by iteratively selecting tweets from a larger pool of tweets as described in the *Creating the Annotation Scheme and Labeled Data Set* section. We refer to these 3202 tweets as the *total labeled data set.* Although part of this data set was combined using keywords and model predictions (see below), a second subsample of 1000 tweets was selected randomly. We refer to these 1000 tweets as the *randomly selected labeled data set.*

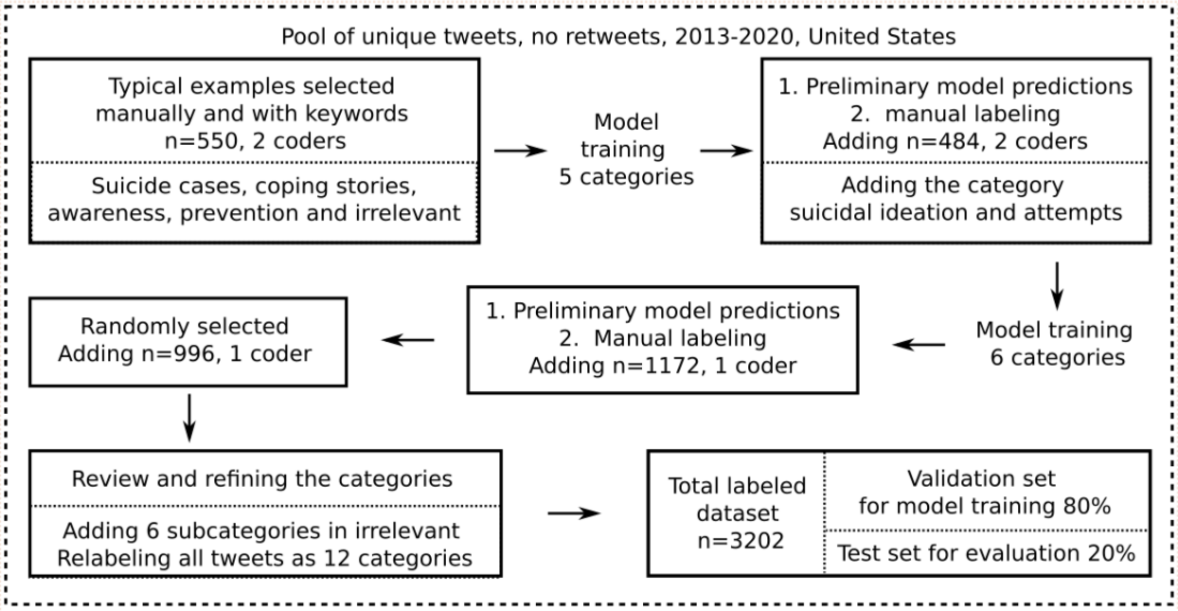
In the course of the study, we combined two other data sets: the first to compare model and human interrater reliability and the second for a face validity check and a follow-up study (Niederkrotenthaler et al, unpublished data, May 2022; see Multimedia Appendix 1 for details). Both are described in detail in the *EvaluatingReliability and Face Validity of Model Predictions for BERT* section.

Creating the Annotation Scheme and Labeled Data Set

Creating the annotation scheme and the labeled data set was an iterative human-in-the-loop process building on preliminary classifiers and annotations. We started with 5 broad categories which appeared most relevant, given previous research on traditional media (see the *Introduction* section). We then added additional categories when tweets did not fit into the existing categories but might nevertheless be associated with suicides. Given that the tweets of interest are relatively rare compared with irrelevant tweets, we used the following stepwise procedure to identify examples, which is also illustrated in Figure 1.

1. We manually selected approximately 100 tweets for each of the 5 main categories (550 tweets in total): *suicide cases*, *coping stories*, *awareness*, *prevention*, and *irrelevant tweets.* To gather the first set of tweets, we searched the data set for typical examples, both randomly and with keywords that might indicate each particular category. We iteratively expanded the list of keywords by inspecting the most frequent terms in the resulting tweets in a systematic way [31]. The full list of keywords is provided in Multimedia Appendix 1; examples are *committed or found dead* for suicide cases, *recover** or *hope* for coping stories, *lifeline or prevention* for prevention, *awareness*, and *please retweet* or *please copy* for awareness.
2. We used a preliminary machine learning model to make predictions based on the first training data set of 550 tweets to identify potential examples for each category. Next, two authors with domain expertise (TN and HM) continued annotating 100 tweets from each of the 5 predicted categories (484 after removing duplicates and missing labels from a coder). The interrater reliability for these 500 tweets was a Cohen κ of 0.75. On the basis of a careful inspection of all disagreements, we refined the definitions for all categories and adjusted the labels of all previously annotated tweets accordingly. Annotating these tweets, we further noticed a novel type of message not described in research on traditional news reporting, namely purely negative descriptions of suicidal experiences without any hint at coping, hope, or recovery. We updated the annotation scheme to include this new category *suicidal ideation and attempts*, resulting in 6 main categories. The total training set of tweets included 1034 tweets.
3. At this stage of the labeling process, we found that two dimensions were generally helpful in differentiating between categories: message type (eg, a personal story, a news story, and a call to action) and the underlying perspective about suicide (ie, if the tweet applies a problem- and suffering- or solution- and coping-centered perspective). For each message type, we noticed that some tweets implicitly or explicitly frame suicide only as a problem or from an exclusively negative or suffering perspective (categories: suicidal ideation and attempts, suicide cases, and awareness), whereas other tweets implied that coping was possible or suggested ways of dealing with the problem (categories: coping stories and prevention).
4. Repeating step 2, we trained our best preliminary model to make new predictions for the 6 categories based on all labeled tweets. Each coder annotated a different set of tweets for each predicted label until we reached a minimum of 200 training examples for the smallest categories (suicidal and coping stories). This resulted in 2206 tweets in total.
5. To mitigate bias from the search terms we used to assemble our initial training set and to estimate the distribution of tweets across categories on Twitter, HM labeled a random sample of 996 tweets (initially 1000, with 4 were not labeled owing to a displaying error in the used spreadsheet). These were then added to the training set, resulting in a total sample of 3202 tweets.
6. After reviewing the entire training set, we finally refined the categories to allow for the following distinctions: for stories about coping and suicidal experiences, we differentiated the perspective from which an experience was described (first or third person), which experience was described (the one of a concerned or a bereaved individual), and whether a tweet was shared by news media or individual users. For reporting of cases, we distinguished tweets about individuals who had actually died by suicide from tweets about someone saving the life of an individual who was about to take his or her life. Finally, we organized the tweet categories according to 2 dimensions further described in the *Annotation Scheme* section.

Figure 1. Creating the labeled data set and annotation scheme. Each box describes how tweets were selected from the large pool of available tweets, how many tweets were added to the training data set in each step (after removing duplicates), and how many coders labeled each tweet. When we used preliminary model predictions to identify potential candidates for each category, we deleted the model labels before manual coding. After rounds with 2 coders, we checked interrater reliability, adapted the annotation scheme until all disagreements were clarified, and relabeled the respective sample.



Annotation Scheme

Overview

The annotation scheme divides tweets into 12 categories, including 10 categories of interest and 2 irrelevant categories. Each category can be described in terms of two dimensions: the message type (eg, a personal story, a news story, and a call to action) and the underlying perspective about suicide (ie, if the tweet applies a problem- or solution-centered perspective). The perspective distinguishes messages that implicitly or explicitly frame suicide only as a problem or from an exclusively negative or suffering perspective from messages that imply that coping is possible or suggest ways of dealing with the problem. The organization of the tweet categories along these 2 dimensions is presented in Table 1. Detailed instructions for annotating tweets are provided in Multimedia Appendix 2. These include prioritization rules for how to deal with ambiguous tweets that may fit into more than one category.

Table 1. Annotation scheme of content categories organized along two dimensions: message type and underlying perspective about suicide.

Message type	Underlying perspective	
	Problem and suffering	Solution and coping
Personal experiences first or third person	Suicidal ideation and attempts[a]	Coping (Papageno)[a]
News about experiences and behavior	News suicidal ideation and attempts	News coping
Experience of bereaved	Bereaved negative	Bereaved coping
Case reports	Suicide cases (Werther)[a]	Lives saved
Calls for action	Awareness[a]	Prevention[a]
Irrelevant[a]		
Suicide other	Murder-suicides, history, fiction, not being suicidal, and opinions	
Off-topic[b]	Bombings, euthanasia, jokes, metaphors, and band or song names	

[a] The 6 main categories classified in machine learning task 1.

[b] Task 2 distinguished the off-topic category from all other categories (see *Classification Tasks*).

Description of Content Categories

For each message type (except for irrelevant messages), there is a category for more problem- or suffering-focused tweets and for more solution- or coping-focused tweets:

1. Personal stories describing the experience of an affected individual either in first- or third-person perspective:
    - Suicidal ideation and attempts: Personal stories about an individual's negative experiences with suicidal thoughts, related suffering (eg, depression), suicidal communication and announcements, or suicide attempts without a sense of coping or hope

- Coping: Personal stories about an individual's experience with suicidal thoughts or a suicide attempt, with a sense of hope, recovery, coping, or mentioning an alternative to suicide. The sentiment does not have to be positive. A neutral tone or talking about difficult experiences with a sense of coping or mentioning recovery is sufficient. Previous research has suggested that such messages may have a Papageno effect.

2. News reports about suicidal experiences and behavior except cases, often about celebrities:
    - News suicidal ideation and attempts: About suicidal experiences without any mention of coping, including reports on suicidal ideation, suicide attempts, announcements of suicide, and someone being put on "suicide watch"
    - News coping: About attempted or successful coping with or recovering from a suicidal crises.
3. Tweets describing the experience of a person who has lost someone to suicide from the first- or third-person perspective:
    - Bereaved negative: Describes the suffering or purely negative experience of a person who has lost someone to suicide, including depression, grief, and loss. These tweets necessarily refer to a suicide case but are labeled as bereaved as long as they focus on the experience of bereaved individuals.
    - Bereaved coping: Describes the experience of a bereaved person with a sense of hope, recovery, or coping. The sentiment does not have to be positive. A neutral tone, or talking about difficult experiences with a sense of coping or mentioning recovery is sufficient.
4. Reports of a particular completed or prevented suicide cases, often news reports:
    - Suicide cases: About an individual suicide or a timely or geographical suicide cluster. Suicide cases have priority over the definition criteria of other categories (except tweets focusing on bereaved individuals, which are always related to a suicide case). Previous research suggests that such messages on individual suicide deaths (especially about celebrities) may have a Werther effect.
    - Lives saved: News report or personal message about someone saving a life. In contrast to prevention tweets, these lives are often being saved coincidentally.
5. Calls for action are general statements calling for actions addressing the problem of suicide and intending to spread problem awareness or prevention-related information:
    - Awareness: Tweets intending to spread awareness for the problem of suicide, often focusing on high suicide rates or associations with bullying, racism, depression, and Veterans without hinting at any solution. These are often the reports of research findings or suicide statistics.
    - Prevention: Tweets spreading information about a solution or an attempt to solve the problem of suicide, including prevention at an individual (eg, do not leave people alone in crisis situations) or public health level (eg, safety nets on bridges). Hinting at a solution or a way of dealing with the problem is sufficient. No specific action needs to be described. These tweets often include a helpline number. Announcements of prevention events and broad recommendations for actions also count—donations and prayers with a focus on a solution for suicide, being there for someone, telling people that they matter, taking a course about suicide prevention, and warning signs to watch out for.
6. Irrelevant, including messages that do not fit into any of the above categories:
    - Suicide other: Anything about actual suicide but not clearly related to any other above category, including murder-suicides, confident statements that something was not a suicide, convincing statements of not being suicidal, historical tweets about suicides that were a minimum of 40 years ago (eg, about the suicide of Hitler), movies, books, novels, and fiction about suicide.
    - Off-topic: Messages that use the term suicide in a context other than suicide. This includes messages on euthanasia, suicide bombing and suicide attacks, messages that are (suspected) jokes, irony, sarcasm, flippant remarks or really unclear in terms of authenticity, and messages that use suicidal or suicide to exaggerate an emotional experience (unclear if serious) or as a metaphor (eg, political, financial, or career suicide, suicide workout, and suicidal immigration policies), and messages about "suicidal animals" (eg, killed by car).

Analysis

Software

Data analysis was performed using R for intercoder reliability, descriptive statistics, and figures (version 3.6.3, R Foundation for Statistical Computing). The main libraries used in R were tidyverse, caret, and DescTools [32-34]. For training deep learning models, we used Python 3.6 (Python Software Foundation). The main packages were the ktrain wrapper [35] for the deep learning library TensorFlow Keras [36] and the scikit-learn library [37] for term frequency-inverse document frequency (TF-IDF) and support vector machines (SVM). For links to code and data, refer to the *Data and Code Availability* section.

Text Preprocessing

We applied standard preprocessing strategies (eg, [38]) and replaced all URLs with a general marker token "*http*," all mentions (tags of Twitter users) with "*@user*," and lowercased all words. The latter allowed using the smaller, more resource-efficient Bidirectional Encoder Representations from Transformers (BERT)–lowercase model (see the *BERT-base* section). We kept emoji, stop words, and punctuation separated into single tokens, given that they can indicate the emotional connotation of a message (eg, expressing excitement or surprise [39] or frequent singular pronouns indicating suicidal ideation [15]). We report the effects of additional standard different preprocessing steps, namely removing digits, punctuation, stop words, and lemmatization, in Multimedia Appendix 1. The basic preprocessing strategy yielded the most consistently high-performance scores on the validation set and was therefore used for all analyses. After preprocessing, the mean length of tweets in our labeled data set was 25 tokens, the 95th and 99th percentile were 57 and 67 tokens, respectively (Figure S1 in Multimedia Appendix 1). On the basis of this, we used 80 tokens as the maximum sequence length for model input.

Classification Tasks

Task 1: 6 Main Categories

We trained our models to classify among categories with at least 200 tweets to allow sufficient training data. From the categories of interest, these were messages about (1) *personal experiences of coping*, (2) *personal experiences of suicidal ideation and attempts*, (3) *suicide cases*, (4) *awareness*, and (5) *prevention*. We assigned all tweets from smaller categories (suicidal and coping news, negative and coping experiences of the bereaved, and lives saved) to the category *suicide other*, which belongs to the larger category of *irrelevant* tweets. In this task, we did not differentiate between irrelevant tweets that were about suicide (suicide other) and off-topic tweets, which used the word suicide in some other way. Instead, we subsumed suicide other and off-topic tweets in the category (6) *irrelevant*.

Task 2: Detecting Content About Actual Suicide

This binary classification distinguishes tweets that are (1) *about actual suicide* in the meaning of someone taking their own life, from tweets that are (2) *off-topic*, that is, use the word suicide in some other context. In our annotation schema, this task therefore separates the off-topic category from all other categories. The resulting label predictions allow to estimate the total volume of tweets about actual suicide, thereby improving the total volume estimates only based on keyword searches.

Machine Learning Models and Model Training

Train-, Validation-, and Test Set

Before training models, we divided the data set of 3202 tweets into training (2049/3202, 63.99%), validation (512/3202, 15.99%), and test sets (641/3202, 20.02%), stratifying per tweet category to have a similar distribution in all sets. The training set was used for fitting the parameters of the classifier using 5-fold cross-validation. The validation set was used to tune the hyperparameters (eg, learning rate) and evaluate the model developed on the training data. After model training, we used the test set only once per model to estimate its ability to generalize to novel texts.

Majority Classifier

We used a naïve classifier that always predicts the majority class as a baseline to compare the other models.

TF-IDF and SVM

TF-IDF represents the text of tweets using weighted word *frequencies (f)*, which reflect how important a *term (t)* is to a *document (d, here a tweet)* in a corpus (all tweets). We slightly adjust the original formula for TF-IDF by adding 1 in the numerator and denominator, to ensure each word occurs at least once and prevent 0 division [40]: $tf\text{-}idf(t,d) = tf(t,d) \times \log([N+1]/[df+1])$.

The resulting value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus that contain the word. This helps to adjust the weight of uncommon words that are more important for distinguishing different documents from each other than words that occur in every single document. After building the TF-IDF representation, we trained a SVM classifier using all term values as features.

To identify the best TF-IDF representation and SVM classifier, we ran a grid search across the following dimensions. For TF-IDF, we (1) included only unigrams or unigrams+bigrams and (2) reduced the text to its n top features ordered by term frequency, where $n \in \{10{,}000; 25{,}000; 50{,}000\}$. For the SVM, we tested different hyper-parameters, namely (1) regularization parameter $C \in \{0, 1\}$, which determines the strength of the regularization and (2) class weight $cw \in \{\text{balanced, none}\}$, which determines whether the weights of the classes are automatically adjusted inversely proportional to class frequencies. We further tested (3) a linear and a radial basis function kernel and (4) decision function shapes one versus one and one versus rest. Optimal results were achieved including both unigrams and bigrams as text representation, 10,000 top features, and an SVM with C=0.82 in task 1 and C=0.46 in task 2, cw=balanced, a linear kernel, L2 penalty, and a one-versus-one decision function shape.

BERT Base

We used a transfer learning approach based on a pretrained BERT-base-uncased model [41]. BERT is an autoencoding deep contextual language representation model developed by Google AI, which has 12 transformer layers, 12 self-attention heads, and a hidden size of 768. It is designed to pretrain bidirectional representations of word sequences, that is, it learns from both the left-side and right-side context of a word in all of its layers. BERT was pretrained with masked language modeling: a percentage (approximately 15%) of words in the sentence is randomly masked, and the model tries to predict the masked words from the sequence of other words. BERT was further trained to predict the next sentence from the previous sentence in the data.

XLNet Base

A known limitation of BERT is that it neglects the dependence between the different masked words in a sentence. When predicting a word from a sequence that does not include the other masked words, BERT lacks information about the dependence between the masked words and the predicted word. Unlike autoregressive models, BERT further predicts all masked words simultaneously, and thus lacks some information about the order of words. XLNet [42] has a similar architecture as that of BERT but addresses these shortcomings through permutation language modeling, predicting each word from all possible permutations of other words in the sentence. It thereby improves both on previous autoregressive models by using all words in the sentence and on BERT by considering the order dependence of words. In addition, it incorporates some techniques from Transformers-XL [43], which also allows it to learn from the longer context before each word (relative positional encoding and segment recurrence mechanism).

Fine-tuning of BERT and XLNet

We fine-tuned the pretrained BERT-base uncased model and the XLNet-base model to our training data set as in the study by Liu et al [16], added one dense output layer to reduce the dimensions of the model's last layer to the number of labels in the classification task, and trained all the parameters simultaneously. We ran a hyper-parameter search to determine good learning rate (LR) candidates and subsequently tested each LR by running 3 experiments with different seeds ∊ {1,2,3}. We aimed to find the maximal LR associated with a still-falling loss (before the loss diverging) by training for 5 epochs with learning rates ∊ {2e-5, 3e-5, 5e-5}. The reported results for BERT in Task 1 (6 classes) were the result of fine-tuning with a LR=2e-5 for 7 epochs and seed=1. The results for task 2 (about actual suicide) were based on a BERT model trained with a LR=1e-5, 10 epochs, and seed=1. The reported results for XLNet were based on model training with LR=2e-5, 8 epochs, and seed=1 in both tasks.

Metrics for Comparing Machine Learning Models

We used various evaluation metrics to compare different machine learning models. *Accuracy* indicates the percentage of correct predictions (true positive and true negative). It is a global metric calculated for all the classes in a data set. In data sets with large class imbalances, it can be high even if it always predicts only the majority class (eg, the irrelevant category in task 1). In this case, the model may not have learned anything despite its high accuracy. *Precision* indicates the proportion of correct "positive" predictions out of all predictions; for example, how many of all predicted coping tweets were actually labeled as coping tweets by human raters. *Recall* indicates the proportion of all "true" cases (eg, all actual coping tweets) that the model detects. The F_1-score is the harmonic mean between precision and recall ($F_1 = 2 \times [\text{precision} \times \text{recall}]/[\text{precision} + \text{recall}]$). Precision, recall, and F_1-scores were calculated for each category and can be averaged across classes to produce a macroaverage. For category-specific precision and recall, we provide 95% binomial CIs calculated using the Clopper-Pearson method.

To compare models, we report macroaverages of model performance scores for both the validation and test sets, that is, we calculate the mean of the performance measures of each class to have an aggregate measurement robust to class imbalance. Although good scores on the training set indicate that the model has learned patterns existing in the training set, good scores on the test set additionally indicate an ability to generalize to novel data.

For determining the model to make predictions for a follow-up study (Niederkrotenthaler et al, unpublished data, May 2022; see Multimedia Appendix 1 for details), we decided a priori that we would prioritize precision over recall for task 1, which aims to identify specific categories of tweets. The rationale behind this is that our follow-up study focused on identifying specific Twitter signals (ie, the percentage of coping tweets) that are associated with suicide cases and helpline calls. In such a situation, a false negative is less costly than a false positive, that is, missing a tweet is less costly than falsely including a tweet in a certain category. Prioritizing precision ensures that we only count a tweet when it belongs to a category with a high probability. Furthermore, because of the large number of tweets, the proportion assigned to each category should accurately reflect the true proportion, even if not all the tweets are recognized. In contrast, task 2 makes predictions that aim to capture the entire discussion about actual suicides on Twitter. When choosing the best model for task 2, we focused on the F_1-score. Here, we aimed to capture the total volume of tweets about suicide as fully as possible, as well as at accurate predictions at the tweet level. False positives are less critical as a problem in task 2 than in task 1, because we look at total tweet volume and do not try to distinguish between the specific effect of a certain tweet category. This is best captured with the F_1-score, which balances recall and precision. In any case, none of these a priori decisions had consequences for our results, given that BERT and XLNet performed very similarly and much better than the other models.

Evaluating Reliability and Face Validity of Model Predictions for BERT

Comparing Model and Human Interrater Reliability

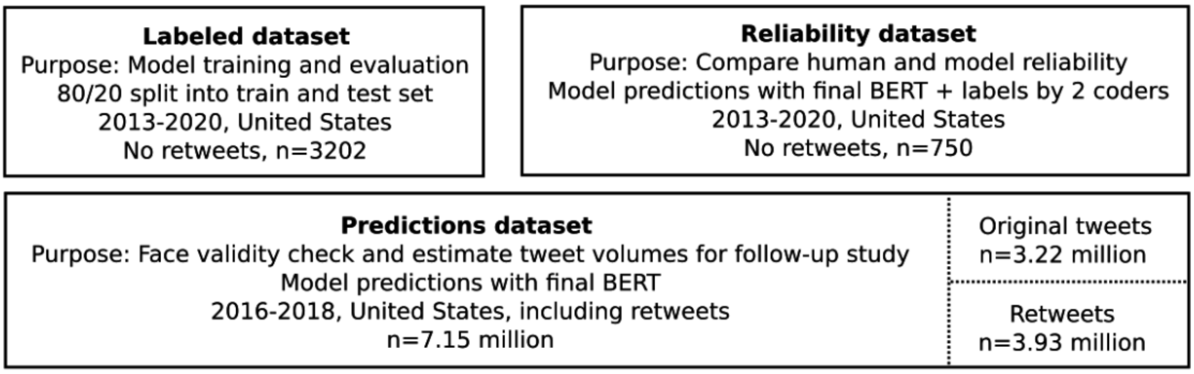
To compare the models' reliability to human interrater reliability on novel data, we made predictions using one of the best models (BERT) for tweets from the full data set that were not part of the labeled data set. We selected 150 tweets per predicted label for each of the 5 relevant main categories. In all, 2 independent human coders manually labeled these tweets until we reached at least 80 tweets per main category. The final set of 750 labeled tweets comprised the reliability data set.

Face Validity Check With the Predictions Data Set

For a face validity check and a follow-up study [30], we estimated the daily volume of tweets per category that Twitter users may have been exposed to between January 1, 2016, and December 31, 2018. For this, we created a data set with the same keywords and exclusion terms as the machine learning data set but including retweets (to account for the full volume) and for a shorter period of 3 years (determined by the follow-up study [30]). This resulted in 7,150,610 tweets in English from users in the United States. We used the machine learning model BERT to predict the category labels for these tweets and calculated the daily percentage of tweets per category. We refer to this data set with model predictions for approximately 7 million tweets as the prediction data set. As a face validity check, we plotted the time series of tweet volumes per category and identified the events associated with the largest frequency peaks. We investigated word frequencies on these days, read the tweets containing the most frequent terms, and Googled these terms plus the date, or the tweet in quotes, to find (news) reports about the event. The follow-up study [30] investigated the associations of these daily tweet volumes with suicide cases and helpline calls in the United States. It has access to suicide case data from the Centers for Disease Control and Prevention and call data from the United States suicide prevention lifeline for the years 2016 to 2018, which was the reason for estimating tweet volumes for this period.

Properties of all data sets used in this study, including the labeled machine learning data set and those for comparing model and human performance and the face validity check are depicted in Figure 2.

Figure 2. Overview of characteristics of data sets. Each box describes the purpose of the data set, further details on how it was used or created, and the sample size. Only the predictions data set includes retweets, as it aims to capture the full volume of tweets posted on a given day. BERT: Bidirectional Encoder Representations from Transformers.



Results

Frequency of Tweets per Category

Table 2 displays the proportion of tweets per main category in our labeled data set and in 2 different samples used to estimate the natural frequency of categories on Twitter. First, we used a subsample of the labeled data set of 1000 randomly selected tweets (ie, selected without keywords or model predictions, 996 after 4 labeling; see *Creating the Annotation Scheme and Labeled Data Set* section) to estimate the frequency of original tweets, without counting retweets. For the second estimate, we used predictions by the best model (BERT) to label tweets in the prediction data set, which included retweets. The 2 estimates were similar for suicidal ideation and attempts and suicide cases.

The percentages per category in Table 2 demonstrate that we managed to include proportionally more rare tweet categories, such as coping and suicidal ideation stories in our training set. Nevertheless, irrelevant tweets, particularly off-topic tweets, still make up a majority of tweets in our data set.

Table 2. Distribution of tweets across categories for manual labels and model predictions.

Category label	Total labeled sample (n=3202)	Subset of labeled tweets, randomly selected (n=1000)	Estimated frequency in predictions data set (including retweets; n=7.15 million), n (%)[a]	
			Task 1	Task 2
Suicidal ideation and attempts, n (%)	284 (8.87)	63 (6.33)	367,135.56 (5.13)	5,471,499 (76.52)
Coping, n (%)	205 (6.4)	26 (2.71)	90,328.99 (1.26)	5,471,499 (76.52)
Awareness, n (%)	314 (9.81)	126 (12.54)	1,577,650 (22.06)	5,471,499 (76.52)
Prevention, n (%)	457 (14.27)	71 (7.13)	1,109,223.6 (15.51)	5,471,499 (76.52)
Suicide cases, n (%)	514 (16.05)	129 (12.95)	1,155,277.92 (16.16)	5,471,499, (76.52)
Irrelevant, n (%)	1428 (44.5)	581 (58.33)	2,850,994 (39.88)	5,471,499 (76.52)
Subcategories of irrelevant, n (%)				
News suicidal	68 (2.12)	20 (2.01)	2,850,994 (39.88)	5,471,499 (76.52)
News coping	27 (0.84)	5 (0.5)	2,850,994 (39.88)	5,471,499 (76.52)
Bereaved negative	34 (1.06)	7 (0.7)	2,850,994 (39.88)	5,471,499 (76.52)
Bereaved coping	34 (1.06)	5 (0.5)	2,850,994 (39.88)	5471499 (76.52)
Live saved	13 (0.41)	2 (0.2)	2,850,994 (39.88)	5,471,499 (76.52)
Suicide other	440 (13.74)	206 (20.68)	2,850,994 (39.88)	5,471,499 (76.52)
Off-topic	812 (25.36)	336 (33.73)	2,850,994 (39.88)	1,679,111 (23.48)

[a] For the predictions data set: Absolute values and percentages were weighted (ie, divided) by the model's recall (proportion of all true cases the model detects). Sample values (n) and percentage for the irrelevant category were calculated by subtracting the sum of all other categories from the total sample size and 100, respectively. If several cells contain the same values, this is because they were subsumed to one higher-level category (*irrelevant* in task 1, *about suicide* in task 2) in the respective classification task.

Model Performance

Task 1: 6 Main Categories

Performance scores averaged across all 6 tweet categories (Table 3) show that all deep learning models performed very similarly and substantially better than the TF-IDF and SVM approaches. However, TF-IDF and SVM were clearly better than a naïve majority classifier. It reached scores from 0.61 to 0.66, which were nearly identical on the validation and test sets. For BERT and XLNet, all scores were at or above 0.70, and only 0.1 to 0.3 lower on the test than the validation set, indicating a good ability to generalize to new tweets. The macroaverage performance scores in all 5 runs for BERT and XLNet were approximately 10% higher than the TF-IDF and SVM macroaverages (Table S2 in Multimedia Appendix 1).

Given that the macroaverage performances were substantially lower for the majority classifier, we focused on the 3 other models for intraclass scores (Table 4 and Figure 3). To choose a model for making predictions, we focused on F_1-scores and precision (see the section on evaluation metrics). F_1-scores were higher for BERT and XLNet than for TF-IDF and SVM for all relevant categories, with clear differences for some categories (suicidal, coping, and awareness) and very small differences for others (suicide cases and prevention). For BERT and XLNet, F_1-scores were almost identical for all categories. BERT yielded higher precision for coping and prevention tweets, 2 crucial categories for a follow-up publication (Niederkrotenthaler et al, unpublished data, May 2022; see Multimedia Appendix 1 for details). Therefore, we chose BERT as the model to make predictions for further analyses. It should be noted here that CIs are quite large because of the limited size of the test set per class and entirely overlap for BERT and XLNet and somewhat overlap for most categories with TF-IDF and SVM. Nonetheless, the performance scores in the 5 runs of BERT and XLNet were higher than those of TF-IDF and SVM in almost all cases for all relevant categories (Table S3 in Multimedia Appendix 1). Only in the case of precision for prevention tweets, TF-IDF and SVM performed similarly well in 3 out of 5 runs.

Overall, BERT correctly classified 73% of the tweets in the test set. F_1-scores lay between 0.70 and 0.85 for the different categories of interest (Table 4 and Figure 3), with the exception of the suicidal ideation and attempt category, with an F_1-score of 0.51. More specifically, recall for suicidal ideation and attempt was relatively low (0.45), indicating difficulties in detecting all such tweets, whereas precision was higher, with 0.58. All performance scores were particularly good (>0.81) for prevention tweets and high for tweets about suicide cases (>0.75). For coping tweets, BERT achieved very high precision (0.76) but lower recall (0.69), which resembles the pattern observed for suicidal tweets. Performance scores for awareness tweets were approximately 70%.

Table 3. Macroaveraged performance metrics and accuracy cross all 6 categories on the validation and test set.

Model	Validation set (n=513)				Test set (n=641)			
	Precision	Recall	F_1	Accuracy	Precision	Recall	F_1	Accuracy
Majority classifier	0.07	0.17	0.10	0.45	0.07	0.17	0.10	0.44
TF-IDF[a] and SVM[b]	0.61	0.63	0.62	0.66	0.61	0.65	0.62	0.66
BERT[c,d]	0.73	0.71	0.71	0.76	0.72	0.69	0.70	0.73
XLNet[d]	0.74	0.73	0.73	0.77	0.71	0.71	0.71	0.74

[a] TF-IDF: term frequency-inverse document frequency.

[b] SVM: support vector machine.

[c] BERT: Bidirectional Encoder Representations from Transformers.

[d] Given that the performance of both deep learning models with fixed seeds and parameters varied slightly from run to run owing to internal segmentation, we ran these models 5 times. We report the average of all 5 runs in this section and include the metrics for each individual run in Table S2, in Multimedia Appendix 1.

Table 4. Intraclass performance metrics on the test set.

Category	TF-IDF[a] and SVM[b]			BERT[c,d]			XLNet[d]		
	Precision (95% CI)	Recall (95% CI)	F_1	Precision (95% CI)	Recall (95% CI)	F_1	Precision (95% CI)	Recall (95% CI)	F_1
Suicidal ideation (n=57)	0.32 (21.93-43.58)	0.44 (30.74-57.64)	0.37	0.58 (43.25-73.66)	0.45 (32.36-59.34)	0.51	0.60 (46.11-74.16)	0.54 (40.66-67.64)	0.55
Coping (n=42)	0.44 (31.55-57.55)	0.64 (48.03-78.45)	0.52	0.76 (59.76-88.56)	0.69 (52.91-82.38)	0.72	0.71 (54.80-83.24)	0.74 (57.96-86.14)	0.73
Awareness (n=63)	0.65 (51.60-76.87)	0.62 (48.80-73.85)	0.63	0.71 (58.05-81.80)	0.70 (56.98-80.77)	0.70	0.69 (56.74-79.76)	0.74 (62.06-84.73)	0.72
Prevention (n=91)	0.83 (74.00-90.36)	0.82 (73.02-89.60)	0.83	0.81 (71.93-88.16)	0.89 (80.72-94.60)	0.85	0.82 (72.27-88.62)	0.87 (78.10-93.00)	0.84
Suicide cases (n=103)	0.70 (60.82-78.77)	0.74 (64.20-81.96)	0.72	0.75 (65.14-82.49)	0.77 (67.34-84.46)	0.76	0.78 (68.31-85.52)	0.75 (65.24-82.80)	0.76
Irrelevant (n=285)	0.74 (67.78-79.18)	0.63 (57.27-68.77)	0.68	0.64 (57.76-69.11)	0.65 (59.06-70.45)	0.64	0.68 (61.96-73.46)	0.64 (57.99-69.44)	0.66

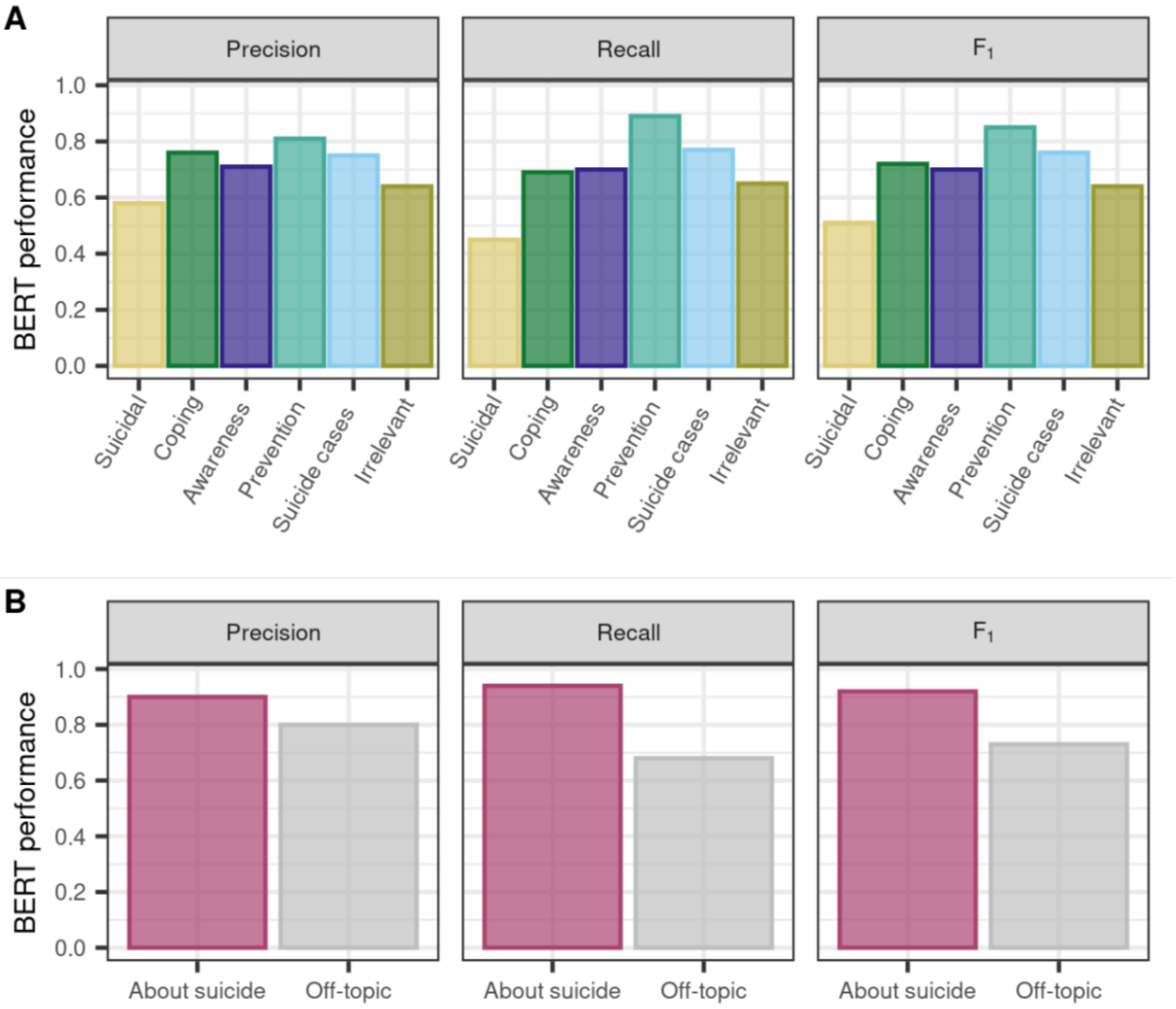
[a] TF-IDF: term frequency-inverse document frequency.

[b] SVM: support vector machine.

[c] BERT: Bidirectional Encoder Representations from Transformers.

[d] Scores are averages across 5 model runs for BERT and XLNet. Table S3 in Multimedia Appendix 1 shows separate runs.

Figure 3. Performance scores per category for Bidirectional Encoder Representations from Transformers (BERT) for the 6 main categories (A) and for tweets about actual suicide versus off-topic tweets (B).



Task 2: About Actual Suicide

Best performances for separating tweets about actual suicide from off-topic tweets (Table 5) were observed with BERT. However, XLNet performances were very similar, with largely overlapping CIs. With TF-IDF and SVM, recall for about suicide tweets and precision for off-topic tweets were significantly lower than the deep learning scores, whereas precision for about suicide and recall for off-topic was not significantly different. The model with overall highest scores, BERT, correctly labeled 88.5% of tweets as about suicide versus off-topic, with very similar scores on the validation and test sets. F_1-scores for about suicide versus off-topic tweets in the test set were 0.92 and 0.73, respectively (Table 6). All metrics were at least 10% higher for tweets about suicide than for the off-topic tweets. In particular, recall was very high for tweets about suicide (94%), which indicates that volume estimates for tweets related to suicide would be quite complete. The precision for tweets about suicide was 90%, indicating that positive predictions of the model were very reliable.

Table 5. Macroaveraged performance metrics and accuracy for task 2 (about suicide vs off-topic) on the validation and test sets.

Model	Validation set (n=513)				Test set (n=641)			
	Precision	Recall	F_1	Accuracy	Precision	Recall	F_1	Accuracy
Majority classifier	0.37	0.50	0.43	0.75	0.37	0.50	0.43	0.75
TF-IDF[a] and SVM[b]	0.74	0.77	0.75	0.80	0.75	0.77	0.76	0.81
BERT[c]	0.85	0.81	0.83	0.88	0.85	0.81	0.83	0.88
XLNet	0.84	0.78	0.81	0.87	0.83	0.80	0.81	0.87

[a] TF-IDF: term frequency-inverse document frequency.

[b] SVM: support vector machine.

[c] BERT: Bidirectional Encoder Representations from Transformers.

Table 6. Intraclass performance metrics for deep learning models in task 2 (about suicide vs off-topic) on the test set.

Test set and model	About suicide (n=478)			Off-topic (n=163)		
	Precision (95% CI)	Recall (95% CI)	F_1	Precision (95% CI)	Recall (95% CI)	F_1
TF-IDF[a] and SVM[b]	0.89 (85.74-91.71)	0.85 (80.96-87.64)	0.87	0.60 (53.03-67.49)	0.69 (61.63-76.30)	0.65
BERT[c,d]	0.90 (87.42-92.81)	0.94 (91.64-96.07)	0.92	0.80 (71.62-85.67)	0.68 (60.35-75.17)	0.73
XLNet[d]	0.90 (87.12-92.59)	0.93 (90.68-95.38)	0.92	0.76 (68.60-83.06)	0.67 (59.72-74.60)	0.71

[a] TF-IDF: term frequency-inverse document frequency.

[b] SVM: support vector machine.

[c] BERT: Bidirectional Encoder Representations from Transformers.

[d] Scores are averages across 5 model runs for BERT and XLNet. Table S5 in Multimedia Appendix 1 shows separate runs.

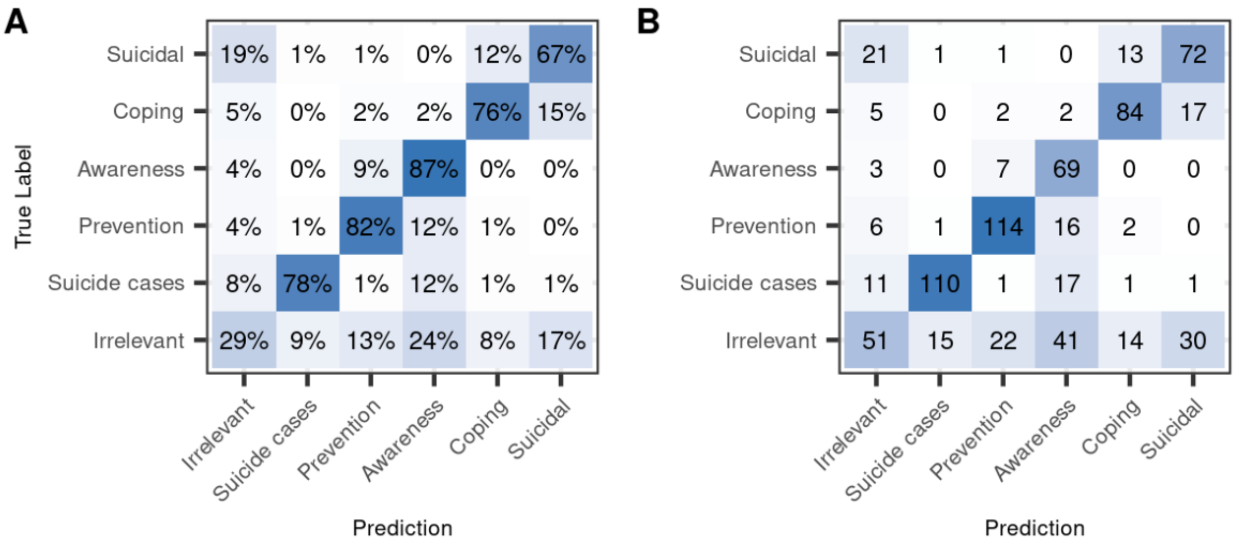
Comparing Model and Human Interrater Reliability

Task 1: 6 Main Categories

The interrater reliability (Cohen κ) for the 6 main categories was 0.70 (95% CI 0.67-0.74) between 2 human coders and 0.60 (95% CI 0.56-0.64) and 0.63 (95% CI 0.59-0.67) between each human coder and the BERT model, respectively. The lower agreement with BERT compared with between humans was mainly driven by the irrelevant class. Excluding it from analysis yielded κ=0.85 (95% CI 0.82-0.89) between human raters and κ=0.81 (95% CI 0.77-0.85) and 0.80 (95% CI 0.76-0.84) between BERT and each human rater. These overlapping CIs indicate a nonsignificant difference and show that BERT achieved near human-level accuracy for the relevant categories.

Precision and recall comparisons between model and human performance per tweet category are shown in Table S6 in Multimedia Appendix 1, and the confusion matrix for coder 1 and BERT is shown in Figure 4. First, we report the metrics for the model versus each coder, with the coder as the ground truth. Second, we report the same metrics for coder 2 compared with coder 1 as the ground truth. Model precision was clearly lower than between-human precision for *suicidal ideation and attempts* and *awareness* messages, more comparable for *coping stories*, and very similar for *prevention* and *suicide case tweets*. Recall is clearly higher between human raters for *suicidal and coping* stories and similar for *suicide cases*. For awareness and prevention tweets, the model actually achieves better recall than human coders. Thus, the model seems quite good at detecting awareness tweets but is not very precise in return.

Figure 4. Confusion matrix of true and predicted labels in the reliability data set. (A) percentages and (B) count of tweets per true and predicted category. The diagonal from bottom left to top right represents correct predictions. True labels are labels by coder 1, and predicted labels are by Bidirectional Encoder Representations from Transformers (BERT).



Task 2: About Actual Suicide

When categorizing tweets as being about actual suicide versus off-topic, human interrater reliability was κ=0.44 (95% CI 0.29-0.58) compared with κ=0.15 (95% CI −0.07 to 0.37) and κ=0.21 (95% CI −0.01 to 0.44) between each coder and BERT. These low κ coefficients were mainly driven by low performances for the irrelevant off-topic category between both human coders (coder 1-coder 2: precision=0.52, recall=0.44, and F_1=0.48), which were even lower when comparing human to model labels (coder 1-BERT: precision=0.26, recall=0.13, and F_1=0.17; coder 2-BERT: precision=0.39, recall=0.16, and F_1=0.23). In contrast, the performance for the suicide category was very high when comparing human labels (precision=0.96, recall=0.97, and F_1=0.96), as well as when comparing human and model labels (coder 1-BERT: precision=0.94, recall=0.98, and F_1=0.96; coder 2-BERT: precision=0.94, recall=0.98, and F_1=0.96). This shows that the 2 coders and the model agreed which tweets were about actual suicide and detected most tweets that the other coder had labeled as about suicide. However, they agreed less when judging whether a tweet was not about actual suicide, hinting at the inherent difficulty of judging whether something is serious, sarcastic, or metaphorical. In any case, for future studies correlating tweets about suicide with behavior in the population, only the about suicide category, which can be reliably detected by humans and the machine learning model, is relevant.

Error Analysis

Figure 4 shows the confusion matrix of the true and predicted labels for BERT for the 6 main categories in the reliability data set. Most misclassifications were predictions of the label irrelevant. Such false negatives are less problematic than misclassifications between relevant categories, as we prioritized precision over recall. Among the relevant categories, there were 5 cases in which coder 1 and the model labeled >9 but a maximum of 15% of tweets differently: (1 and 2) confusions between coping and suicidal tweets in both directions, (3 and 4) confusions between awareness and prevention tweets in both directions, and (5) tweets about suicide cases misclassified as awareness tweets.

For the 13 “true” suicidal labels where coder 1 and the model disagreed (12% of the 108 suicidal tweets), only 2 of the model labels were clear errors, all other tweets were ambiguous. Coder 2 and the model agreed on the coping label for one-third of these tweets (4/13, 31%), indicating the difficulty of clearly separating personal stories about suicidal ideation and coping even for humans. The model’s label more closely matched the category definition than coder 1’s label in at least 5 of the 13 cases (38.4%). Many of the ambiguous tweets described suicidal ideation in the past, implicitly hinting that the suicidal phase was over when the tweet was written. Out of 17 misclassifications of coder 1’s coping tweets (15% of the 110 coping tweets), coder 2 and the model agreed on the suicidal label in 6 cases (35.2%), suggesting that many of these misclassified tweets were ambiguous. Although 12 of these 17 (71%) misclassifications were actual model errors, most of them were understandable, given that coping was described implicitly by means of suicidal ideation in the past or that strong suicidal ideation was expressed along with a way in which the person deals with it.

Misclassifications of awareness as prevention tweets (7/79, 9%) were errors by coder 1 rather than the model in 4 of 7 cases (57%), indicating that model performances are higher for awareness tweets as the scores in Table 3 suggest. In contrast, when the model labeled prevention tweets as awareness, these were mostly clear mistakes (only 3 out of 16, 19%, were ambiguous errors by the coder). Finally, suicide cases mislabeled as awareness were mostly actual errors by the model, but ambiguous in 4 out of 17 cases (24%) and actually correct in 2 out of 17 cases (12%).

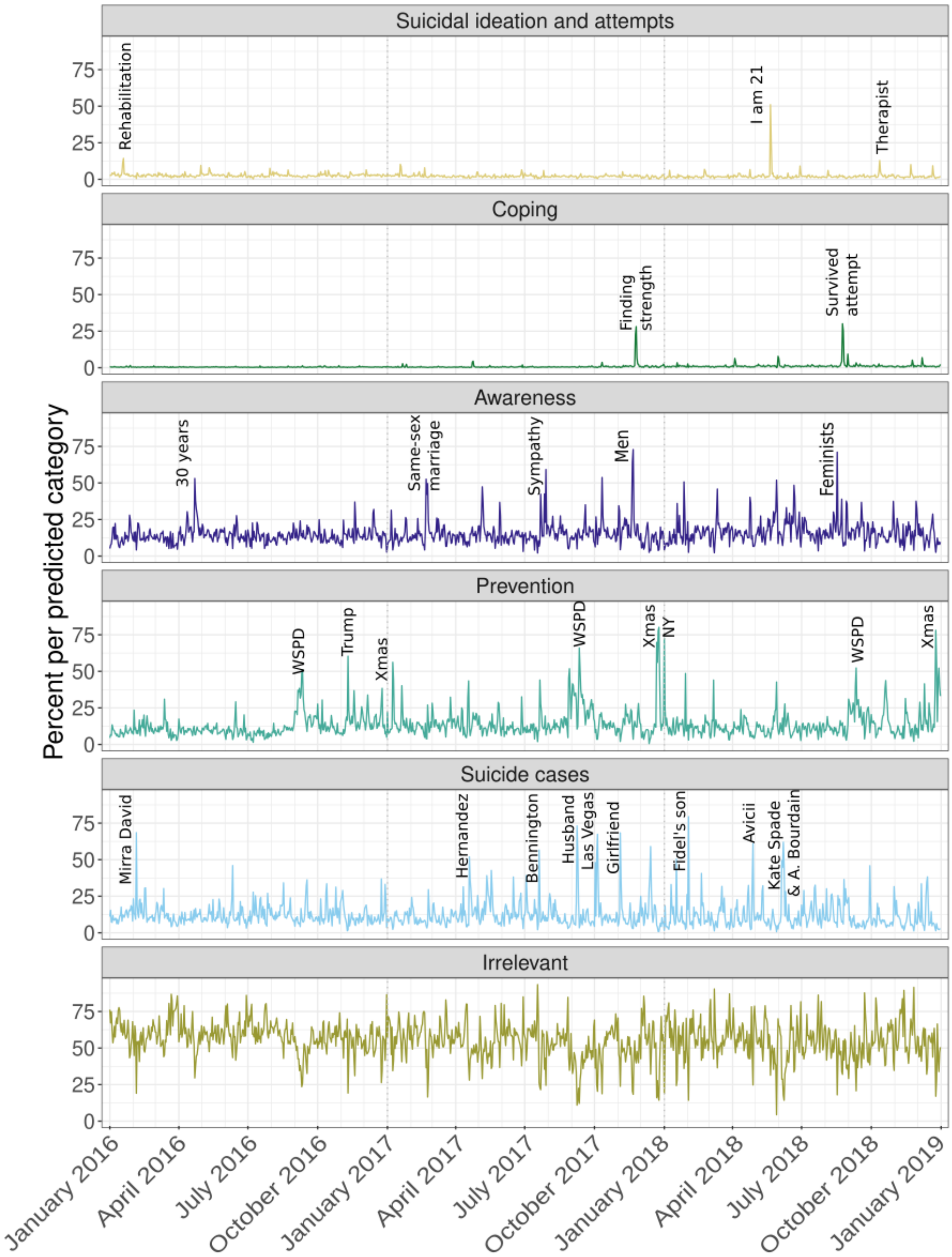
There was further a strong confusion between stories of suicidal ideation and a particular type of irrelevant tweet: tweets manually labeled as not serious or unclear if serious (dimension 3 in Multimedia Appendix 2). Of all tweets predicted to be personal stories of suicidal ideation or attempts, 10.8% (13/120) were not serious or were unclear, compared with only 1%-4% in the other predicted categories. Of all nonserious or unclear tweets, 45% (18/40) were correctly classified as irrelevant, 32% (13/40) were wrongly classified as suicidal compared with 3%-10% wrongly assigned to the other categories. The 13 (32%) nonserious tweets that were misclassified as suicidal included 4 exaggerations, 2 sarcastic remarks, 2 tweets with song lyrics or band names with the terms suicidal ideation or thoughts, 1 metaphoric use, and 2 statements about not being suicidal.

Face Validity Check With Daily Time Series Peaks per Category

Figure 5 illustrates the daily percentage of tweets in each predicted category in the prediction data set. As a face validity check, we identified the events that were mainly associated with each of the largest peaks in the time series of tweets per predicted category. These events are labeled with the following keywords in Figure 5: (1) *Suicidal ideation and attempts*: rehabilitation—4000 retweets of a rather sarcastic tweet from someone getting "punished" in rehabilitation because he said "fucking lit" that another patient was about to commit suicide; I am 21—a personal story retweeted 2200 times, describing someone's successful journey from a difficult childhood, through a long suicidal crisis, to a University degree and a full time job, falsely labeled as suicidal, as it describes not just a suicidal crisis, but also coping; therapist—900 retweets of someone reporting no suicidal thoughts as their therapist asked, although they have them very frequently. (2) C*oping*: finding strength—approximately 6000 retweets of a story of someone finding strength 3 weeks after a suicide attempt; survived attempt—a marine corps Veteran with posttraumatic stress disorder tweets about his survived suicide attempt, approximately 5000 retweets. (3) *Awareness*: sympathy—a tweet saying people who died by suicide need care while still alive, rather than sympathy when they are dead, retweeted 3500 times; men—7000 retweets of a tweet mentioning that suicide is the largest cause of young men; feminists—6000 retweets of a tweet saying that feminists who want equality should also consider that boys are double as likely to die from suicide than girls; 30 years—retweets and discussion of a federal data analysis results that suicide in the United States had risen to the highest levels in nearly 30 years; same-sex marriage—many tweets on research finding that suicide rates drop after legalization of same-sex marriage. (4) *Prevention*: Trump—increased calls to suicide hotline after Trump's election. (5) *Suicide cases* [44]: Aaron Hernandez (American football player); Chester Bennington (singer of Linkin Park); husband—retweets of tweet by a woman remembering her husband's suicide; Las Vegas—many retweets of a reply correcting a tweet by Trump, by stating that the shooter killed himself; girlfriend—many retweets of a tweet about a girlfriend who killed herself; Fidel's son—Fidel Castro Diaz-Balart; NY—new year; WSPD—World Suicide Prevention Day; and Xmas—Christmas.

For coping, prevention, and suicide case tweets, all highly frequent tweets were correctly classified tweets. Both highly shared coping tweets were from individuals who had survived a suicide attempt. Prevention peaks were related to the yearly World Suicide Prevention Day, to increased prevention efforts around Christmas and the New Year, and to increased lifeline calls after Trump's election [45]. All the identified peaks of tweets about suicide cases were related to actual instances of someone taking their own life. For awareness tweets, all but one peak were driven by actual awareness tweets. This single tweet (labeled same-sex marriage in Figure 3) was ambiguous, as it reported a research finding similar to a typical awareness tweet, but the finding was somewhat prevention related. Most awareness peaks were driven by tweets that cite a statistic about suicides. Of the tweets driving the 3 largest peaks in the suicidal category, only 1 was clearly suicidal ideation, another was a somewhat cynical tweet about someone else wanting to commit suicide, and a third was a clear confusion with an actual coping tweet. Thus, this face validity check reflects the high precision of the model for prevention, awareness, and coping tweets, as well as the lower performance for suicidal ideation and attempt tweets.

Figure 5. Daily percent of tweets per predicted category in the predictions data set (n=7.15 million). The daily value subsumes original and retweets per category. Key words for event peaks are explained in the main text.



Discussion

Overview of This Study

Owing to the effort required for manual annotation of texts, previous research on media and suicide prevention was limited by small sample sizes or by data sets put together using keyword search. Keywords either capture only a particular type of text (eg, containing celebrity names) or lump together a variety of different texts that contain broad search terms (eg, "suicide" [2,10-12]) In addition, research on the correlation of social media content with suicide cases in the population remains extremely scarce [9-13]. This study extends media research on suicide prevention by focusing on a broad range of suicide-related content on social media and by developing a reliable and efficient content labeling method based on machine learning, enabling fine-grained analysis of large data sets. We first developed a comprehensive annotation scheme for suicide-related content that includes new content types more typical on social than traditional media, such as personal stories of coping or suicidal ideation, or calls for action addressed at follower networks. On the basis of this systematic labeling scheme, we then tested the ability of different machine learning algorithms to distinguish 5 content types that seem particularly relevant based on previous research [2,4,6,30]. We further applied these methods to separate tweets about actual suicide, that is, in the meaning of someone taking their own life, from tweets that use the word suicide in some other way or context (binary classification). Our results for these 2 classification tasks show that machine learning methods, particularly deep learning models, achieve performances comparable with both human performance and with state-of-the-art methods in similar tasks [24,46].

This study is one of the first to automatically classify social media data other than suicidal ideation into categories relevant for suicide prevention. Only 3 studies, 2 (67%) of which used the same data set, have previously applied machine learning to distinguish specific types of social media posts other than suicidal ideation [9,24,25]. We extend these studies in several ways. Rather than classifying emotions in tweets about specific celebrities [9] or using a relatively small set of 816 tweets put together with a focus on suicidal ideation and celebrity names [24,25], we trained models to categorize any type of tweets containing suicide-related terms in a much larger data set than in previous studies. Furthermore, our larger data set enabled us to use deep learning models that can account for differences in the meaning of words across contexts, rather than only considering word frequencies. Finally, our annotation scheme introduced more fine-grained and particularly prevention-relevant categories. Specifically, it includes personal coping stories, for which some research on traditional media suggests preventive effects [4-6], and distinguishes awareness from prevention-focused tweets [27].

Principal Findings

Regarding the machine learning results, pretrained deep learning models fine-tuned to our data clearly outperformed a naïve majority classifier and a linear SVM classifier based on the word frequency representation TF-IDF. BERT and XLNet achieved F_1-scores of 0.70 and 0.71 in the 6-category classification and 0.83 and 0.81 in the binary about suicide versus off-topic classification in the test set. These scores were only slightly lower or even identical to those in the validation set, indicating good generalization to novel data. The clear advantage of deep learning models over TF-IDF and SVM suggests that there is crucial information about meaning in the context of words beyond what mere word frequencies can capture in tweets about suicide. Performance of the deep learning models was better than the more traditional approaches, but was very similar between BERT and XLNet. Advantages of XLNet over BERT include its ability to learn from long contexts and to consider dependencies between all words in the sentence. It seems that these advantages cannot be fully exploited given the limited number of words in tweets.

The 6 investigated tweet categories separated five important categories, including personal stories about either (1) coping or (2) suicidal ideation and attempts, calls for action that spread (3) problem awareness or (4) prevention-related information and (5) tweets about suicide cases, from other tweets (6) irrelevant to this categorization. The performance scores per category were nearly indistinguishable for BERT and XLNet. The model that performed better depended on the metric and the category and varied between model runs. BERT had slightly higher precision than XLNet for 2 important categories for a follow-up publication Niederkrotenthaler et al, unpublished data, May 2022; see Multimedia Appendix 1 for details) and was therefore chosen as the model to make predictions and test reliability. Although our data set included a much broader set of tweets than previous studies focusing on similar prevention-related tweet categories [24,25], our machine learning performances were comparable or better than in these previous studies, with the exception of the suicidal ideation category.

In general, BERT and XLNet were better at classifying tweets that are also easier to distinguish for humans, including more homogeneous classes such as prevention and suicide cases. These often included similar keywords, such as prevention, hotline, lifeline, or the phrase “committed/commits suicide” (see word clouds in Multimedia Appendix 1). For these categories, BERT performance was very similar to the human interrater performances. BERT and human performance were also comparable for coping stories. The model’s performance was lower only for more subjective classes such as suicidal ideation and attempt stories. Error analysis suggests confusions with sarcastic, joking, exaggerated, or metaphorical uses of the word suicide as one part of the explanation. Such nonserious messages are difficult to distinguish from genuine suicidal ideation for both the model and humans. The gap between human and model performance is the largest for suicidal tweets, suggesting that this distinction is even more difficult for the model. Both humans and the model missed many ambiguous expressions of suicidal ideation (low recall). In contrast, between-human precision is much higher than the model’s precision. This shows that there are many suicidal ideation tweets that humans can clearly identify, whereas model reliability can still be improved for these types of tweets.

The analysis of the most common model errors demonstrates that the mistakes were mostly not trivial. Most confusions of suicidal and coping tweets by the model were tweets in which the feelings of the tweet author were quite ambiguous. This suggests that much higher performance scores are difficult to achieve for these personal stories about suicide. Nonetheless, for tweets about suicidal ideation in the past, which implicitly express that coping occurred, there may be room for improvement through adding more training examples. Furthermore, the error analysis suggests possible improvements for tweets regarding prevention and suicide cases. In contrast, the model actually helped detect errors by the human coder for awareness tweets labeled as prevention.

When distinguishing tweets about actual suicide from off-topic tweets, the model achieved excellent performance scores, particularly for tweets about actual suicide, with no difference between the 2 deep learning models. In other words, tweets labeled as “about suicide” are reliably actual tweets about suicide, and most such tweets are detected by the models. Thus, the use of any of these models for future research is very promising.

Using the final BERT models for both classification tasks, we estimated the percentage of tweets per category out of all suicide-related tweets in the United States from 2016 to 2018. Overall, approximately 6% were personal stories of concerned individuals, with approximately 5% on suicidal ideation or attempts and approximately 1% on coping stories. Estimates for awareness, prevention, and suicide case tweets were approximately 22%, 16%, and 16% of tweets, respectively. We plotted the daily volume per tweet category and investigated tweets during peaks in the time series. Most of these tweets were correctly classified by our models, and peaks often coincided with events matching a particular category (eg, the World Suicide Prevention Day or a celebrity suicide), which highlights the face validity of our model predictions. Finally, approximately three-fourths of all suicide-related tweets actually referred to someone taking their own life, whereas the rest used the term in another meaning or context (eg, euthanasia, suicide bombers, jokes, metaphors, and exaggerations).

Limitations and Future Work

Despite our data set being more comprehensive than any existing data set on the topic, one of the limitations of our study is the size of the training data set, which is crucial for training deep learning models. In particular, this concerns the rarer categories that we have not yet used for machine learning in this study (eg, bereaved experiences and lives saved). The data set could further benefit from adding more examples of coping messages that describe suicidal ideation and behavior in the past, thereby implicitly indicating coping (see category definition in Multimedia Appendix 2). Furthermore, some tweets in the categories suicide other and off-topic might warrant to be investigated separately, given recent findings of the possible protective effects of flippant remarks and humor or negative portrayals of suicide in the form of murder-suicides [11]. Similarly, the suicide case category may warrant being separated into suicide news and condolence messages, which may have protective effects [9], and tweets about suicide cases may warrant filtering out those about celebrities [2]. Higher classification performance for the category suicidal ideation in the study by Burnap et al [24] showed that a focus on this category during data collection could improve our model.

Finally, a number of limitations apply to automated text analyses, such as machine learning. First, there are no traces of images, videos, or content of the URLs shared in the text of tweets, although this additional information can crucially affect the meaning of a tweet. Second, some things are only implicitly expressed or very subjective, and thus difficult to capture with such methods, but also difficult to reliably recognize for humans. For instance, it is difficult to clearly differentiate coping from suffering, even for humans who have some knowledge about how such experiences look like in the real world. It is even more difficult to capture such subjective experiences using word frequencies. Deep learning models such as BERT and XLNet, having been trained on huge amounts of text produced by humans, may be able to capture some of these nuances but require large amounts of training examples. Third, a machine learning model can only recognize example tweets that are sufficiently similar to the examples in the training set and only predict the predefined categories. In contrast, a human coder might recognize new ways of expressing the same meaning or the need to introduce a new category. We partially addressed the latter limitation through an extensive labeling process, ensuring that we captured all typical message categories by including a random set of tweets. Nonetheless, including more and different examples for suicidal ideation and coping stories in future studies would likely improve model performance.

Conclusions and Practical Implications

The field of media and suicide research has only recently begun to evolve to consider social media content as relevant in the assessment of media effects. This study makes 2 major contributions to this field. First, it provides a systematic overview of different content types that are common on social media, which may be useful as a content labeling scheme for future research on the topic. Some of the categories identified have been found to be relevant to suicide prevention, particularly in other media types. For social media content, these associations with indicators of behavior, particularly suicidal behaviors and help-seeking, remain to be tested accordingly. Second, the machine learning methods enable researchers to assess large amounts of social media data and subsequently correlate it with available behavioral data of interest; for example, suicides or help-seeking data. In this way, this work enables systematic large-scale investigations of associations between these behaviors and fine-grained message characteristics of social media posts (eg, Niederkrotenthaler et al, unpublished data, May 2022; see Multimedia Appendix 1 for details). Such large-scale investigations will contribute to accumulating robust evidence on which characteristics are actually harmful and protective. Furthermore, future applications of the developed models might include the screening of social media content to detect other types of content associated with suicide cases that have not been described in previous research. The classification performances of the developed models demonstrate the strong potential of machine learning, particularly deep learning, for media suicide effects research.

Acknowledgments

This study was supported by Project V!brant Emotional Health grant “Suicide Prevention media campaign Oregon” to TN and a grant from the Vienna Science and Technology Fund to DG (grant VRG16-005). The authors would like to thank Konstantin Hebenstreit for assistance with data collection and cleaning.

Data and Code Availability

The data sets generated and analyzed during this study are available on GitHub. The code and data for the statistical analyses and figures (intercoder reliability and frequency of categories) as well as CIs for machine learning are available on the web [47]. The code for training the machine learning models is available on the web [48]. Raw data only include tweet IDs, not tweet text, to protect the identity of the tweet authors. Tweets can be redownloaded via the Twitter API using these IDs.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary information with additional details on methods, results tables, search term lists, and word clouds.

[PDF File (Adobe PDF File), 2543 KB - jmir_v24i8e34705_app1.pdf]

Multimedia Appendix 2

Annotation scheme.

[PDF File (Adobe PDF File), 92 KB - jmir_v24i8e34705_app2.pdf]

References

1. Ritchie H, Roser M, Ortiz-Ospina E. Suicide. Our World in Data. 2015. URL: https://ourworldindata.org/suicide [accessed 2022-05-05]
2. Niederkrotenthaler T, Braun M, Pirkis J, Till B, Stack S, Sinyor M, et al. Association between suicide reporting in the media and suicide: systematic review and meta-analysis. BMJ 2020 Mar 18;368:m575 [FREE Full text] [doi: 10.1136/bmj.m575] [Medline: 32188637]
3. Phillips DP. The influence of suggestion on suicide: substantive and theoretical implications of the Werther effect. Am Sociol Rev 1974;39(3):340-354. [Medline: 11630757]
4. Niederkrotenthaler T, Voracek M, Herberth A, Till B, Strauss M, Etzersdorfer E, et al. Role of media reports in completed and prevented suicide: Werther v. Papageno effects. Br J Psychiatry 2010 Sep;197(3):234-243. [doi: 10.1192/bjp.bp.109.074633] [Medline: 20807970]
5. Niederkrotenthaler T, Till B. Effects of suicide awareness materials on individuals with recent suicidal ideation or attempt: online randomised controlled trial. Br J Psychiatry 2020 Dec;217(6):693-700. [doi: 10.1192/bjp.2019.259] [Medline: 31843026]
6. Sinyor M, Schaffer A, Nishikawa Y, Redelmeier DA, Niederkrotenthaler T, Sareen J, et al. The association between suicide deaths and putatively harmful and protective factors in media reports. CMAJ 2018 Jul 30;190(30):E900-E907 [FREE Full text] [doi: 10.1503/cmaj.170698] [Medline: 30061324]
7. Till B, Tran US, Voracek M, Niederkrotenthaler T. Beneficial and harmful effects of educative suicide prevention websites: randomised controlled trial exploring Papageno v. Werther effects. Br J Psychiatry 2017 Aug;211(2):109-115. [doi: 10.1192/bjp.bp.115.177394] [Medline: 28522433]
8. Till B, Arendt F, Scherr S, Niederkrotenthaler T. Effect of educative suicide prevention news articles featuring experts with vs without personal experience of suicidal ideation: a randomized controlled trial of the Papageno effect. J Clin Psychiatry 2018 Nov 20;80(1):17m11975. [doi: 10.4088/JCP.17m11975] [Medline: 30549483]
9. Fahey RA, Matsubayashi T, Ueda M. Tracking the Werther Effect on social media: emotional responses to prominent suicide deaths on twitter and subsequent increases in suicide. Soc Sci Med 2018 Dec;219:19-29 [FREE Full text] [doi: 10.1016/j.socscimed.2018.10.004] [Medline: 30342383]
10. Sinyor M, Williams M, Zaheer R, Loureiro R, Pirkis J, Heisel MJ, et al. The relationship between suicide-related twitter events and suicides in Ontario from 2015 to 2016. Crisis 2021 Jan;42(1):40-47 [FREE Full text] [doi: 10.1027/0227-5910/a000684] [Medline: 32366171]
11. Sinyor M, Williams M, Zaheer R, Loureiro R, Pirkis J, Heisel MJ, et al. The association between Twitter content and suicide. Aust N Z J Psychiatry 2021 Mar;55(3):268-276. [doi: 10.1177/0004867420969805] [Medline: 33153274]
12. Ueda M, Mori K, Matsubayashi T, Sawada Y. Tweeting celebrity suicides: users' reaction to prominent suicide deaths on Twitter and subsequent increases in actual suicides. Soc Sci Med 2017 Sep;189:158-166 [FREE Full text] [doi: 10.1016/j.socscimed.2017.06.032] [Medline: 28705550]
13. Won HH, Myung W, Song GY, Lee WH, Kim JW, Carroll BJ, et al. Predicting national suicide numbers with social media data. PLoS One 2013 Apr 22;8(4):e61809 [FREE Full text] [doi: 10.1371/journal.pone.0061809] [Medline: 23630615]
14. Colombo GB, Burnap P, Hodorog A, Scourfield J. Analysing the connectivity and communication of suicidal users on Twitter. Comput Commun 2016 Jan 01;73(Pt B):291-300 [FREE Full text] [doi: 10.1016/j.comcom.2015.07.018] [Medline: 26973360]
15. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:2098-2110 [FREE Full text] [doi: 10.1145/2858036.2858207] [Medline: 29082385]
16. Liu X, Liu X, Sun J, Yu NX, Sun B, Li Q, et al. Proactive Suicide Prevention Online (PSPO): machine identification and crisis management for Chinese social media users with suicidal thoughts and behaviors. J Med Internet Res 2019 May 08;21(5):e11705 [FREE Full text] [doi: 10.2196/11705] [Medline: 31344675]
17. Ophir Y, Tikochinski R, Asterhan CS, Sisso I, Reichart R. Deep neural networks detect suicide risk from textual Facebook posts. Sci Rep 2020 Oct 07;10(1):16685 [FREE Full text] [doi: 10.1038/s41598-020-73917-0] [Medline: 33028921]

18. Sawhney R, Manchanda P, Mathur P, Shah R, Singh R. Exploring and learning suicidal ideation connotations on social media with deep learning. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2018 Presented at: WASSA '18; October 31-November 1, 2018; Brussels, Belgium p. 167-175. [doi: 10.18653/v1/w18-6223]
19. Bernert RA, Hilberg AM, Melia R, Kim JP, Shah NH, Abnousi F. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. Int J Environ Res Public Health 2020 Aug 15;17(16):5929 [FREE Full text] [doi: 10.3390/ijerph17165929] [Medline: 32824149]
20. Castillo-Sánchez G, Marques G, Dorronzoro E, Rivera-Romero O, Franco-Martín M, De la Torre-Díez I. Suicide risk assessment using machine learning and social networks: a scoping review. J Med Syst 2020 Nov 09;44(12):205 [FREE Full text] [doi: 10.1007/s10916-020-01669-5] [Medline: 33165729]
21. 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]
22. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
23. Yin Z, Sulieman L, Malin B. A systematic literature review of machine learning in online personal health data. J Am Med Inform Assoc 2019 Jun 01;26(6):561-576 [FREE Full text] [doi: 10.1093/jamia/ocz009] [Medline: 30908576]
24. Burnap P, Colombo G, Amery R, Hodorog A, Scourfield J. Multi-class machine classification of suicide-related communication on Twitter. Online Soc Netw Media 2017 Aug;2:32-44 [FREE Full text] [doi: 10.1016/j.osnem.2017.08.001] [Medline: 29278258]
25. Chiroma F, Liu H, Cocea M. Suicide related text classification with prism algorithm. In: Proceedings of the 2018 International Conference on Machine Learning and Cybernetics. 2018 Presented at: ICMLC '18; July 15-18, 2018; Chengdu, China p. 575-580. [doi: 10.1109/icmlc.2018.8527032]
26. Shearer E, Matsa KE. News Use Across Social Media Platforms 2018. Pew Research Center. 2018 Sep 10. URL: https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/ [accessed 2021-02-17]
27. Till B, Tran US, Niederkrotenthaler T. The impact of educative news articles about suicide prevention: a randomized controlled trial. Health Commun 2021 Dec;36(14):2022-2029. [doi: 10.1080/10410236.2020.1813953] [Medline: 32867541]
28. Niederkrotenthaler T, Reidenberg DJ, Till B, Gould MS. Increasing help-seeking and referrals for individuals at risk for suicide by decreasing stigma: the role of mass media. Am J Prev Med 2014 Sep;47(3 Suppl 2):S235-S243. [doi: 10.1016/j.amepre.2014.06.010] [Medline: 25145745]
29. Sumida N, Walker M, Mitchell A. 3. The role of social media in news. Pew Research Center. 2019 Apr 23. URL: https://www.journalism.org/2019/04/23/the-role-of-social-media-in-news/ [accessed 2021-02-18]
30. Pirkis JE, Burgess PM, Francis C, Blood RW, Jolley DJ. The relationship between media reporting of suicide and actual suicide in Australia. Soc Sci Med 2006 Jun;62(11):2874-2886. [doi: 10.1016/j.socscimed.2005.11.033] [Medline: 16387400]
31. King G, Lam P, Roberts ME. Computer-assisted keyword and document set discovery from unstructured text. Am J Polit Sci 2017 Apr 03;61(4):971-988. [doi: 10.1111/ajps.12291]
32. Signorell A. DescTools: Tools for Descriptive Statistics. The Comprehensive R Archive Network. 2022 May 9. URL: https://cran.r-project.org/package=DescTools [accessed 2021-11-01]
33. Kuhn M. caret: Classification and Regression Training. The Comprehensive R Archive Network. 2022 Apr 19. URL: https://CRAN.R-project.org/package=caret [accessed 2021-11-01]
34. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. J Open Source Softw 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
35. Maiya AS. ktrain: a low-code library for augmented machine learning. J Mach Learn Res 2022 May;23:1-6 [FREE Full text]
36. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. TensorFlow. 2015. URL: https://www.tensorflow.org/ [accessed 2021-10-09]
37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
38. Ma G. Tweets Classification with BERT in the Field of Disaster Management. Department of Civil Engineering, Stanford University. 2019. URL: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/custom/15785631.pdf [accessed 2021-09-07]
39. Preoţiuc-Pietro D, Schwartz HA, Park G, Eichstaedt J, Kern M, Ungar L, et al. Modelling valence and arousal in Facebook posts. In: Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2016 Presented at: WASSA '16; June 16, 2016; San Diego, CA, USA p. 9-15. [doi: 10.18653/v1/w16-0404]
40. Aizawa A. An information-theoretic perspective of tf–idf measures. Inf Process Manag 2003 Jan;39(1):45-65. [doi: 10.1016/S0306-4573(02)00021-3]
41. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL-HLT '19; June 2-7, 2019; Minneapolis, MN, USA p. 4171-4186 URL: https://aclanthology.org/N19-1423.pdf

42. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le QV. XLNet: generalized autoregressive pretraining for language understanding. In: Proceedings of the Advances in Neural Information Processing Systems 32. 2019 Presented at: NeurIPS '19; December 8-14, 2019; Vancouver, Canada URL: https://papers.nips.cc/paper/2019/hash/dc6a7e655d7e5840e66733e9ee67cc69-Abstract.html
43. Dai Z, Yang Z, Yang Y, Carbonell J, Le QV, Salakhutdinov R. Transformer-XL: attentive language models beyond a fixed-length context. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.: Association for Computational Linguistics; 2019 Jun 2 Presented at: ACL '19; July 28-August 2, 2019; Florence, Italy p. 2978-2988 URL: http://arxiv.org/abs/1901.02860 [doi: 10.18653/v1/p19-1285]
44. List of suicides. Wikipedia - The Free Encyclopedia. 2021. URL: https://en.wikipedia.org/w/index.php?title=List_of_suicides&oldid=1053465646 [accessed 2021-11-04]
45. Walters J. Suicide hotline calls reached record high as Trump victory became clear. The Guardian. 2016 Nov 12. URL: https://www.theguardian.com/us-news/2016/nov/12/suicide-prevention-hotline-record-calls-donald-trump-victory-lgbt [accessed 2021-11-04]
46. Barbieri F, Camacho-Collados J, Neves L, Espinosa-Anke L. TweetEval: unified benchmark and comparative evaluation for tweet classification. In: Conference on Empirical Methods in Natural Language Processing.: Association for Computational Linguistics; 2020 Oct 26 Presented at: EMNLP '20; November 16-20, 2020; Virtual p. 1644-1650 URL: https://aclanthology.org/2020.findings-emnlp.148/ [doi: 10.18653/v1/2020.findings-emnlp.148]
47. Metzler H. TwitterSuicideR. GitHub. URL: https://github.com/hannahmetzler/TwitterSuicideR [accessed 2022-08-05]
48. Metzler H, Baginski H. TwitterSuicideML. GitHub. URL: https://github.com/HubertBaginski/TwitterSuicideML [accessed 2022-08-05]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
LR: learning rate
SVM: support vector machine
TF-IDF: term frequency−inverse document frequency

Edited by R Kukafka; submitted 11.12.21; peer-reviewed by E Malova, A Teles, Z Xie; comments to author 04.04.22; revised version received 02.06.22; accepted 08.06.22; published 17.08.22.

Please cite as:
Metzler H, Baginski H, Niederkrotenthaler T, Garcia D
Detecting Potentially Harmful and Protective Suicide-Related Content on Twitter: Machine Learning Approach
J Med Internet Res 2022;24(8):e34705
URL: https://www.jmir.org/2022/8/e34705
doi:10.2196/34705
PMID:35976193

Original Paper

Increased Online Aggression During COVID-19 Lockdowns: Two-Stage Study of Deep Text Mining and Difference-in-Differences Analysis

Jerome Tze-Hou Hsu[1,2]; Richard Tzong-Han Tsai[1,3], PhD

[1]Center for Geographic Information Science, Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan
[2]Taipei Municipal Jianguo High School, Taipei, Taiwan
[3]Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan

Corresponding Author:
Richard Tzong-Han Tsai, PhD
Center for Geographic Information Science
Research Center for Humanities and Social Sciences
Academia Sinica
128 Academia Road, Section 2
Nankang
Taipei, 11529
Taiwan
Phone: 886 2 27898150
Fax: 886 2 27894160
Email: rthtsai@gate.sinica.edu.tw

Abstract

Background: The COVID-19 pandemic caused a critical public health crisis worldwide, and policymakers are using lockdowns to control the virus. However, there has been a noticeable increase in aggressive social behaviors that threaten social stability. Lockdown measures might negatively affect mental health and lead to an increase in aggressive emotions. Discovering the relationship between lockdown and increased aggression is crucial for formulating appropriate policies that address these adverse societal effects. We applied natural language processing (NLP) technology to internet data, so as to investigate the social and emotional impacts of lockdowns.

Objective: This research aimed to understand the relationship between lockdown and increased aggression using NLP technology to analyze the following 3 kinds of aggressive emotions: anger, offensive language, and hate speech, in spatiotemporal ranges of tweets in the United States.

Methods: We conducted a longitudinal internet study of 11,455 Twitter users by analyzing aggressive emotions in 1,281,362 tweets they posted from 2019 to 2020. We selected 3 common aggressive emotions (anger, offensive language, and hate speech) on the internet as the subject of analysis. To detect the emotions in the tweets, we trained a Bidirectional Encoder Representations from Transformers (BERT) model to analyze the percentage of aggressive tweets in every state and every week. Then, we used the difference-in-differences estimation to measure the impact of lockdown status on increasing aggressive tweets. Since most other independent factors that might affect the results, such as seasonal and regional factors, have been ruled out by time and state fixed effects, a significant result in this difference-in-differences analysis can not only indicate a concrete positive correlation but also point to a causal relationship.

Results: In the first 6 months of lockdown in 2020, aggression levels in all users increased compared to the same period in 2019. Notably, users under lockdown demonstrated greater levels of aggression than those not under lockdown. Our difference-in-differences estimation discovered a statistically significant positive correlation between lockdown and increased aggression (anger: P=.002, offensive language: P<.001, hate speech: P=.005). It can be inferred from such results that there exist causal relations.

Conclusions: Understanding the relationship between lockdown and aggression can help policymakers address the personal and societal impacts of lockdown. Applying NLP technology and using big data on social media can provide crucial and timely information for this effort.

(J Med Internet Res 2022;24(8):e38776) doi:10.2196/38776

KEYWORDS

natural language processing; lockdown; online aggression; infoveillance; causal relationship; social media; neural networks; computer; pandemic; COVID-19; emotions; internet; sentiment analysis; Twitter; content analysis; infodemiology

Introduction

Background

On March 13, 2020, the United States declared a state of emergency in response to the COVID-19 pandemic. Many states imposed lockdown measures to slow down the spread of the virus. However, lockdown (stay-at-home) policies affect many aspects of human life. The frustration and loneliness people experience under extended periods of confinement may predictably have negative psychological impacts [1-3]. Furthermore, frustration can manifest itself through increased aggressiveness [4]. In a time when people live closely beside intimate family members, emotional problems, such as suicidal thoughts and aggressiveness, may lead to destructive behaviors and have an immediate impact on society [5,6]. Whether scientific investigations corroborate such observations can have significant policy implications for public or private governance. Unsurprisingly, the relationship between lockdown and adverse psychological effects has attracted increasing attention from multiple disciplines of researchers. However, there have been few robust tests of the causal relationship between lockdown and aggressive emotions. This research used machine learning to produce robust data. Then, we used a statistical difference-in-differences analysis to estimate the causal relationship between lockdown and increased online aggression. The application of machine learning technologies in social science research can provide new information in a much broader scope at a much higher speed.

Related Works

Negative Impacts of Lockdown

At the individual level, studies have shown that lockdown is associated with suicidal ideation, anxiety disorder, nightmares, depression, loneliness, and poor mental health [7-12]. At the societal level, a lockdown's adverse effects are manifested through significant increases in divorces, sexual violence [13], and domestic violence [14]. All these effects pose considerable threats to the stability and well-being of individuals and society. Therefore, it is an urgent task to understand these harmful actions under COVID-19 lockdowns.

Other research in psychology has focused on the deterioration of mental health before and under lockdown [7]. The authors observed an increase in certain health behaviors 1 month into lockdown by comparing prelockdown and postlockdown survey data. However, the authors did not analyze the causal relationship between lockdown and these behaviors.

Emotion is one of the main drivers of human action. It is reasonable that a more aggressive state of mind leads to aggressive behaviors like domestic and sexual violence. The influential frustration-aggression theory [15] suggests that aggressive behavior results from frustration caused by thwarting individual goals. In the early months of the pandemic, lockdown led to many canceled plans and unaccomplished goals. Therefore, a causal relationship between lockdown and increased aggression is reasonable.

Notably, a study investigated the correlation between lockdown and increased aggression [2]. Killgore et al conducted a questionnaire survey [2]. They used the Buss Perry Aggression Questionnaire to measure aggression levels in patients under and not under lockdown during the initial months of the pandemic in the United States. They found a statistically significant increase in the following 4 kinds of aggression between lockdown and nonlockdown groups: physical aggression, verbal aggression, anger, and hostility. However, owing to limitations in traditional questionnaire methods, such as the lack of data before the pandemic, this research could not assess the causal relationship between lockdown and increased aggression. Furthermore, because of practical limitations in survey administration, the authors had to survey different participants in every sampling, which provides an additional source of uncertainty.

COVID-19 Twitter Sentiment Analysis

Online analysis of tweets using natural language processing (NLP) has provided valuable information in health-related research. General sentiment analysis has been performed to examine people's emotions under lockdown [16,17].

Some are related to specific topics, such as vaccination [18-20], while others are related to specific regions or countries [21]. However, most studies did not analyze the relationship between lockdown and emotions to the best of our knowledge. Su et al [22] analyzed the psycholinguistic features in 2 different cities going into lockdown. While this can capture specific rising trends in tweet words, the lexicon frequency analysis method does not capture each word's context. Thus, it cannot predict emotions as accurately as neural network models [23].

Our Study

This paper addresses the weaknesses of the current psychology and NLP research. Most of the recent literature in psychology has not offered meaningful evidence to the causality between lockdown and aggressive emotions. On the other hand, current NLP research in sentiment analysis mainly focuses on optimizing methods for machines to capture emotions in vast volumes of digitalized human discourse. However, the current body of NLP literature in the public health area rarely probes into causal relations of social phenomena.

We used new NLP technology to conduct a virtual longitudinal study of online Twitter users and their tweets to investigate the impacts of lockdowns on the following 3 kinds of aggression: anger, offensive language (offensive), and hate speech (hate). Our infoveillance method allows us to discover trends in aggression levels that can provide important information for policy makers and health professionals. Moreover, data before and after lockdown allows us to estimate the potential causal relationship between lockdown and increased aggression using the difference-in-differences analysis, an established econometric method to understand the causal relationship in nonexperimental time-series data [24]. This interdisciplinary method yields robust results in understanding the relationship between lockdown and increased aggression, and it opens up new potentials for applying NLP and internet technology to support medical research.

Methods

Overview

First, we sampled a group of Twitter users across the United States as our subjects of analysis. Then, we used Twitter's application programming interface to obtain all the tweets the sampled users posted between January 1, 2019, and October 1, 2020. Our objective was to use a neural network model to detect different levels of aggressive emotions during different periods in these tweets. We selected the Bidirectional Encoder Representations from Transformers (BERT) model, a state-of-the-art language model that can understand the meanings of emotions through contexts and nuances better than previous lexicon-based models [23,25]. To train the BERT model for emotion detection, we collected training data based on established data sets [26-28]. Afterward, we classified the tweet emotions using our trained BERT model. We conducted an observational analysis to compare and contrast aggression levels between different geographies, lockdown statuses, and times. After observing an increase in aggression after lockdown, we measured the relationship between lockdown status and aggression levels using the Poisson regression as a difference-in-differences estimation.

Twitter Data Sampling

In the United States, states retain the power to implement lockdown policies. Therefore, this research used the state lockdown status to determine whether an individual user was under lockdown at a particular time. We randomly sampled Twitter users geographically tagged with the states in the United States as our longitudinal internet study participants. After sampling the users, we sampled every tweet they posted in 2019 and the first 6 months after President Trump declared a national emergency in March 2020. Our sampling yielded a spatiotemporal data set of 1,281,362 tweets posted from January 1, 2019, to October 1, 2020, by 11,455 Twitter users. The sampled users came from all across the United States, including users from all 50 states. In this study, we used these tweets to investigate the relationship between lockdown and social media aggressiveness. All tweets followed a data preprocessing protocol [29] before being analyzed by the BERT model for emotion detection.

Training Data Collection

In order to detect aggressive emotions in the tweets, we trained a BERT neural network binary classification model for each of the 3 aggressive emotions. For each model, we collected different training data sets. Our definition of each emotion is identical to that of the training data set. Table 1 contains the definitions for the 3 aggressive emotions, with Table 2 providing sample text for each.

Table 1. Definition for each emotion.

Aggressive emotion	Definition
Anger	A strong feeling of displeasure or antagonism [26]
Offensive language	Speech that contains unacceptable language (profanity) and is potentially harmful to a disadvantaged group [27]
Hate speech	Language that expresses hatred toward a targeted group and is intended to be derogatory, insulting, and humiliating [27,28]

Table 2. Sample text containing each aggressive emotion.

Aggressive emotion	Sample text
Anger	I hope this all ends soon. This is hell
Offensive language	Are people really this stupid?
Hate speech	@user The rot starts from the top......Trumps wankers are all racist......F*ck them all!

Anger

We selected the GoEmotions data set to train the anger classification model [26]. It is one of the largest manually annotated data sets of 58,000 English Reddit comments. In the data set, each sentence is annotated to identify the presence of 28 relatively common emotions. To train the BERT model for binary classification, we selected the 6000 sentences that contain anger and a random sample of 6000 other sentences that do not contain anger. We selected 1000 other annotated comments for testing. Based on GoEmotions, anger is defined as "A strong feeling of displeasure or antagonism."

Offensive Language

Offensive language is speech that contains unacceptable language (profanity) and is potentially harmful to a disadvantaged group. We selected the "Automated Hate Speech Detection and the Problem of Offensive Language" (AHSD) [27] data set as our training data set, which contains 24,802 human-labeled tweets. We randomly sampled 7750 sentences for training and 613 for testing. The study distinguishes hate speech with real harmful intentions from general offensive lexicons. For example, many teenagers often use terms like f*ck and b*tch in a casual manner that does not intend harm. AHSD provides annotated data for offensive language and more harmful hate speech.

Hate Speech

Unlike general offensive language, hate speech is a more specific language that causes intentional harm. To train our hate model, we merged the AHSD data set, as mentioned earlier, with the Large Scale Crowd Sourcing [28] data set, which provides an additional 2067 tweets labeled hateful by humans. We randomly sampled 6450 sentences for training and 639 for testing. Both of these data sets identify hate speech as the language that expresses hatred toward a targeted group and is intended to be derogatory, insulting, and humiliating. This definition has been widely used in previous research [30-32].

Model Training

We used the BERT model [29] to identify emotions in tweets. This pretrained neural network model is one of the most powerful models in emotion understanding. With its abundant pretraining data from the entire English Wikipedia, the model already had a basic understanding of the English language before we conducted the final fine-tuning. The model's contextual embedding allows it to understand words regarding context, taking its language understanding ability beyond traditional lexicon analysis. Our model architecture is constructed under python modules pytorch 1.8.1 and transformer 4.11.0. Using the training data, we obtained great performing models for all 3 of our target emotions (specific statistics are shown below).

Model Evaluation

We first tested our model predictions on the testing set (train-test split). The results are shown in Table 3, along with the confusion matrices in Figure 1. Then, we tested our model on the sampled Twitter data set used for further analysis. To evaluate the model performance on our self-sampled Twitter data, we selected 1080 tweets, 540 from people under lockdown and 540 from people not under lockdown, with 5-6 tweets randomly selected from each week. Then, 2 native English speakers annotated the tweets based on the definition for each emotion above. The Cohen kappa values between annotators and our model's performance are reported in Tables 4 and 5, respectively. Through this, we can validate our model competence on Twitter data used in further down-stream analysis.

Table 3. Model performance on the testing set.

Model	Precision	Recall	F1
Anger	0.869	0.826	0.847
Offensive language	0.953	0.988	0.970
Hate speech	0.956	0.920	0.933

Figure 1. Confusion matrices for our models: anger (A), offensive (B), and hate (C). The bottom-right and top-left quadrants are where the models predicted correctly, which represent true negatives and true positives, respectively. A darker quadrant color indicates greater prediction.

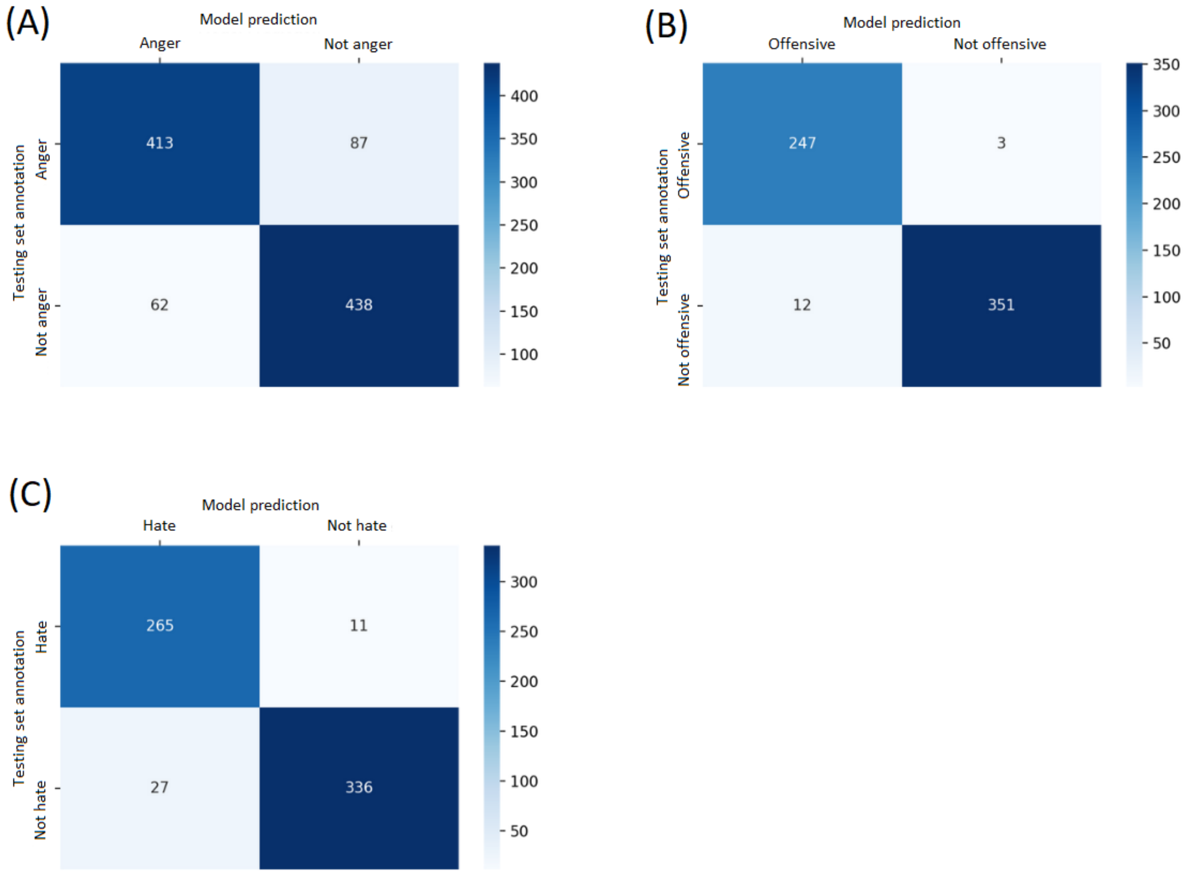


Table 4. Cohen kappa interrater agreement between the raters.

Emotion	Kappa
Anger	0.928
Offensive language	0.937
Hate speech	0.890

Table 5. Model performance on the sampled Twitter data set.

Model	Precision	Recall	F1
Anger	0.795	0.888	0.839
Offensive language	0.843	0.922	0.880
Hate speech	0.810	0.872	0.839

Data Analysis Methods

Overview

To understand aggression levels in tweets, we measured the proportion of tweets that contain aggression among all randomly sampled tweets. First, we used separate BERT models for each aggressive emotion (ie, anger, offensive, and hate) to analyze our sampled Twitter data. The analysis resulted in 3 data sets, one for each aggressive emotion. In each data set, for each of the 50 states, we calculated the percentage of tweets containing the aggressive emotion for the 92 weeks from January 2019 to October 2020. This analysis resulted in 3 data sets with 4600 data points each.

Although our data consist of aggressive tweet counts in different spatiotemporal settings, we analyzed the proportion of aggressive tweets among total tweets, rather than the count of aggressive tweets, to investigate the aggression level on Twitter. This is because an increase in aggressive tweet counts may be due to an increase in total tweets posted, which does not necessarily indicate a higher level of aggression. Measuring the proportion of aggressive tweets more accurately depicts the aggression level on Twitter.

Over 3 stages of observation and analysis, we looked at the data from different perspectives. In the first part, we compared aggression levels between groups of different lockdown statuses in the first 6 months of the pandemic. After that, we focused on the states that had undergone lockdown, and we looked at their aggression levels before and after lockdown. Finally, we used difference-in-differences analysis to estimate the impact of lockdown on the increase of aggression.

Observing the Difference in Aggression Levels Between Groups Under and Not Under Lockdown

To understand the impact of lockdown on aggressive emotions, we investigated the aggressive tweet proportions for each specific time and location, and compared the proportions in people under and not under lockdown. For that purpose, we designed our first objective. For each week in the 92 weeks from January 1, 2019, to October 1, 2020, we separated states under lockdown from those not under lockdown into 2 groups. Then, we separately aggregated the number of aggressive tweets and total tweets. We calculated the aggressive tweet percentage for each of the 2 groups every week for 92 weeks based on the combined data. Note that the users in each state represented the patients under lockdown and not under lockdown based on the state's current lockdown status.

Observing Aggression Trends in States That Had Undergone Lockdown From the Weeks Before and After Lockdown

In the previous section, we observed and compared aggressive emotions between groups under and not under lockdown. In this section, we focused on understanding the trends in states that had undergone lockdown. More specifically, we looked at the increase in aggression after lockdown by comparing data before and after lockdown. We chronologically aligned the data in each lockdown state based on the initial week of lockdown. More specifically, for every state that had ever undergone lockdown, the week that lockdown started was denoted as week 0. Other weeks were numbered accordingly (ie, the first week after week 0 was week 1, the week before week 0 was week −1, and so forth). Using this method, we visualized the increase in aggression after the lockdown. Note that this was solely an observation of aggression trends before and after lockdown. It did not measure the net impact of lockdown status on aggression levels. To specifically measure the impact and investigate the causal relationship, we applied the difference-in-differences estimation in the next section to quantify the difference in aggression levels between the lockdown and nonlockdown groups in a statistical manner.

Difference in Differences Using Poisson Regression

The traditional way to investigate a causal relationship is an experiment conducted on randomly assigned subjects, in which participants are randomly separated into 2 groups. One group receives treatment, and the other does not. However, in many cases, including ours, an experiment is not viable owing to practical or ethical reasons. For example, we cannot randomly assign people and put them under lockdown for an extended time. Some social scientific researchers use multivariate regression to solve this problem, when the independent variable of interest, X, and other correlated variables, Z_1, …, Z_k, act together to determine the outcome, Y. Although this method can control for the effect of the selected Z variables, some other potentially relevant variables might be lacking in data or difficult to identify, leaving a possibility that important variables are not considered. To address this problem, scientists have used the difference-in-differences method.

As Callaway and Sant'Anna indicated, "Difference in differences (DID) has become one of the most popular research designs used to evaluate causal effects in policy interventions" [33]. Difference in differences compares the difference between the treatment group and the control group at a particular time (T1) with that between them at another time (T2), with the 2 times separated by a particular intervention. This method compares the difference at T1 (D_{T1}) with the difference at T2 (D_{T2}) and measures whether the difference between D_{T1} and D_{T2} (difference in differences) has causal relations with the intervention. In short, difference in differences measures whether the intervention causally impacts the difference between D_{T1} and D_{T2} [34]. In this research, the treatment group refers to those Twitter users under lockdown, and the control group refers to those not under lockdown. The intervention is lockdown. Our objective was to compare the difference in the level of aggression between these 2 groups (D_{T2}) with that between the 2 groups before lockdown (D_{T1}) and measure whether the difference between D_{T1} and D_{T2} is causally related to lockdown.

To implement the difference-in-differences estimation in time-series data, we use fixed effect models. Fixed effect models address unseen variables by controlling for the average in each geographic and temporal data group (data group, in short). The average in each data group is constituted by many factors, including those Z variables that we may or may not know. In other words, the effects of the Z variables we need to control are captured in the average of each data group. By subtracting the average from the outcome in each data group, fixed effect models control for the influence of miscellaneous Z variables and measure the net increase in the Y variable to which the X variable contributes. Taking "fixed effect of the states" as an example, the difference in aggression levels caused by different political tendencies and racial compositions was captured in the average aggression level in each state. After subtracting the average, we now measured how aggression levels increased with respect to each state's norm. Models with fixed effects used in this research "come closer than does ordinary regression analysis to achieving unbiased estimates of causal effect" [35].

To implement the fixed effect model above, we needed to add the fixed effects to a regression model best suited for our data. Because our original observation was the number of aggressive tweets posted in a specific spatiotemporal setting, our data represented a type of count data. Therefore, we selected the classic Poisson model for count data with fixed effects [36]. The following equation was initially used:

$$\ln(\mathit{Aggressive\ Tweet\ Count}_{s,t}) = \alpha_0 + \alpha_1 \mathit{under\ lockdown} + \mu_{state} + \sigma_{time} + \varepsilon \quad \mathbf{(1)}$$

In this case, *Aggressive Tweet* $Count_{s,t}$ is the number of aggressive tweets in a specific state (s) under a specific time (t). α_0 is the constant in standard regression models. Variable α_1 signifies the treatment effect of *under lockdown* on the aggressive tweet count. *Under lockdown* is a binary variable (the explanatory variable in this experiment) that has a value of 0 for not under lockdown and 1 for under lockdown. The model also included state and time fixed effects as follows: μ_{state} and σ_{time}. These 2 variables do not have a specific range but rather represent the average of their corresponding groups of data (eg, data in a state or in a specific week). ε is the error term included in all statistical regressions.

As aforementioned, however, an increase in aggressive tweet counts may be due to an increase in the total tweets posted, which does not necessarily indicate a higher level of aggression. Measuring the proportion of aggressive tweets more accurately depicts the aggression level on Twitter. Therefore, to measure the proportion of aggressive tweets using this count-based model, we exposed the estimation to the total number of tweets by adding the term ln(*Total*), with the coefficient fixed to 1, to the equation. This action is designed for situations like ours and is supported by the *exposure()* option in Stata 17 software [37]. More specifically, our equation was now as follows:

$$\ln(Aggressive\ Tweet\ Count_{s,t}) = \alpha_0 + \alpha_1 under\ lockdown + \mu_{state} + \sigma_{time} + \varepsilon + \ln(Total_{s,t}) \quad \mathbf{(2)}$$

To understand the mechanism of how adding $\ln(Total_{s,t})$ allows us to estimate the proportion rather than the count, we can look at the equation in the following way: when we subtract $\ln(Total_{s,t})$ on both sides of the equation, the estimation is equivalent to modeling the proportion of aggressive tweets. The equation is as follows:

$$\ln(Aggressive\ Tweet\ Count_{s,t} / Total_{s,t}) = \alpha_0 + \alpha_1 under\ lockdown + \mu_{state} + \sigma_{time} + \varepsilon \quad \mathbf{(3)}$$

Our model comes closer to capturing the unbiased causal effect of the independent variable of interest on the dependent variable in observational data [35]. Since most other independent factors that might affect the result, such as seasonal and regional factors, have been absorbed by time and state fixed effects, a significant result in this difference-in-differences analysis can not only indicate a concrete positive correlation but also strongly suggest a causal relationship. All analyses were conducted using Stata BE Edition 17.0 (StataCorp).

Results

Observing the Difference in Aggression Levels Between Groups Under and Not Under Lockdown

This analysis compared aggression levels between states under and not under lockdown. The United States declared a state of emergency on March 13, 2020. Among 42 states that had ever imposed lockdown, 40 started lockdown in the 2 weeks between March 20 and April 4, 2020. Figure 2 shows data from April to October 2020, when the pandemic was getting severe in the United States and some states began lockdown. It illustrates the weekly difference in aggression levels between groups under and not under lockdown. Figure 3 zooms out the timeframe to include data from 2019, putting data under the pandemic into a broader perspective.

Figure 2. Weekly aggressive tweet percentages since April 1, 2020, for all 3 aggressive emotions: anger (A), offensive (B), and hate (C). States under lockdown (magenta) and not under lockdown (cyan) are indicated.

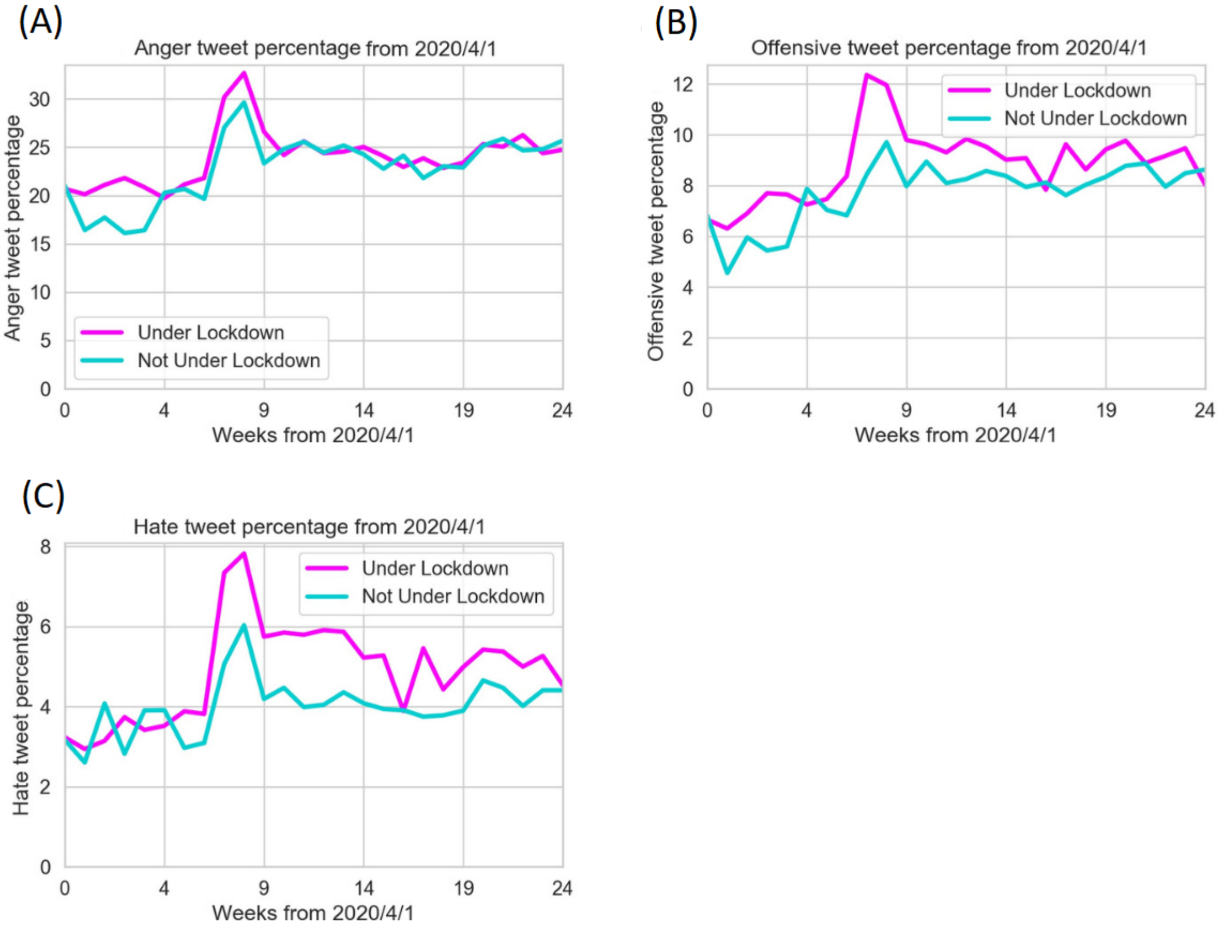
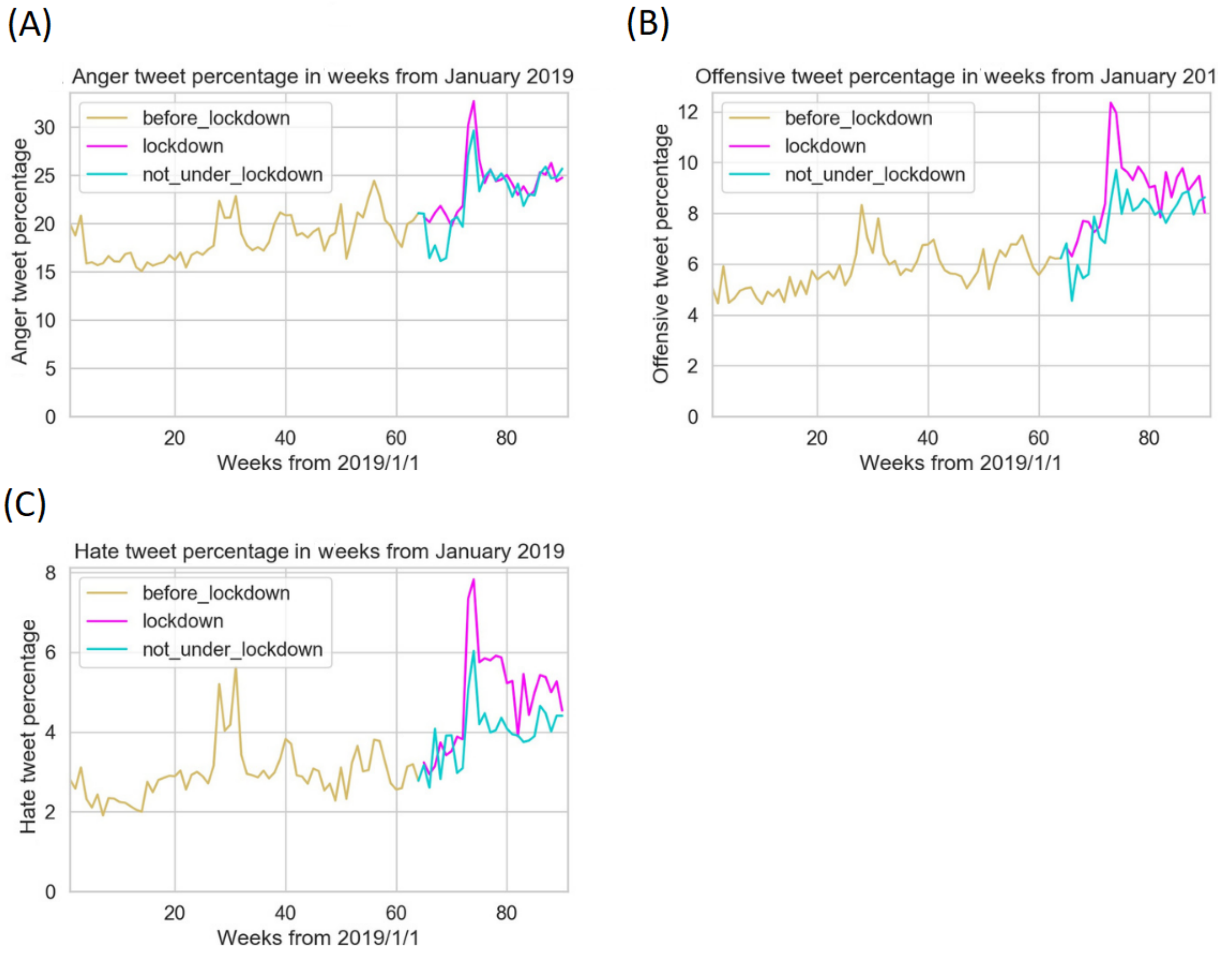


Figure 3. Weekly aggressive tweet percentages since January 1, 2019, for all 3 aggressive emotions: anger (A), offensive (B), and hate (C). States under lockdown (magenta), states not under lockdown (cyan), and data before any lockdown started (yellow) are indicated.



Anger

In terms of the intense feeling of displeasure or antagonism, there was a sudden increase in tweet count in May 2020, particularly evident in those under lockdown. In the first 9 weeks since April, average anger levels were 2%-3% higher in the group under lockdown than in the other group. Figure 3 shows that the percentage of angry tweets fluctuated around 20% in 2019. Coming into May 2020, the percentage rose to as high as 34%. In the summer of 2020, anger levels decreased, and the 2 groups demonstrated similar angry tweet percentages.

Offensive Language

For unacceptable language that can potentially harm a disadvantaged group, tweet proportions increased sharply for 7 weeks since April 2020 (Figure 2). Offensive levels fluctuated around 6% in 2019 (Figure 3). In the 7th week, offensive tweet percentages in under and not under lockdown groups surged to 12.2% and 9.8%, respectively. Afterward, the numbers started to gradually decrease. Although similar trends were seen in both the under and not under lockdown groups, the tweet percentages under lockdown were consistently 2%-3% higher than the values in the other group.

Hate Speech

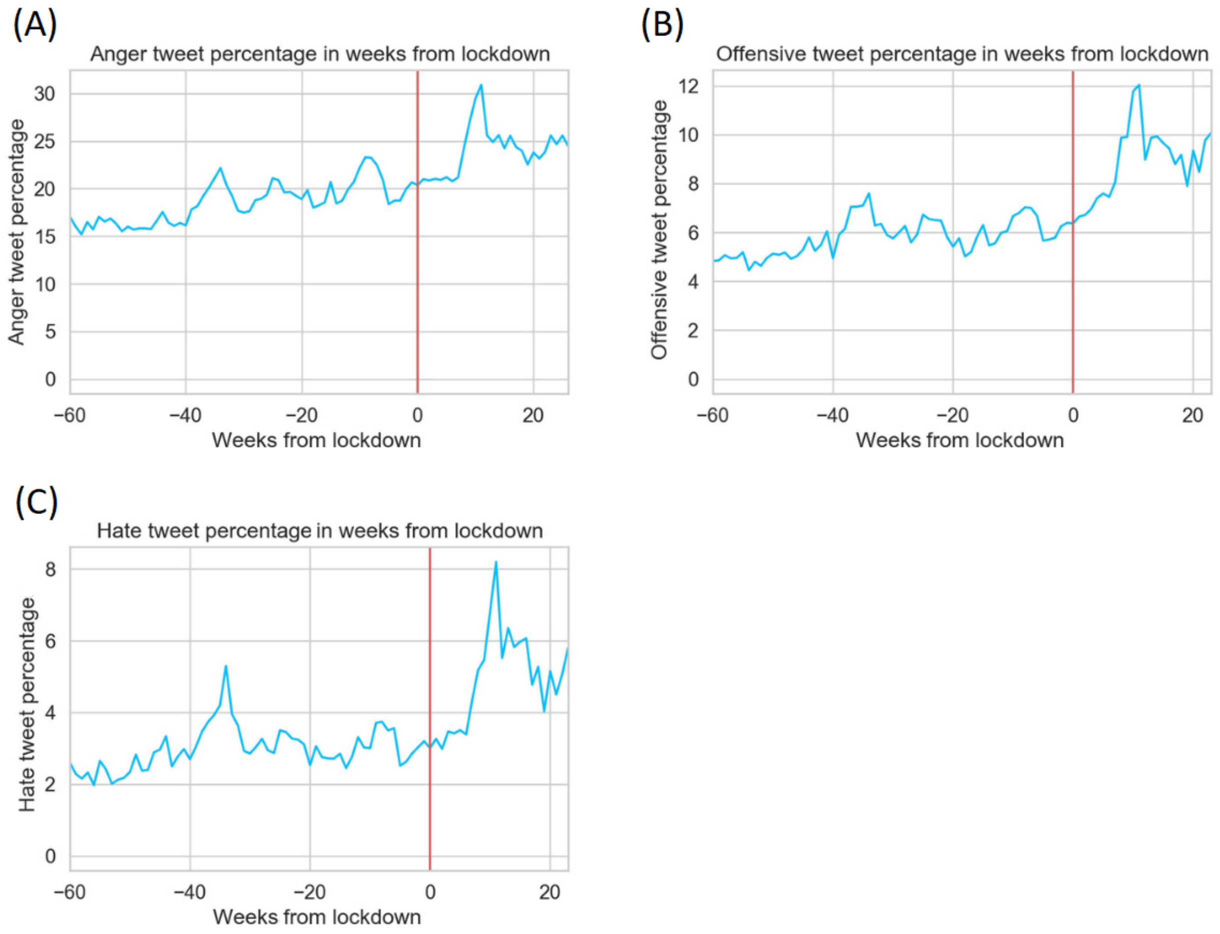
For derogatory, humiliating, and insulting speech intended to express hatred to a targeted group, tweet proportions reached the peak in the 8th week from April 1, 2020 (Figure 2). Hate speech percentages surged from around 3% when the pandemic started to 8% and 6% for people under and not under lockdown. Percentages decreased in the summer of 2020, gradually stabilizing in the months thereafter. Similar to trends in offensive language, hate tweet percentages were considerably greater than the values in 2019 after the decrease (Figure 3). After the initial surge, hate speech percentages under lockdown were 1%-2% higher than the values in the other group.

Aggression levels surged in all 3 kinds of aggressive emotions, reaching the peak around 6-8 weeks from April 1, 2020 (Figure 2). People under lockdown demonstrated a more aggressive tendency throughout the process than those not under lockdown. People who were not under lockdown experienced similar trends to those under lockdown, but to a less drastic degree. After the initial peak of increased aggression, all 3 kinds of emotional tweet percentages decreased to a relatively stable plateau. This stabilization might suggest that people are gradually getting used to the situation, and emotions are relatively eased compared with the sudden displeasure in the early days of lockdown. Despite the temporary decrease, aggressive tweet percentages were considerably higher than the values in 2019 (Figure 3).

Observing Aggression Trends From the Weeks Before and After Lockdown

We selected states that had undergone lockdown and compared their aggression levels before and after lockdown. We visualized the increasing trends of aggression after lockdown. Figure 4 shows the weekly changes in tweets containing the target emotions. In all 3 kinds of aggressive emotions, there was a visible surge in tweet percentage within 10 weeks after lockdown. The table below shows the average weekly tweet percentages (60 weeks before lockdown and 22 weeks after lockdown). In all 3 emotions, the percentage rose after lockdown (anger, 18.51% to 23.77%; offensive, 5.80% to 8.79%; hate, 2.97% to 4.85%). These descriptive data give us a basic grasp of the potential connection between lockdown and increased aggression. Note that this part does not conclude the increase is totally caused by lockdown but rather shows the general trends of aggression before and after lockdown that might be caused by multiple factors. In the next part, we conducted a difference-in-differences analysis to precisely estimate the net impact of lockdown on the increase of aggression.

Figure 4. Aggressive tweet percentages for anger (A), offensive (B), and hate (C) before and after lockdown. The vertical red line at week 0 denotes the start of lockdown. Note that because states might have started lockdown at different times, week 0 can differ in different states. Nevertheless, states generally started their lockdown between March 20, 2020, and April 4, 2020.



Difference in Differences Using Poisson Regression

In this section, we conducted a difference-in-differences analysis using a Poisson regression model. Lockdown was associated with an increase in aggressive tweet proportions for all 3 kinds of aggressive emotions. In a log-linear model, the original coefficient, α_1, between X and Y denotes the increase of ln(Y) for every unit increase of X, which is difficult to interpret due to the presence of logarithm. The incidence rate ratio (IRR) is the exponentiated coefficient of the independent variable of interest, $e^{\alpha 1}$, that demonstrates the increase of Y for every unit increase of X. For all 3 aggressive emotions, the IRR between the aggressive tweet proportion and lockdown status was greater than 1 (anger, 1.049; offensive, 1.168; hate, 1.114), indicating that after the initial lockdown, there were on average 4.9%, 16.8%, and 11.4% increases in emotional tweets for anger, offensive, and hate, respectively. All 3 of the results demonstrated high statistical significance (anger: *P*=.002, offensive: *P*<.001, hate: *P*=.005). Difference-in-differences results are shown in Table 6. Under the control of state and time fixed effects, most possible factors that can lead to misinterpretation were nullified. Therefore, we could measure the net impact of lockdown status on aggression levels. Our estimation strongly suggested a causal relationship between lockdown and increased aggression in all 3 categories of aggressive emotions.

Table 6. Results of Poisson regression for emotional tweet proportion and lockdown status.

Lockdown status	Incidence rate ratio	Standard error	z	P value	95% CI
Under lockdown (anger)	1.049745	0.0163949	3.11	.002	1.018099-1.082375
Under lockdown (offensive)	1.168261	0.0319202	5.69	<.001[a]	1.107345-1.232529
Under lockdown (hate)	1.114432	0.0432653	2.79	.005	1.032780-1.202541

[a] STATA regression yielded *P*=.000.

Discussion

Principal Findings

Infoveillance Study on Aggressive Emotions Under Lockdown

Understanding the trends of aggressive emotions is the first step to understanding various social problems associated with aggressive behaviors during the pandemic. Inspired by the questionnaire study by Killgore et al [2], we used NLP as an infoveillance method to observe the trends of online aggression in the first few months of the pandemic. We hope this method can support traditional psychology surveys by utilizing computer technology to provide a more efficient way of understanding crowd emotions.

By using statewide lockdown status to analyze tweets, we can capture the peaks and valleys of aggression levels throughout a prolonged time period. We can also identify the difference in aggression levels between groups with different lockdown statuses. There were a few particularly noticeable peaks in aggression levels in the observed timeframe. These aberrations might be able to explain the effects of various social events on public sentiment. From the start of April 2020, when most states imposed lockdown, to the second half of May in the same year, aggression levels rose by a magnitude not found in 2019, with people under lockdown demonstrating a more acute rise than others. As the nationwide deaths from COVID-19 skyrocketed from below 100 per day in late March to over 2000 in mid-April, emotions and lifestyles were impacted unprecedentedly. Through lockdowns, death tolls steadily decreased in the next few months, reaching a lower equilibrium in June. The peaks of all 3 kinds of aggressive emotions were observed in the most severe month of the nationwide pandemic when states were experiencing high death tolls and civil unrest.

After a roughly 2-month period of lockdown, aggression levels reduced. This drop might be due to pandemic fatigue, making people feel less stressed and demonstrating less aggression [38]. According to a report by the World Health Organization, “At the beginning of a crisis, most *people* are able to tap into their surge capacity – a collection of mental and physical adaptive systems that humans draw on for short-term survival in acutely stressful situations. However, when dire circumstances drag on, they have to adopt a different style of coping, and fatigue and demotivation may be the result” [39]. Despite the decrease, those under lockdown still demonstrated a higher aggression level than those not under lockdown.

Although aggression trends roughly follow the same pattern in all 3 kinds of aggressive emotions, each has some slightly different characteristics that reveal the uniqueness of each emotion. Anger was the most common emotion among the 3 emotions. After the initial peak, the aggression lines between groups of different lockdown statuses intertwined in the next few months. Offensive and hate were seen less often than anger. However, offensive and hate levels among those under lockdown were consistently higher than the levels among those not under lockdown in the first 6 months of the pandemic.

Our infoveillance study captured the fluctuation of people’s emotions over a specific timeframe, providing vital information for policymakers and public health professionals.

Discovery of a Causal Relationship Between Lockdown and Increased Aggression

Our estimation suggested a causal relationship between lockdown and increased aggression. The Poisson regression analysis designated for count data is suited to estimate the number of aggressive tweets posted in a time period. Time and state fixed effects are able to address the undesired effects of factors other than lockdown status on the outcome. Using this rigorous statistical model, we can show the net impact of lockdown status on the increase of aggressive tweets. The highly significant results in all 3 kinds of aggressive emotions (anger: *P*=.002, offensive: *P*<.001, hate: *P*=.005) matched with the observation in our infoveillance study, that is, people under lockdown have higher aggression levels. Aggressive emotions under lockdown can cause social problems such as domestic violence and divorce. Our findings provide essential information for understanding the causes of aggressive emotions during the pandemic.

Potential Policy Implications

Statistics and scientific evidence play crucial roles in rational policymaking during the pandemic [40]. Using big data to detect potential causal relations between lockdown and aggression may guide governments to implement mental health support policies during lockdowns. In the past, mental health support has come in various different forms, including but not limited to domestic violence protection [41], school counseling [42], and psychological consulting [43]. We hope that our spatiotemporal detection of aggression trends may facilitate more efficient allocation of public resources to areas that are most in need. Moreover, we hope to inspire future researchers to use machine learning to detect social trends that invite proper policy responses. Moreover, we hope our causal analysis can raise social and political awareness of the importance of mental health policies during the pandemic.

Comparison With Prior Work

Killgore et al [2] discovered an increase in aggression levels after lockdown that was particularly evident among those under lockdown. Our research used the Poisson regression model with fixed effects to precisely measure the net impact of lockdown and aggression. This widely established econometric method points to a causal relationship between lockdown and increased aggression [33]. Compared with traditional questionnaire surveys that can only collect data for 1 subject at a time, our data from Twitter are much more versatile. They can be used in other subject research by adjusting the analysis method.

Previous NLP sentiment analysis studies focused on using machines to understand emotions in vast volumes of text [16,17]. However, few of them applied this technology to investigate causal relations of social phenomena in the public health area. Inspired by traditional questionnaire research in psychology, we applied NLP technology to a longitudinal internet study of emotions. This interdisciplinary effort provides crucial information to understand the factors contributing to increased aggression. It opens up new opportunities for NLP technology to make psychology and public health research efficient and timely.

Limitations and Future Work

Our research has several limitations as well as potential for improvement in future work. First, Twitter data overrepresent younger users who have better access to mobile apps and live in a culture that promotes social media. Such users might not accurately reflect the whole population, as certain groups of different demographic and socioeconomic statuses might be underrepresented [44,45]. However, this limitation is not unique to this study but is present in all studies involving Twitter data. Second, due to time and computational limits, the number of tweets we sampled was not very large considering the total number of available tweets. In the future, we can use this research procedure with an increased number of tweet samples to detect aggression levels in space and time at a more granular scale. Nevertheless, the current number of sampled tweets was sufficient for this research to show aggression fluctuations at the statewide scale and draw statistically significant claims. Third, to determine the lockdown status of each user, this research could only use the lockdown status of the geo-tagged US state of the user. Since we were unable to ask the users about their lockdown status at a personal level, this might have led to some inaccuracies in determining the lockdown status of users. This limitation is inherent to social event studies on the internet. Therefore, our study is conducted under the assumption that users followed the lockdown policies in their state. Furthermore, different types and stringencies of lockdown policies emerged in response to the rapidly changing pandemic circumstances. This research only measured the initial lockdown where people were restricted to staying at home. Future research is open to measuring the effect of lockdown policies at different levels and nuances on aggressive emotions and behaviors. Another source of uncertainty comes from the Twitter user location labels, since the location is per user rather than per tweet ("a user who moves from one city to another [and updates his location] will have all of his tweets considered as being from the latter location" [46]). Moreover, the studies by Gore et al and Frank et al showed that the sentiment of a tweet is highly correlated with the geographical area (ie, city) it was composed in [47,48]. Finally, our research method is not restricted to measuring aggressive emotions. Future research can easily apply our methodology to other emotions and research topics. Using NLP technology to help psychology and public health research has vast potential in the future.

Conclusions

Infoveillance studies can be immensely useful in the modern world. With recent advancements in NLP, models can be trained to accurately understand emotions in text. NLP technology can be applied to analyze emotions in large volumes of social media data. This large-scale spatiotemporal data of public emotions can be further analyzed to investigate the correlations and causal relations between emotional trends and certain policies like lockdowns. Applying computer technology to social scientific research has vast potential in the future.

Acknowledgments

The authors wish to thank Dr Yun-Chien Chang, Dr Han-Wei Ho, Dr Chang-Ching Lin, Bo-Wei Chen, Dian-Han Yang, Hao-Hsuan Ting, Sin-En Lu, Lifen Chou, Pan Weili, Peggy Pei-Chin Tsai, Chyn Lin, Chih-Cyuan Liang, and Ashley Chang for their help with this paper. This research was supported in part by the National Science and Technology Council under grants 109-2221-E-008-062-MY3 and 109-2221-E-008-058-MY3.

Data Availability

The source code for this study is available on GitHub [49].

Conflicts of Interest

None declared.

References

1. Shah SGS, Nogueras D, van Woerden HC, Kiparoglou V. The COVID-19 pandemic: A pandemic of lockdown loneliness and the role of digital technology. J Med Internet Res 2020 Nov 05;22(11):e22287 [FREE Full text] [doi: 10.2196/22287] [Medline: 33108313]

2. Killgore WD, Cloonan SA, Taylor EC, Anlap I, Dailey NS. Increasing aggression during the COVID-19 lockdowns. J Affect Disord Rep 2021 Jul;5:100163 [FREE Full text] [doi: 10.1016/j.jadr.2021.100163] [Medline: 34075370]
3. Muñoz-Fernández N, Rodríguez-Meirinhos A. Adolescents' concerns, routines, peer activities, frustration, and optimism in the time of COVID-19 confinement in Spain. J Clin Med 2021 Feb 16;10(4):798 [FREE Full text] [doi: 10.3390/jcm10040798] [Medline: 33669433]
4. Breuer J, Elson M. Frustration–Aggression Theory. In: Sturmey P, editor. The Wiley Handbook of Violence and Aggression. Hoboken, NJ: Wiley; 2017.
5. Sher L. The impact of the COVID-19 pandemic on suicide rates. QJM 2020 Oct 01;113(10):707-712 [FREE Full text] [doi: 10.1093/qjmed/hcaa202] [Medline: 32539153]
6. Moreira DN, Pinto da Costa M. The impact of the Covid-19 pandemic in the precipitation of intimate partner violence. Int J Law Psychiatry 2020 Jul;71:101606 [FREE Full text] [doi: 10.1016/j.ijlp.2020.101606] [Medline: 32768122]
7. Niedzwiedz CL, Green MJ, Benzeval M, Campbell D, Craig P, Demou E, et al. Mental health and health behaviours before and during the initial phase of the COVID-19 lockdown: longitudinal analyses of the UK Household Longitudinal Study. J Epidemiol Community Health 2021 Mar 25;75(3):224-231 [FREE Full text] [doi: 10.1136/jech-2020-215060] [Medline: 32978210]
8. Marroquín B, Vine V, Morgan R. Mental health during the COVID-19 pandemic: Effects of stay-at-home policies, social distancing behavior, and social resources. Psychiatry Res 2020 Nov;293:113419 [FREE Full text] [doi: 10.1016/j.psychres.2020.113419] [Medline: 32861098]
9. 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]
10. Pesonen A, Lipsanen J, Halonen R, Elovainio M, Sandman N, Mäkelä JM, et al. Pandemic dreams: Network analysis of dream content during the COVID-19 lockdown. Front Psychol 2020 Oct 1;11:573961 [FREE Full text] [doi: 10.3389/fpsyg.2020.573961] [Medline: 33117240]
11. 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 Sep 24;15(9):e0239698 [FREE Full text] [doi: 10.1371/journal.pone.0239698] [Medline: 32970764]
12. Majumdar P, Biswas A, Sahu S. COVID-19 pandemic and lockdown: cause of sleep disruption, depression, somatic pain, and increased screen exposure of office workers and students of India. Chronobiol Int 2020 Aug 13;37(8):1191-1200. [doi: 10.1080/07420528.2020.1786107] [Medline: 32660352]
13. Usta J, Murr H, El-Jarrah R. COVID-19 lockdown and the increased violence against women: Understanding domestic violence during a pandemic. Violence and Gender 2021 Sep 01;8(3):133-139. [doi: 10.1089/vio.2020.0069]
14. Sanga S, McCrary J. The impact of the coronavirus lockdown on domestic violence. SSRN Journal 2020:1-27 [FREE Full text] [doi: 10.2139/ssrn.3612491]
15. Berkowitz L. Frustration-aggression hypothesis: examination and reformulation. Psychol Bull 1989 Jul;106(1):59-73. [doi: 10.1037/0033-2909.106.1.59] [Medline: 2667009]
16. Kruspe A, Häberle M, Kuhn I, Zhu XX. Cross-language sentiment analysis of European Twitter messages duringthe COVID-19 pandemic. arXiv. 2020. URL: https://arxiv.org/abs/2008.12172 [accessed 2022-07-13]
17. Li X, Zhou M, Wu J, Yuan A, Wu F, Li J. Analyzing COVID-19 on online social media: Trends, sentiments and emotions. arXiv. 2020. URL: https://arxiv.org/abs/2005.14464 [accessed 2022-07-13]
18. Mehedi Shamrat FMJ, Chakraborty S, Imran MM, Muna JN, Billah MM, Das P, et al. Sentiment analysis on Twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. IJEECS 2021 Jul 01;23(1):463. [doi: 10.11591/ijeecs.v23.i1.pp463-470]
19. Jang H, Rempel E, Roe I, Adu P, Carenini G, Janjua NZ. Tracking public attitudes toward COVID-19 vaccination on tweets in Canada: Using aspect-based sentiment analysis. J Med Internet Res 2022 Mar 29;24(3):e35016 [FREE Full text] [doi: 10.2196/35016] [Medline: 35275835]
20. 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]
21. Mohamed Ridhwan K, Hargreaves CA. Leveraging Twitter data to understand public sentiment for the COVID‐19 outbreak in Singapore. International Journal of Information Management Data Insights 2021 Nov;1(2):100021. [doi: 10.1016/j.jjimei.2021.100021]
22. Su Y, Xue J, Liu X, Wu P, Chen J, Chen C, et al. Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: A psycholinguistic analysis on Weibo and Twitter. Int J Environ Res Public Health 2020 Jun 24;17(12):4552 [FREE Full text] [doi: 10.3390/ijerph17124552] [Medline: 32599811]
23. Kotelnikova A. Comparison of deep learning and rule-based method for the sentiment analysis task. 2020 Presented at: 2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon); October 6-9, 2020; Vladivostok, Russia. [doi: 10.1109/FarEastCon50210.2020.9271333]
24. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: CRC Press; 2020.

25. Catelli R, Pelosi S, Esposito M. Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. Electronics 2022 Jan 26;11(3):374. [doi: 10.3390/electronics11030374]
26. Demszky D, Movshovitz-Attias D, Ko J, Cowen A, Nemade G, Ravi S. GoEmotions: A dataset of fine-grained emotions. arXiv. 2020. URL: https://arxiv.org/abs/2005.00547 [accessed 2022-07-13]
27. Davidson T, Warmsley D, Macy M, Weber I. Automated hate speech detection and the problem of offensive language. arXiv. 2017. URL: https://arxiv.org/abs/1703.04009 [accessed 2022-07-13]
28. Founta AM, Djouvas C, Chatzakou D, Leontiadis I, Blackburn J, Stringhini G, et al. Large scale crowdsourcing and characterization of twitter abusive behavior. 2018 Presented at: 12th International AAAI Conference on Web and Social Media; June 25-28, 2018; Stanford, CA.
29. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. 2018. URL: https://arxiv.org/abs/1810.04805 [accessed 2022-07-13]
30. Badjatiya P, Gupta S, Gupta M, Varma V. Deep learning for hate speech detection in tweets. In: WWW '17 Companion: Proceedings of the 26th International Conference on World Wide Web Companion. 2017 Presented at: 26th International Conference on World Wide Web Companion; April 3-7, 2017; Perth, Australia. [doi: 10.1145/3041021.3054223]
31. Warner W, Hirschberg J. Detecting hate speech on the world wide web. In: LSM '12: Proceedings of the Second Workshop on Language in Social Media. 2012 Presented at: Second Workshop on Language in Social Media; June 7, 2012; Montreal, Canada.
32. Schmidt A, Wiegand M. A survey on hate speech detection using natural language processing. In: Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media. 2017 Presented at: Fifth International Workshop on Natural Language Processing for Social Media; April 3, 2017; Valencia, Spain. [doi: 10.18653/v1/w17-1101]
33. Callaway B, Sant'Anna PH. Difference-in-differences with multiple time periods. Journal of Econometrics 2021 Dec;225(2):200-230. [doi: 10.1016/j.jeconom.2020.12.001]
34. Angrist JD, Pischke JS. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press; 2009.
35. Kempf-Leonard K. Encyclopedia of Social Measurement. Amsterdam, Netherlands: Elsevier; 2004.
36. Hilbe JM. Modeling Count Data. Cambridge, United Kingdom: Cambridge University Press; 2014.
37. Stata 17 Base Reference Manual. College Station, TX: Stata Press; 2021.
38. Bodenmann G, Meuwly N, Bradbury TN, Gmelch S, Ledermann T. Stress, anger, and verbal aggression in intimate relationships: Moderating effects of individual and dyadic coping. Journal of Social and Personal Relationships 2010 Apr 22;27(3):408-424. [doi: 10.1177/0265407510361616]
39. Pandemic fatigue: reinvigorating the public to prevent COVID-19: policy considerations for Member States in the WHO European Region. World Health Organization. URL: https://apps.who.int/iris/handle/10665/335820 [accessed 2022-07-15]
40. Dattner I, Gal R, Goldberg Y, Goldshtein I, Huppert A, Kenett RS, et al. The role of statisticians in the response to COVID-19 in Israel: a holistic point of view. Isr J Health Policy Res 2022 Apr 20;11(1):22 [FREE Full text] [doi: 10.1186/s13584-022-00531-y] [Medline: 35443682]
41. Chandan JS, Taylor J, Bradbury-Jones C, Nirantharakumar K, Kane E, Bandyopadhyay S. COVID-19: a public health approach to manage domestic violence is needed. The Lancet Public Health 2020 Jun;5(6):e309. [doi: 10.1016/s2468-2667(20)30112-2]
42. Pincus R, Hannor-Walker T, Wright L, Justice J. COVID-19's effect on students: How school counselors rise to the rescue. NASSP Bulletin 2020 Dec 11;104(4):241-256. [doi: 10.1177/0192636520975866]
43. Kavoor AR, Chakravarthy K, John T. Remote consultations in the era of COVID-19 pandemic: Preliminary experience in a regional Australian public acute mental health care setting. Asian J Psychiatr 2020 Jun;51:102074 [FREE Full text] [doi: 10.1016/j.ajp.2020.102074] [Medline: 32294583]
44. Wilson K, Atkinson K, Deeks S. Opportunities for utilizing new technologies to increase vaccine confidence. Expert Rev Vaccines 2014 Aug 14;13(8):969-977. [doi: 10.1586/14760584.2014.928208] [Medline: 24931799]
45. Karami A, Kadari RR, Panati L, Nooli SP, Bheemreddy H, Bozorgi P. Analysis of geotagging behavior: Do geotagged users represent the Twitter population? IJGI 2021 Jun 02;10(6):373. [doi: 10.3390/ijgi10060373]
46. Mislove A, Lehmann S, Ahn YY, Onnela JP, Rosenquist J. Understanding the demographics of Twitter users. In: Proceedings of the International AAAI Conference on Web and Social Media. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17–21, 2011; Barcelona, Spain.
47. Gore RJ, Diallo S, Padilla J. You are what you tweet: Connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015 Sep 2;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
48. Frank MR, Mitchell L, Dodds PS, Danforth CM. Happiness and the patterns of life: a study of geolocated tweets. Sci Rep 2013 Sep 12;3(1):2625 [FREE Full text] [doi: 10.1038/srep02625] [Medline: 24026340]
49. Increased Online Aggression During COVID-19 Lockdowns: Two-Stage Study of Deep Text Mining and Difference-in-Differences Analysis. GitHub. URL: https://github.com/JeromeTH/Increased-Online-Aggression-During-COVID-19-Lockdowns [accessed 2022-08-02]

Abbreviations

AHSD: Automated Hate Speech Detection and the Problem of Offensive Language
BERT: Bidirectional Encoder Representations from Transformers
IRR: incidence rate ratio
NLP: natural language processing

Edited by A Mavragani; submitted 17.04.22; peer-reviewed by A Klein, R Gore; comments to author 06.05.22; revised version received 19.06.22; accepted 23.06.22; published 09.08.22.

Please cite as:
Hsu JTH, Tsai RTH
Increased Online Aggression During COVID-19 Lockdowns: Two-Stage Study of Deep Text Mining and Difference-in-Differences Analysis
J Med Internet Res 2022;24(8):e38776
URL: https://www.jmir.org/2022/8/e38776
doi:10.2196/38776
PMID:35943771

Original Paper

Teledermatology to Facilitate Patient Care Transitions From Inpatient to Outpatient Dermatology: Mixed Methods Evaluation

Samantha M R Kling[1*], RD, PhD; Erika A Saliba-Gustafsson[1*], PhD; Marcy Winget[1], PhD; Maria A Aleshin[2], MD; Donn W Garvert[1], MS; Alexis Amano[1], MSc; Cati G Brown-Johnson[1], PhD; Bernice Y Kwong[2], MD; Ana Calugar[2], MHSA; Ghida El-Banna[2], MPP; Jonathan G Shaw[1], MS, MD; Steven M Asch[1,3], MPH, MD; Justin M Ko[2], MBA, MD

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

[2]Department of Dermatology, Stanford University School of Medicine, Stanford, CA, United States

[3]Center for Innovation to Implementation (Ci2i), Veterans Affairs Palo Alto Heath Care System, Menlo Park, CA, United States

[*]these authors contributed equally

Corresponding Author:
Samantha M R Kling, RD, PhD
Division of Primary Care and Population Health
Department of Medicine
Stanford University School of Medicine
3180 Porter Dr, MC 5702
Palo Alto, CA, 94304
United States
Phone: 1 650 723 1146
Email: skling@stanford.edu

Abstract

Background: Both clinicians and patients have increasingly turned to telemedicine to improve care access, even in physical examination–dependent specialties such as dermatology. However, little is known about whether teledermatology supports effective and timely transitions from inpatient to outpatient care, which is a common care coordination gap.

Objective: Using mixed methods, this study sought to retrospectively evaluate how teledermatology affected clinic capacity, scheduling efficiency, and timeliness of follow-up care for patients transitioning from inpatient to outpatient dermatology care.

Methods: Patient-level encounter scheduling data were used to compare the number and proportion of patients who were scheduled and received in-clinic or video dermatology follow-ups within 14 and 90 days after discharge across 3 phases: June to September 2019 (before teledermatology), June to September 2020 (early teledermatology), and February to May 2021 (sustained teledermatology). The time from discharge to scheduling and completion of patient follow-up visits for each care modality was also compared. Dermatology clinicians and schedulers were also interviewed between April and May 2021 to assess their perceptions of teledermatology for postdischarge patients.

Results: More patients completed follow-up within 90 days after discharge during early (n=101) and sustained (n=100) teledermatology use than at baseline (n=74). Thus, the clinic's capacity to provide follow-up to patients transitioning from inpatient increased from baseline by 36% in the early (101 from 74) and sustained (100 from 74) teledermatology periods. During early teledermatology use, 61.4% (62/101) of the follow-ups were conducted via video. This decreased significantly to 47% (47/100) in the following year, when COVID-19–related restrictions started to lift (*P*=.04), indicating more targeted but still substantial use. The proportion of patients who were followed up within the recommended 14 days after discharge did not differ significantly between video and in-clinic visits during the early (33/62, 53% vs 15/39, 38%; *P*=.15) or sustained (26/53, 60% vs 28/47, 49%; *P*=.29) teledermatology periods. Interviewees agreed that teledermatology would continue to be offered. Most considered postdischarge follow-up patients to be ideal candidates for teledermatology as they had undergone a recent in-person assessment and might have difficulty attending in-clinic visits because of competing health priorities. Some reported patients needing technological support. Ultimately, most agreed that the choice of follow-up care modality should be the patient's own.

Conclusions: Teledermatology could be an important tool for maintaining accessible, flexible, and convenient care for recently discharged patients needing follow-up care. Teledermatology increased clinic capacity, even during the pandemic, although the timeliness of care transitions did not improve. Ultimately, the care modality should be determined through communication with patients to incorporate their and their caregivers' preferences.

(J Med Internet Res 2022;24(8):e38792) doi:10.2196/38792

KEYWORDS

teledermatology; telemedicine; telehealth; video visits; care transitions; care coordination; discharge planning; follow-up; inpatient; outpatient; mixed methods; dermatology; mobile phone; smartphone

Introduction

In 2020, the COVID-19 pandemic drove telemedicine to the forefront of health care [1,2]; dermatology care was no exception. Prepandemic teledermatology had gained popularity in some specific use cases [3,4]; however, the pandemic gave rise to new policies that overcame previous restrictions to ensure continued access to care, facilitating a rapid pivot to telemedicine for outpatients, including patients transitioning from inpatient care [5-7].

The highly visual nature of dermatology is well suited for this cost-effective and efficient care modality [4,8-10], which is well received by clinicians and patients [10-14]. Although convenience and improved access to care are the primary benefits, especially for rural and underserved populations, teledermatology also boasts time and cost savings, greater flexibility for dermatologists and patients, fewer no-shows, and better continuity of care [4,5,10,15-25]. However, teledermatology has some shortcomings that affect care delivery, including suboptimal image quality, patient privacy, diagnostic accuracy, network connectivity, patient technological literacy, and access to digital devices [8,9,11,15]. In addition, the inaccessibility of in-clinic tools and treatments (eg, dermoscopy, biopsy, and cryotherapy) makes managing certain conditions challenging [9]. These limitations may disproportionally affect patients with low socioeconomic status, Medicare beneficiaries, older adults, and non–English-preferring patients [9,11], who may also be at risk for delayed care transitions.

Nevertheless, teledermatology may be particularly beneficial for patients transitioning from inpatient dermatology consultation services to outpatient dermatology care. Currently, high-risk patients who are hospitalized often experience numerous comorbidities and may experience difficulties accessing in-clinic follow-up care. As a result, they risk receiving fragmented care and being lost to follow-up, which could have serious health consequences [26-28]. Teledermatology may improve follow-up access for these patients by increasing the capacity of dermatology clinics and improving the efficiency of scheduling and care provision. As video visits become a fixture in health care expected by patients and clinicians, it is essential to understand whether teledermatology supports timely care transitions. We retrospectively evaluated teledermatology use and its impact on the clinic's capacity, scheduling efficiency, and timeliness of follow-up care for patients transitioning from inpatient to outpatient dermatology care and explored dermatology clinicians' and schedulers' perceptions of teledermatology for this patient population.

Methods

Setting

Stanford University's Department of Dermatology encompasses 13 outpatient clinics with 16 subspecialties and provides inpatient consultative services in a quaternary hospital; that is, consultation requests placed by the patient's admitting team, such as general medicine or oncology. Consultations are delivered by 5 dermatologists and 2 dermatology residents on monthly rotations. The team consults >1500 inpatients per year, many of whom have complex, high-risk skin conditions in immunocompromised states and have multiple clinical teams involved in their care. Approximately 40% of these patients require postdischarge outpatient follow-up.

Intervention: Teledermatology

The department rapidly implemented teledermatology across all ambulatory clinics in response to the statewide COVID-19 stay-at-home orders in March 2020 [10]. Clinicians were provided with video visit–enabled hardware to enable the remote provision of teledermatology. All clinicians and staff completed the web-based training developed for the institution's rollout. Initially, clinicians and staff were encouraged to convert all nonurgent or emergent in-clinic visits into video visits. Once in-clinic capacity began to expand in spring 2020, department-developed clinical criteria guided appropriate video visit use for all patients except (1) patients with high skin cancer risk requiring full skin examination, including melanoma; (2) patients requiring specialized examinations (scalp and genitals); and (3) patients requiring procedural interventions. As of July 2022, teledermatology had remained a fixture and was offered to patients transitioning from inpatient to outpatient care.

Mixed Methods

Overview

Outcomes derived from quantitative scheduling data and qualitative interviews are defined in Table 1. Data were consolidated throughout the analysis and interpreted in parallel to understand converging and diverging issues regarding teledermatology use and its impact on the clinic's capacity, clinical appropriateness, sustainability, and the remaining barriers for patients transitioning from inpatient to outpatient care.

Table 1. Outcomes, definitions, and data sources used to evaluate the use, impact, and sustainability of teledermatology for patients transitioning from inpatient to outpatient dermatology follow-up care.

Outcomes and definitions	Data sources
Clinic's capacity	
Number and proportion of patients after discharge	
Scheduled follow-up within 90 days after discharge	Patient-level scheduling data
Completed follow-up within 90 days after discharge	Patient-level scheduling data
Teledermatology use	
Number and proportion of follow-up visits completed over video within 90 days after discharge	Patient-level scheduling data
Acceptability of teledermatology for postdischarge follow-up patients among clinicians, residents, schedulers, and patients	Clinician and scheduler interviews
Clinical appropriateness	
Perceived fit or compatibility of teledermatology within this setting, particularly for patients transitioning from inpatient to outpatient dermatology care	Clinician and scheduler interviews
Teledermatology to support timely care transitions	
Scheduling efficiency	
Days from hospital discharge to initial scheduling for in-clinic and video visits	Patient-level scheduling data
Days from hospital discharge to finalized scheduling for in-clinic and video visits	Patient-level scheduling data
Perceived impact of teledermatology on scheduling efficiency	Clinician and scheduler interviews
Timeliness of follow-up visits	
Days from hospital discharge to follow-up visit completion for in-clinic and video visits	Patient-level scheduling data
Number and proportion of patients who attended follow-up within 14 days after discharge (local benchmark)	Patient-level scheduling data
Incomplete follow-up visits	
Number and proportion of patients who scheduled but did not complete a teledermatology or in-clinic visit	Patient-level scheduling data
Perceived impact of teledermatology on follow-up visit completion	Clinician and scheduler interviews
Remaining barriers to video visit coordination	
Perceived long-term sustainability of video visits and the barriers need to be addressed to improve clinician, scheduler, and patient experience	Clinician and scheduler interviews

Quantitative: Inclusion Criteria, Data Collection, and Analysis

Patients who received a dermatology consultation in the inpatient or emergency department settings were discharged in 1 of the 3 study periods and potentially needed follow-up with outpatient dermatology. The three study periods were (1) June 1 to September 30, 2019 (baseline [before teledermatology]); (2) June 1 to September 30, 2020 (early teledermatology); and (3) February 1 to May 31, 2021 (sustained teledermatology). Follow-up scheduling and care were recorded for 90 days after discharge; visits scheduled >90 days after discharge were likely unrelated to the patient's hospitalization. Eligible patients and relevant events were retrospectively identified using the electronic health records and scheduling data. Inpatient dermatology consults were identified using Current Procedural Terminology codes (Multimedia Appendix 1).

Teledermatology use and its impact on the clinic's capacity, scheduling efficiency, and timeliness of follow-up care were compared across periods and visit modalities (video and in-clinic) using the outcomes described in Table 1. Descriptive statistics were calculated to describe patient characteristics and assess differences across the 3 study periods and by visit modality. Statistical significance was assessed using the Kruskal-Wallis test for patient age, chi-square test for categorical (ie, proportional) outcomes, and generalized linear models for continuous outcomes (eg, days from discharge). Differences in teledermatology use by patient age and distance between patient residence and outpatient dermatology clinic were determined using chi-square tests. Clinically meaningful (<70 years vs ≥70 years) or median-based (<21 miles vs ≥21 miles) categories were used. *P* values were adjusted for multiple comparisons with an adaptive, 2-stage linear step-up procedure, and significance was set at *P*<.05 [29].

Qualitative: Data Collection and Analysis

We designed a semistructured interview guide to capture perceptions of teledermatology for follow-up care of patients transitioning from the inpatient setting. Clinicians and schedulers were eligible if they were involved in transitioning patients from inpatient to outpatient dermatology. All eligible clinicians and schedulers (ie, 5 dermatologists, 5 dermatology residents, and 13 schedulers) were invited via email (plus 2 reminders) to participate in a 30-minute phone interview. Ultimately, 15 interviews (5/5, 100% dermatologists; 5/5, 100% residents; and 6/13, 46% schedulers) were conducted between April and May 2021 by 2 experienced qualitative researchers (EAS-G and AA), ranging from 30 to 60 minutes. The interviews were audio recorded and subsequently transcribed.

Data were analyzed deductively and inductively using Microsoft Excel. Deductive codes were derived from the Proctor implementation outcomes [30]. We used multiphase matrix analysis by leveraging rapid analytic procedures to achieve consensus coding of transcripts and extract early themes [31]. EAS-G and AA independently summarized transcripts after each interview; summaries were reviewed, and consensus discussions were held. Summaries were then consolidated into a matrix to identify and compare themes across interviewees. To ensure anonymity, all identifiable information was removed from transcripts, summaries, and reports.

Ethics Approval

This retrospective quality improvement evaluation received a nonresearch determination by Stanford University's Institutional Review Board (IRB-60382). Interviewees provided informed verbal consent before initiating the interviews and were assured that all responses would remain confidential. Detailed interview notes were taken if consent for recording was not provided.

Results

Patient Characteristics

Patient characteristics are summarized in Table 2. Briefly, 194 patients, 218 patients, and 256 patients were discharged following an inpatient dermatology consultation during the baseline, early teledermatology, and sustained teledermatology phases, respectively. The median patient age was similar across the 3 periods (61.0, 60.5, and 55.5 years for baseline, early teledermatology, and sustained teledermatology, respectively; *P*=.11). Approximately half of the patients lived ≥21 miles from the dermatology clinic during each study phase, and most had public insurance.

Table 2. Characteristics of patients who potentially needed outpatient postdischarge follow-up dermatology care following an inpatient dermatology consultation during one of three periods: baseline (N=194), early teledermatology (N=218), and sustained teledermatology (N=256).

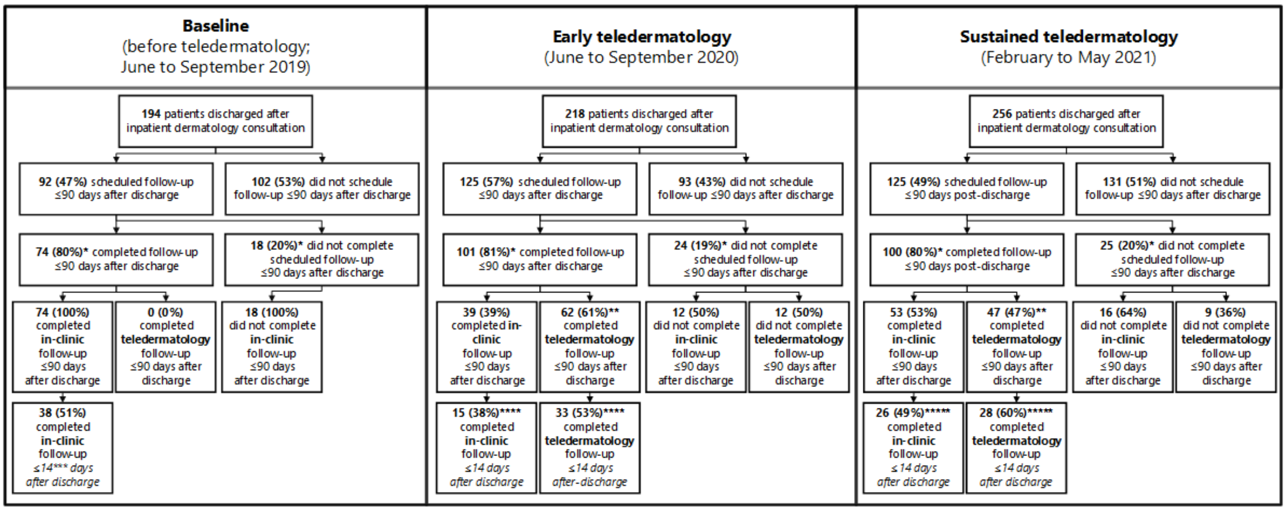
Patient characteristics	Baseline (before teledermatology; June to September 2019), n (%)	Early teledermatology (June to September 2020), n (%)	Sustained teledermatology (February to May 2021), n (%)
Patients with inpatient dermatology consultation	194 (100)	218 (100)	256 (100)
Sex			
Female	100 (51.5)	116 (53.2)	135 (52.7)
Male	94 (48.5)	102 (46.8)	121 (47.3)
Age group (years)			
0-29	19 (9.8)	27 (12.4)	34 (13.3)
30-49	37 (19.1)	48 (22)	69 (27)
50-69	83 (42.8)	82 (37.6)	92 (35.9)
≥70	55 (28.4)	61 (28)	61 (23.8)
Distance from outpatient clinic (miles)[a]			
0-20	90 (46.4)	104 (47.7)	133 (52)
≥21	104 (53.6)	114 (52.3)	123 (48)
Insurance type			
Private	42 (21.6)	28 (12.8)	40 (15.6)
Public	148 (76.3)	187 (85.8)	210 (82)
Other or no insurance identified	4 (2.1)	3 (1.4)	6 (2.3)
Patient hospital stay			
Emergency department	25 (12.9)	40 (18.3)	56 (21.9)
Inpatient	169 (87.1)	178 (81.7)	200 (78.1)
Dermatology specialty for follow-up			
Dermatology	63 (32.5)	91 (41.7)	97 (37.9)
Dermato-oncology	39 (20.1)	41 (18.8)	31 (12.1)
No follow-up	92 (47.4)	86 (39.4)	128 (50)

[a] Median distance between the patient's zip code and the primary dermatology clinic in Palo Alto, CA, United States, was 21 miles.

Clinic Capacity

More patients were scheduled for outpatient dermatology visits within 90 days after discharge during the early (n=125) and sustained (n=125) teledermatology phases than at baseline (n=92), indicating a 36% increase (n=125 from 92 and n=125 from 92 patients for early and sustained teledermatology, respectively) in the scheduling capacity (Figure 1). Similarly, the number of follow-up visits completed within 90 days after discharge was higher in the early teledermatology (n=101) and sustained teledermatology (n=100) than at baseline (n=74), indicating a sustained increase in capacity. The proportion of patients who completed their follow-up did not differ across the 3 evaluation periods, as shown in Figure 1.

Figure 1. Number and percentage of patients discharged following an inpatient dermatology consultation and who were scheduled for and completed an outpatient dermatology follow-up visit within 90 days after discharge in the clinic or via video. *P=.99 indicating no difference between baseline, early teledermatology, and sustained teledermatology phases. **P=.04 indicating significant difference between the early and sustained teledermatology phases. ***A local benchmark. ****P=.15 indicating no difference between the teledermatology and in-clinic follow-ups during the early teledermatology phase. *****P=.29 indicating no difference between the teledermatology and in-clinic follow-ups during the sustained teledermatology phase.



Teledermatology Use

Teledermatology use was highest in the early teledermatology period, with 61.4% (62/101) of follow-ups completed via video. This decreased significantly to 47% (47/100) of follow-ups in the sustainability period (P=.04), indicating a more targeted yet still substantial use (Figure 1). Interviewees remarked that patient acceptance of teledermatology and technology capabilities varied during early implementation; however, acceptance increased as it became the standard of care (see exemplary quotes in Textbox 1).

Clinicians and schedulers believed that older patients may prefer clinic visits, whereas patients residing farther from the clinic may favor video visits (Textbox 1). However, video visit use did not differ by patient age during the early or sustained teledermatology phase or by distance from the clinic during the early teledermatology phase (Table 3). However, in the sustained period, significantly more patients residing ≥21 miles away from the clinic had follow-up video visits (26/41, 63% patients) than those living closer (21/59, 36% patients; P=.01), confirming the interviewees' view that patients living farther away preferred video to in-clinic visits as teledermatology became optional (ie, no longer mandated because of the pandemic).

Textbox 1. Exemplary quotes from interviews with dermatologists, residents, and scheduling staff describing the use of teledermatology for patients transitioning from inpatient to outpatient dermatology care.

Teledermatology use

- “In the pandemic, people were frequently upset that they had to do a video visit. They wanted to be seen in person [...] Now that people are more used to the virtual world, they seem to be more okay with doing video visits. But I still think that for certain people, they just really don’t want anything to do with them [...]” [Resident 5]
- “...[video visits] made things a little bit easier because we are able to get over that hurdle of travel. So for a patient that maybe is hours away where they’re just never going to come back for a 15-30 minute, dermatology visit, where they’re just not able to do that. It has made us better able to at least connect with them.” [Dermatologist 3]

Clinical appropriateness

- “...for some patients video visits are totally fine, like a patient comes in with a drug rash, they’re totally better, you’re just checking it and making sure they’re not flaring again. It’s perfect. But when there’s an issue where they might need a culture or a lab, or they need a little more intensive care, like wound change or something like that, it’s very difficult. It’s challenging. Sometimes we just need in-person.” [Dermatologist 3]
- “...for some of our sick patients, maybe their skin isn’t their priority, and it isn’t a very complex thing that requires inpatient evaluation; it saves them so much time and stress, and also, kind of helps close the loop on our end as well.” [Dermatologist 2]
- “...medically, from our perspective, people that are in hospital tend to have certain conditions so a close follow-up where we can actually see all their skin, as opposed to pictures, is more helpful.” [Resident 4]
- “...a decent amount of the time we’ll do video visits, but we are leaving it up to patients. So even if it’s something that’s not really serious, they’d rather be seen in person, we’ll still accommodate them in person.” [Resident 5]

Teledermatology to support timely care transitions

- “A lot of the clinicians have more video visits than they do in person so it’s a lot easier just to get them in the video.” [Front office scheduler 2]
- “I think it [video visits] makes our work a little easier because we have more options to give the patient...[video visits] give the patient more options because sometimes patients don’t want to come into clinic, they’d rather do a video. [...] on top of that, with video visits, we can get the patients in sooner because with a lot of the video visits, there’s more video visits available where we can get the patients in sooner versus in-person that are booking months out.” [New patient coordinator 3]
- “The biggest difference for me is provider availability particularly for patients that live pretty far away because they’re able to be more flexible in terms of when they can schedule and then they can schedule sooner.” [Resident 3]
- “...it’s probably easier to schedule video visits. I think attendings can squeeze them in a bit faster than in-person. [...] having the option makes it a little easier to schedule in a timely manner. That’s really the only benefit I can think of.” [Resident 5]
- “...no-show rate is much higher for video visits ... significantly higher.” [Resident 5]
- “I find the video visits are a lot easier because they’re more likely to follow-up, because a lot of patients otherwise don't show when they have to come in-person.” [Dermatologist 5]

Remaining barriers to video visit coordination

- “...ideally, if they could leave the hospital with a follow-up appointment, that would actually help even more, but most of the time, that’s just not feasible.” [Resident 3]
- “I...prefer in-person visits to video visits just because you’re relying a lot on the patient uploading photos and then the photos they upload have to be good quality... I have definitely been fooled before where I see photos that a patient’s taken and thought one thing and then when you see them in person, it’s much different. So I think there definitely are limitations to video visit.” [Resident 5]
- “...we have a very large elderly group of patients and elderly people aren’t tech savvy. They’re a main group of people who are vulnerable during COVID but it made it really hard for them to do video visits.” [Front office scheduler 1]
- “...there are barriers, like if we’re not getting the best photos from the patient, or let’s say they have to do labs. There’s a lot of really tricky coordination. They need a lab slip and we have to get it to them somehow. Then they have to get to the lab and call us since we can’t call the lab for the results, and we’re really reliant on their primary care doctors, if they have one.” [Clinician 3]

Table 3. Completion of in-clinic and video outpatient dermatology postdischarge follow-up visits (by patient age and location) for patients who received an inpatient dermatology consultation and were discharged across 3 periods: baseline, early teledermatology, and sustained teledermatology.

Follow-up visit modality	Baseline (before teledermatology; June to September 2019)		Early teledermatology (June to September 2020)				Sustained teledermatology (February to May 2021)			
	Total visits, N	In-clinic, n (%)	Total visits, N	In-clinic, n (%)	Video, n (%)	*P* value[a]	Total visits, N	In-clinic, n (%)	Video, n (%)	*P* value
Patient age (years)						.32				.47
<70	56	56 (100)	72	30 (42)	42 (58)		71	36 (51)	35 (49)	
≥70	18	18 (100)	29	9 (31)	20 (69)		29	17 (59)	12 (41)	
Distance from clinic (miles)[b]						.15				.01
<21	36	36 (100)	47	19 (40)	28 (60)		59	38 (64)	21 (36)	
≥21	38	38 (100)	54	20 (37)	34 (63)		41	15 (37)	26 (63)	

[a] Differences in proportion between in-clinic and video visits were determined using chi-square tests.

[b] The median distance between the patient's zip code and the primary dermatology clinic in Palo Alto, CA, United States was 21 miles.

Clinical Appropriateness

Clinicians expressed interest in continuing to offer video visits to recently discharged patients who have ongoing stable conditions that are well-suited to teledermatologic care or limitations preventing in-clinic care (Textbox 1). Teledermatology was perceived to reduce barriers (eg, lack of time or resources to travel) to attending follow-up in person for high-risk patient populations who may need to attend several follow-up visits and may consequently deprioritize their dermatological issues. A video visit was perceived as better than no follow-up at all for patients who were too debilitated to travel, even if an in-clinic visit was clinically ideal.

Video visits were often considered ideal for quick, simple check-ins for postdischarge patients as they had recently been examined in person, whereas in-clinic visits were considered more appropriate for conditions requiring skin examinations or procedures, laboratory tests, or dressing changes. Nevertheless, a few residents believed that video visits were less suitable for postdischarge patients (Textbox 1). However, most agreed that the choice should ultimately be the patient's own. According to clinicians, patients appreciate having a choice that meets their priorities, needs, and preferences, and accommodating patients to complete the necessary follow-up is worthwhile.

Teledermatology to Support Timely Care Transitions

To understand the impact of teledermatology on the scheduling efficiency and timeliness of care transitions, differences in days from discharge to initial scheduling, finalized scheduling, and completion of an outpatient dermatology follow-up visit across each study period and visit modality were assessed (Table 1). The proportion of incomplete scheduled visits is also reported by the study period and visit modality.

Scheduling Efficiency

The average days from hospital discharge to initial scheduling of in-clinic visits increased from 4.2 (SD 7.2) days at baseline to 8.5 (SD 10.2) days during early teledermatology (*P*=.01; Table 4). This returned to 3.0 (SD 4.8) days in the sustained period, similar to baseline (*P*=.40). In contrast, the average number of days for initial scheduling of video visits was 4.3 (SD 6.8) days and 3.7 (SD 10.1) days for the early and sustainability periods, respectively, similar to the 4.2 (SD 7.2) days at baseline when only in-clinic visits were offered (*P*=.89 and *P*=.76, respectively; Table 4). The results were similar for the other efficiency measures; that is, days from discharge to final scheduling (Table 4). These results potentially reflect the decreased availability and increased difficulty in scheduling in-clinic visits in the early implementation period due to pandemic-related restrictions and concerns; however, the increased flexibility and appointment availability of teledermatology enabled comparable timeliness of follow-up scheduling during the pandemic as to before the pandemic, as reported by schedulers (Textbox 1).

Table 4. Days from inpatient discharge to initial scheduling, final scheduling, and completion of outpatient dermatological follow-up visits for patients who received an inpatient dermatology consultation and were discharged during 3 periods: baseline, early teledermatology, and sustained teledermatology.

Follow-up visit modality	Baseline (before teledermatology; June to September 2019)	Early teledermatology (June to September 2020)				Sustained teledermatology (February to May 2021)			
	In-clinic	In-clinic	*P* value[a]	Video	*P* value	In-clinic	*P* value	Video	*P* value
Patients who completed follow-up ≤90 days after discharge, N	74	39	N/A[b]	62	N/A	53	N/A	47	N/A
Days from inpatient discharge, mean (SD)									
Initial scheduling of outpatient follow-up	4.2 (7.2)	8.5 (10.2)	.01	4.3 (6.8)	.89	3.0 (4.8)	.40	3.7 (10.1)	.76
Final scheduling of outpatient follow-up	9.0 (13.5)	14.9 (17.8)	.02	6.4 (7.6)	.22	7.6 (11.2)	.51	5.7 (12.3)	.18
Completed outpatient follow-up	19.0 (14.8)	25.5 (21.9)	.19	17.3 (15.1)	.94	19.0 (17.7)	.99	18.4 (19.4)	.94

[a] Difference from baseline was determined using least squares means in generalized linear regression and was adjusted for multiple comparisons.

[b] N/A: not applicable.

Timeliness of Follow-up Visits

The average number of days from hospital discharge to completed follow-up visits did not differ across the periods or by visit modality (Table 4). A higher, although nonsignificant, proportion of patients was seen within 14 days after discharge, which is a local follow-up benchmark, via video than in the clinic (Figure 1). During the early teledermatology period, 53% (33/62) of follow-ups were conducted using teledermatology, whereas 38% (15/39) of follow-ups were conducted in the clinic (*P*=.15). Similarly, 60% (28/47) of follow-ups were conducted using teledermatology, whereas 49% (26/53) of follow-ups were conducted in the clinic (*P*=.29) during the sustained teledermatology period (Figure 1). Thus, although teledermatology follow-up visits were scheduled slightly faster (although not significant) than prepandemic in-clinic visits, this did not result in timelier follow-up care.

Incomplete Follow-up Visits

Relatively few patients missed their scheduled follow-up visits during the 3 periods (Figure 1). Of the 24 patients who missed their scheduled visit in the early teledermatology period, 12 (50% patients) had a scheduled teledermatology visit and 12 (50%) had scheduled in-clinic visits. A total of 25 patients did not complete their scheduled visit during the sustained teledermatology period, of whom 9 (36% patients) were scheduled for teledermatology and 16 (64% patients) for an in-clinic visit. This aligned with clinician and resident perceptions that teledermatology facilitated the completion of follow-up care but may not improve cancelation rates (Textbox 1). Some perceived that cancelation rates were higher for video visits, and others said they were higher for clinic visits; the small number of missed visits limits our evaluation of these perceptions.

Remaining Barriers to Video Visit Coordination

Interviewees acknowledged the benefits of teledermatology but indicated that care coordination and video visit setup were sometimes challenging (Textbox 1). Access to a smartphone and a means of taking a high-quality photograph were considered essential, especially for at-risk populations, including older adults, who were believed to benefit the most from improved access through reduced travel and risk during the pandemic. Schedulers perceived that these patients frequently needed help in setting up their devices and uploading their photographs before a visit. Despite this assistance, the photographs submitted sometimes lacked sufficient quality. Care coordination via video was further complicated if the patients required laboratory tests. Some interviewees suggested that this coordination should begin during hospitalization at the patient's bedside to integrate and prioritize the care needs of patients and caregivers in discharge planning and follow-up care scheduling.

Discussion

Principal Findings

Teledermatology was frequently used during the evaluation period; two-thirds of the visits were conducted via teledermatology early in the pandemic, whereas about half of the visits continued to be conducted using teledermatology later in the pandemic, indicating more targeted but nevertheless substantial use. Teledermatology availability increased the clinic's follow-up scheduling capacity for patients transitioning from inpatient to outpatient dermatology care. Teledermatology also provided a flexible option that increased overall clinic capacity while retaining comparable scheduling efficiency and timeliness of care as before the pandemic, even amid a pandemic and strained health care system. However, the scheduling efficiency and timeliness of care transitions did not improve, suggesting that a wider range of efforts are needed to improve these issues. Interviewees viewed teledermatology as an important care modality for providing accessible care, especially for patients with competing medical priorities and limited ability or availability to travel to the clinic, although important logistical and technological limitations were acknowledged for some patients. Ultimately, interviewees believed that patients should make the final choice between in-clinic or video visits.

Comparison With Prior Work

Teledermatology is an important tool for building clinic capacity, as well as improving scheduling timeliness and completion of care [17,18]. A study in an urban safety net hospital setting found that teledermatology implementation increased the total number of cases evaluated per month by approximately 20% and decreased the time to consultation for new patients from 84.6 days to 6.7 days before the COVID-19 pandemic [17]. Teledermatology has also been shown to increase access to and expedite care for patients in many settings, including referrals from primary to specialty care [19,20], within the Veterans Affairs system [21], in medically underserved populations [22,23], and for those needing inpatient consultations [24]. Our study builds on this literature by demonstrating a sustained increased clinic capacity of 36% after implementing teledermatology in the context of recently discharged patients needing follow-up care; the increase in clinic capacity did not come at the cost of less timely care. In addition, teledermatology services allowed safe access to care during the height of the pandemic when in-clinic care was delayed.

Nevertheless, our study was unable to detect improvements in the timeliness of care, which may be because of the urgency of our patient referrals (desired timeline from discharge to follow-up of only 14 days), as has been reported elsewhere [17-24]. Previous studies that found that video visit implementation improved care timeliness have been conducted in settings where patient referrals were nonurgent [17-24]. Although teledermatology supports increased access to outpatient care for patients of dermatology in general and those needing follow-up care after hospitalization, further research is needed to determine whether telemedicine itself supports more timely scheduling and care provision, particularly for care transitions.

Telemedicine has also been shown to promote visit completion and reduce patient cancelations and no-shows compared with in-clinic visits [16,25]. In the outpatient setting of a large academic health care system, 20% of telemedicine visits were canceled compared with 31% of in-clinic visits [16]. Similarly, a study focusing on dermatological care also showed that a lower percentage of virtual consults, specifically e-consults, were either canceled or not attended (ie, no show) than ambulatory consults (18% vs 39%) [25]. In this study, a few patients missed their scheduled follow-up, of whom 50% (12/24) missed a video visit in the early teledermatology period and 36% (9/25) missed a video visit in the sustained period. This latter result, although a different metric, suggests that recently discharged patients may be less likely to miss scheduled video visits than scheduled in-clinic visits, aligning with previous research [16,25]. However, the small sample size limits interpretability, and additional investigations are needed.

Teledermatology is well suited and highly accepted in dermatology, even for high-risk, recently discharged patients [10-14]. In this study, clinicians and scheduling staff recognized that teledermatology is convenient for patients experiencing difficulties related to their current health, with competing medical and care needs, or limited time and resources to access in-person care. In fact, even as in-clinic visits became more available, patients who lived farther from the clinic were significantly more likely to use teledermatology care than those living closer to the clinic, aligning with previous research reporting on the convenience of telemedicine for rural and underserved populations [10-14]. The widely reported flexibility of teledermatology [4,8-10] was recognized to support care transitions and continuity by interviewed clinicians and schedulers. Ultimately, telemedicine may be particularly well suited to this highly visual specialty, as reported here and elsewhere [4,8-10], perhaps even more so for follow-up care of patients recently physically examined and for whom the clinician expects but wants to confirm an improvement in their condition.

Although video visits are a well-accepted and widely used technology, some patients, clinicians, and schedulers still prefer in-person visits due to their limitations [32-34]. As reported here and previously, teledermatology continues to have shortcomings that affect care delivery, including incomplete previsit preparation, poor quality images, limited patient technological literacy, inability to access certain in-person tools and procedures, and patients’ lack of capabilities with digital devices [8,9,11,15]. Patient privacy concerns, diagnostic accuracy, and network connectivity are also well-recognized limitations [8,9,11,15]. The reported inequities in access to telemedicine care, known as the digital divide [9,11], compound the limitations of teledermatology. Although patients with complex medical issues may particularly benefit from the convenience and flexibility of teledermatology care, they may also lack access to a smartphone and the technological capacity to, for example, take and submit high-quality photographs. Thus, efforts to improve care transitions need to not only be attuned to patients who may benefit from teledermatology but also be able to assess whether such patients have access to the needed skills and technology; if not, alternatives or appropriate support for skills and technology must be provided to facilitate equitable access to care for all. Bedside communication for more patient-centered care [35] or employing dedicated care coordinator teams [28,36,37] to ensure that patients’ care needs are met could not only better support timely care transitions but also ensure high levels of patient and caregiver satisfaction, improved patient outcomes, and lower readmissions.

Limitations

This retrospective study has 3 main limitations. First, it was conducted in a single health care setting. Second, Current Procedural Terminology codes (Multimedia Appendix 1) associated with inpatient dermatology consultations were used as a surrogate to identify patients potentially needing follow-up care; ideally, such data would be based directly on clinical recommendations, including the follow-up timeline; however, such data were not systematically available. Third, 2 of the 3 study periods were during the COVID-19 pandemic. We considered many factors that could have influenced our findings, including seasonality, COVID-19 pandemic surges and restrictions, and the presence of other quality improvement initiatives, to identify comparable periods. However, this retrospective evaluation was not able to account for all potential confounding factors, including regular policy changes and vaccine availability.

Conclusions and Future Directions

Telemedicine has moved to the forefront of health care delivery and is anticipated to continue to expand. As telemedicine becomes an established care modality, additional evaluation of its quality, acceptability, and appropriateness for specific use cases and patient populations is needed to ensure the provision and sustainability of appropriate, high-quality care without continuing to widen the care access divide. Teledermatology was viewed as an important tool for maintaining accessible, flexible, and convenient care for patients transitioning from inpatient to outpatient dermatology care. Despite its shortcomings, including photograph quality and varying patient technological capabilities, teledermatology is predicted to be a standard option for patients. However, teledermatology alone does not completely solve care transition delays; it must be coupled with other efforts to improve communication between patients and care teams, patient access to and comfort with video technology, and workflows that support timely and equitable access to follow-up care. Care transitions are a vulnerable time for patients who may easily slip through the cracks and remain a challenge in health care systems [26,28,38-41]. Continued evaluation of alternate approaches to care delivery during care transitions, including telemedicine, as well as reporting of these efforts, is needed to understand their impact on this risky time in the patient care continuum.

Acknowledgments

The authors thank Melissa Dymock, MBA, formerly of Stanford Health Care, for her efforts in designing and executing the electronic health record data extraction and designing a dashboard to support the clinical team in tracking patient care transitions. This project was supported by Stanford Health Care as part of the Improvement Capability Development Program. The funder was not directly involved in the study design, data collection, data analysis, manuscript preparation, and publication decisions.

Data Availability

Data that support these findings are available from the corresponding author upon reasonable request.

Authors' Contributions

SMRK and EAS-G contributed to concept design, data collection and analysis, drafting of the manuscript, and major revisions. MW contributed to the concept design and major revisions. MAA contributed to implementation, design and concept, and revisions. DWG contributed to data collection and analysis, drafting of the manuscript, and revisions. AA contributed to data collection and analysis and revisions. CGB-J contributed to concept design and revisions. BYK, AC, and JMK contributed to implementation, concept design, and revisions. GEB contributed to data collection and revisions. JGS and SMA contributed to concept design and revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Current Procedural Terminology codes associated with inpatient dermatology to identify patients who had an inpatient dermatology consultation and who may have needed outpatient dermatology follow-up care.

[DOCX File , 13 KB - jmir_v24i8e38792_app1.docx]

References

1. 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]
2. Myers US, Birks A, Grubaugh AL, Axon RN. Flattening the curve by getting ahead of it: how the VA healthcare system is leveraging telehealth to provide continued access to care for rural veterans. J Rural Health 2021 Jan;37(1):194-196 [FREE Full text] [doi: 10.1111/jrh.12449] [Medline: 32282955]
3. Pasquali P, Romero-Aguilera G, Moreno-Ramírez D. Teledermatology before, during, and after the COVID-19 pandemic. Actas Dermosifiliogr (Engl Ed) 2021 Apr;112(4):324-329 [FREE Full text] [doi: 10.1016/j.ad.2020.11.008] [Medline: 33242450]
4. Lee JJ, English JC. Teledermatology: a review and update. Am J Clin Dermatol 2018 Apr;19(2):253-260. [doi: 10.1007/s40257-017-0317-6] [Medline: 28871562]
5. Hamad J, Fox A, Kammire MS, Hollis AN, Khairat S. Evaluating the experiences of new and existing teledermatology patients during the COVID-19 pandemic: cross-sectional survey study. JMIR Dermatol 2021 May 5;4(1):e25999 [FREE Full text] [doi: 10.2196/25999] [Medline: 34028471]

6. Perkins S, Cohen JM, Nelson CA, Bunick CG. Teledermatology in the era of COVID-19: experience of an academic department of dermatology. J Am Acad Dermatol 2020 Jul;83(1):e43-e44 [FREE Full text] [doi: 10.1016/j.jaad.2020.04.048] [Medline: 32305442]
7. Asabor EN, Bunick CG, Cohen JM, Perkins SH. Patient and physician perspectives on teledermatology at an academic dermatology department amid the COVID-19 pandemic. J Am Acad Dermatol 2021 Jan;84(1):158-161 [FREE Full text] [doi: 10.1016/j.jaad.2020.09.029] [Medline: 32946971]
8. Beer J, Hadeler E, Calume A, Gitlow H, Nouri K. Teledermatology: current indications and considerations for future use. Arch Dermatol Res 2021 Jan;313(1):11-15 [FREE Full text] [doi: 10.1007/s00403-020-02145-3] [Medline: 33074356]
9. Berman HS, Shi VY, Hsiao JL. Challenges of teledermatology: lessons learned during COVID-19 pandemic. Dermatol Online J 2020 Nov 15;26(11):13030/qt7193305r [FREE Full text] [Medline: 33342172]
10. Pathipati AS, Ko JM. Implementation and evaluation of Stanford Health Care direct-care teledermatology program. SAGE Open Med 2016 Jul 12;4:2050312116659089 [FREE Full text] [doi: 10.1177/2050312116659089] [Medline: 27493756]
11. Farr MA, Duvic M, Joshi TP. Teledermatology during COVID-19: an updated review. Am J Clin Dermatol 2021 Jul;22(4):467-475 [FREE Full text] [doi: 10.1007/s40257-021-00601-y] [Medline: 33835345]
12. Mizes A, Vainder C, Howerter SS, Hu A, Liu A, Harris A, et al. Access to consultative dermatologic care via physician-to-physician asynchronous outpatient teledermatology. Am J Manag Care 2021 Jan;27(1):30-32 [FREE Full text] [doi: 10.37765/ajmc.2021.88574] [Medline: 33471459]
13. Stadler PC, Senner S, Frey S, Clanner-Engelshofen BM, H Frommherz L, French LE, et al. Teledermatology in times of COVID-19. J Dermatol 2021 May;48(5):620-624 [FREE Full text] [doi: 10.1111/1346-8138.15812] [Medline: 33715181]
14. Mounessa JS, Chapman S, Braunberger T, Qin R, Lipoff JB, Dellavalle RP, et al. A systematic review of satisfaction with teledermatology. J Telemed Telecare 2018 May;24(4):263-270. [doi: 10.1177/1357633X17696587] [Medline: 28350281]
15. Haque W, Chandy R, Ahmadzada M, Rao B. Teledermatology after COVID-19: key challenges ahead. Dermatol Online J 2021 Apr 15;27(4):13030/qt5xr0n44p [FREE Full text] [Medline: 33999576]
16. Kubes JN, Graetz I, Wiley Z, Franks N, Kulshreshtha A. Associations of telemedicine vs. in-person ambulatory visits and cancellation rates and 30-day follow-up hospitalizations and emergency department visits. Prev Med Rep 2021 Dec;24:101629 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101629] [Medline: 34976682]
17. Zakaria A, Maurer T, Su G, Amerson E. Impact of teledermatology on the accessibility and efficiency of dermatology care in an urban safety-net hospital: a pre-post analysis. J Am Acad Dermatol 2019 Dec;81(6):1446-1452. [doi: 10.1016/j.jaad.2019.08.016] [Medline: 31415834]
18. Hsiao JL, Oh DH. The impact of store-and-forward teledermatology on skin cancer diagnosis and treatment. J Am Acad Dermatol 2008 Aug;59(2):260-267. [doi: 10.1016/j.jaad.2008.04.011] [Medline: 18485526]
19. Whited JD, Hall RP, Foy ME, Marbrey LE, Grambow SC, Dudley TK, et al. Teledermatology's impact on time to intervention among referrals to a dermatology consult service. Telemed J E Health 2002;8(3):313-321. [doi: 10.1089/15305620260353207] [Medline: 12419025]
20. Carter ZA, Goldman S, Anderson K, Li X, Hynan LS, Chong BF, et al. Creation of an internal teledermatology store-and-forward system in an existing electronic health record: a pilot study in a safety-net public health and hospital system. JAMA Dermatol 2017 Jul 01;153(7):644-650 [FREE Full text] [doi: 10.1001/jamadermatol.2017.0204] [Medline: 28423156]
21. Bezalel S, Fabri P, Park HS. Implementation of store-and-forward teledermatology and its associated effect on patient access in a Veterans Affairs dermatology clinic. JAMA Dermatol 2015 May;151(5):556-557. [doi: 10.1001/jamadermatol.2014.5272] [Medline: 25671336]
22. Naka F, Lu J, Porto A, Villagra J, Wu ZH, Anderson D. Impact of dermatology eConsults on access to care and skin cancer screening in underserved populations: a model for teledermatology services in community health centers. J Am Acad Dermatol 2018 Feb;78(2):293-302. [doi: 10.1016/j.jaad.2017.09.017] [Medline: 29061478]
23. Leavitt ER, Kessler S, Pun S, Gill T, Escobedo LA, Cockburn M, et al. Teledermatology as a tool to improve access to care for medically underserved populations: a retrospective descriptive study. J Am Acad Dermatol 2016 Dec;75(6):1259-1261 [FREE Full text] [doi: 10.1016/j.jaad.2016.07.043] [Medline: 27846951]
24. Georgesen C, Karim SA, Liu R, Moorhead A, Falo Jr LD, English 3rd JC. Inpatient eDermatology (Teledermatology) can help meet the demand for inpatient skin disease. Telemed J E Health 2020 Jul;26(7):872-878. [doi: 10.1089/tmj.2019.0147] [Medline: 31663822]
25. Wang RF, Trinidad J, Lawrence J, Pootrakul L, Forrest LA, Goist K, et al. Improved patient access and outcomes with the integration of an eConsult program (teledermatology) within a large academic medical center. J Am Acad Dermatol 2020 Dec;83(6):1633-1638. [doi: 10.1016/j.jaad.2019.10.053] [Medline: 31678336]
26. Mitchell SE, Laurens V, Weigel GM, Hirschman KB, Scott AM, Nguyen HQ, et al. Care transitions from patient and caregiver perspectives. Ann Fam Med 2018 May;16(3):225-231 [FREE Full text] [doi: 10.1370/afm.2222] [Medline: 29760026]
27. Demiris G, Kneale L. Informatics systems and tools to facilitate patient-centered care coordination. Yearb Med Inform 2015 Aug 13;10(1):15-21 [FREE Full text] [doi: 10.15265/IY-2015-003] [Medline: 26293847]

28. Berry LL, Rock BL, Smith Houskamp B, Brueggeman J, Tucker L. Care coordination for patients with complex health profiles in inpatient and outpatient settings. Mayo Clin Proc 2013 Feb;88(2):184-194. [doi: 10.1016/j.mayocp.2012.10.016] [Medline: 23290738]
29. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 2006 Sep;93(3):491-507. [doi: 10.1093/biomet/93.3.491]
30. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
31. Averill JB. Matrix analysis as a complementary analytic strategy in qualitative inquiry. Qual Health Res 2002 Jul;12(6):855-866. [doi: 10.1177/104973230201200611] [Medline: 12109729]
32. Marchell R, Locatis C, Burgess G, Maisiak R, Liu WL, Ackerman M. Patient and provider satisfaction with teledermatology. Telemed J E Health 2017 Aug;23(8):684-690 [FREE Full text] [doi: 10.1089/tmj.2016.0192] [Medline: 28375822]
33. Pearlman RL, Le PB, Brodell RT, Nahar VK. Evaluation of patient attitudes towards the technical experience of synchronous teledermatology in the era of COVID-19. Arch Dermatol Res 2021 Nov;313(9):769-772 [FREE Full text] [doi: 10.1007/s00403-020-02170-2] [Medline: 33403572]
34. Hadeler E, Gitlow H, Nouri K. Definitions, survey methods, and findings of patient satisfaction studies in teledermatology: a systematic review. Arch Dermatol Res 2021 May;313(4):205-215 [FREE Full text] [doi: 10.1007/s00403-020-02110-0] [Medline: 32725501]
35. Engaging patients in communication at transitions of care. Australian Commission on Safety and Quality in Health Care (ACSQHC). Sydney, Australia: Commonwealth of Australia; 2015. URL: https://www.safetyandquality.gov.au/sites/default/files/migrated/Engaging-Patients-in-Communication-at-Transitions.pdf [accessed 2022-07-05]
36. Cropley S, Sandrs ED. Care coordination and the essential role of the nurse. Creat Nurs 2013;19(4):189-194. [doi: 10.1891/1078-4535.19.4.189] [Medline: 24494384]
37. Doty MM, Fryer AK, Audet AM. The role of care coordinators in improving care coordination: the patient's perspective. Arch Intern Med 2012 Apr 09;172(7):587-588. [doi: 10.1001/archinternmed.2012.212] [Medline: 22412074]
38. Miceli A, Krishnamurthy K. Use of a dermatology-specific discharge form to improve outpatient follow-up after inpatient dermatology consultation. J Am Acad Dermatol 2020 Oct;83(4):1164-1166. [doi: 10.1016/j.jaad.2020.01.057] [Medline: 32007514]
39. Bumpas J, Copeland DJ. Standardizing multidisciplinary discharge planning rounds to improve patient perceptions of care transitions. J Nurs Adm 2021 Feb 01;51(2):101-105. [doi: 10.1097/NNA.0000000000000977] [Medline: 33449600]
40. Zakaria A, Chang AY, Kim-Lim P, Arakaki R, Fox LP, Amerson EH. Predictors of postdischarge follow-up attendance among hospitalized dermatology patients: disparities and potential interventions. J Am Acad Dermatol 2022 Jul;87(1):186-188. [doi: 10.1016/j.jaad.2021.07.027] [Medline: 34311040]
41. Donaho EK, Hall AC, Gass JA, Elayda MA, Lee VV, Paire S, et al. Protocol-driven allied health post-discharge transition clinic to reduce hospital readmissions in heart failure. J Am Heart Assoc 2015 Dec 23;4(12):e002296 [FREE Full text] [doi: 10.1161/JAHA.115.002296] [Medline: 26702083]

Edited by T Leung; submitted 20.04.22; peer-reviewed by R Dellavalle, J Li; comments to author 24.05.22; revised version received 14.06.22; accepted 21.06.22; published 03.08.22.

Please cite as:
Kling SMR, Saliba-Gustafsson EA, Winget M, Aleshin MA, Garvert DW, Amano A, Brown-Johnson CG, Kwong BY, Calugar A, El-Banna G, Shaw JG, Asch SM, Ko JM
Teledermatology to Facilitate Patient Care Transitions From Inpatient to Outpatient Dermatology: Mixed Methods Evaluation
J Med Internet Res 2022;24(8):e38792
URL: https://www.jmir.org/2022/8/e38792
doi:10.2196/38792
PMID:35921146

Original Paper

Factors Associated With the Utilization of Outpatient Virtual Clinics: Retrospective Observational Study Using Multilevel Analysis

Yun-Hsuan Tzeng[1,2], MD; Wei-Hsian Yin[2,3*], MD, PhD; Kuan-Chia Lin[4,5], PhD; Jeng Wei[3], MD, MSD; Hao-Ren Liou[1], BS; Hung-Ju Sung[1], BS; Hui-Chu Lang[4*], PhD

[1]Health Management Center, Cheng Hsin General Hospital, Taipei, Taiwan
[2]School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan
[3]Heart Center, Cheng Hsin General Hospital, Taipei, Taiwan
[4]Institute of Hospital and Health Care Administration, National Yang Ming Chiao Tung University, Taipei, Taiwan
[5]Community Medicine Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan
[*]these authors contributed equally

Corresponding Author:
Hui-Chu Lang, PhD
Institute of Hospital and Health Care Administration
National Yang Ming Chiao Tung University
No 155, Sec 2, Linong Street, Beitou District
Taipei, 112
Taiwan
Phone: 886 2 28267000 ext 67015
Fax: 886 2 28264536
Email: hclang@nycu.edu.tw

Abstract

Background: Although the COVID-19 pandemic has accelerated the adoption of telemedicine and virtual consultations worldwide, complex factors that may affect the use of virtual clinics are still unclear.

Objective: This study aims to identify factors associated with the utilization of virtual clinics in the experience of virtual clinic service implementation in Taiwan.

Methods: We retrospectively analyzed a total of 187,742 outpatient visits (176,815, 94.2%, in-person visits and 10,927, 5.8%, virtual visits) completed at a large general hospital in Taipei City from May 19 to July 31, 2021, after rapid implementation of virtual outpatient clinic visits due to the COVID-19 pandemic. Data of patients' demographic characteristics, disease type, physicians' features, and specialties/departments were collected, and physicians' opinions regarding virtual clinics were surveyed and evaluated using a 5-point Likert scale. Multilevel analysis was conducted to determine the factors associated with the utilization of virtual clinics.

Results: Patient-/visit-, physician-, and department-level factors accounted for 67.5%, 11.1%, and 21.4% of the total variance in the utilization of virtual clinics, respectively. Female sex (odds ratio [OR] 1.27, 95% CI 1.22-1.33, *P*<.001); residing at a greater distance away from the hospital (OR 2.36, 95% CI 2.15-2.58 if distance>50 km, *P*<.001; OR 3.95, 95% CI 3.11-5.02 if extensive travel required, *P*<.001); reimbursement by the National Health Insurance (NHI; OR 7.29, 95% CI 5.71-9.30, *P*<.001); seeking care for a major chronic disease (OR 1.33, 95% CI 1.24-1.42, *P*<.001); the physician's positive attitude toward virtual clinics (OR 1.50, 95% CI 1.16-1.93, *P*=.002); and visits within certain departments, including the heart center, psychiatry, and internal medicine (OR 2.55, 95% CI 1.46-4.46, *P*=.004), were positively associated with the utilization of virtual clinics. The patient's age, the physician's age, and the physician's sex were not associated with the utilization of virtual clinics in our study.

Conclusions: Our results show that in addition to previously demonstrated patient-level factors that may influence telemedicine use, including the patient's sex and distance from the hospital, factors at the visit level (insurance type, disease type), physician level (physician's attitude toward virtual clinics), and department level also contribute to the utilization of virtual clinics. Although there was a more than 300-fold increase in the number of virtual visits during the pandemic compared with the prepandemic period, the majority (176,815/187,742, 94.2%) of the outpatient visits were still in-person visits during the study period. Therefore, it is of great importance to understand the factors impacting the utilization of virtual clinics to accelerate the implementation of telemedicine. The findings of our study may help direct policymaking for expanding the use of virtual clinics, especially in countries struggling with the development and promotion of telemedicine virtual clinic services.

(J Med Internet Res 2022;24(8):e40288) doi:10.2196/40288

KEYWORDS

telemedicine; remote consultation; e-consult; virtual clinic; outpatient; virtual care; virtual consult; physicians; health policy; health care delivery; COVID-19; multilevel analysis; outpatient clinic; telehealth; virtual health; health care system; adoption; attitude; perception

Introduction

During the COVID-19 pandemic, countries worldwide relaxed restrictions on the utilization of telemedicine to reduce the contagion, which has resulted in the expansion of telemedicine in various clinical applications [1-3]. Owing to the severity of the disease outbreak, many countries have implemented strict strategies to restrict movement, such as lockdowns, travel constraints, and quarantine, and have implemented extensive use of telemedicine in place of in-person clinic visits in order to reduce disease exposure and the risk of infection among patients and medical staff [4,5].

Although some initial research on the provision of health care via virtual clinics (virtual consultations, teleconsultations) in the United States and the United Kingdom during the COVID-19 pandemic reported generally high patient and provider satisfaction [6-8], the acceptance of virtual visits appeared to vary widely across different subspecialties and patient populations. A previous study reported that only 32% of patients with head and neck cancer chose to have a virtual visit, even during the COVID-19 pandemic [9]. In different countries, due to differing degrees of technology penetration and COVID-19 outbreak severity, the acceptance of virtual clinics seems to vary. For example, a study conducted in Australia showed that only 61.7% of patients were satisfied with virtual visits, and less than 50% of patients expressed the desire to continue to use it in the future [10]. In addition, research on whether patient demographics are associated with the willingness to use virtual clinics has revealed conflicting results. Although some studies have shown that patients who are female and younger than 65 years of age are more likely to use digital health services [11,12], others have found that female patients are less likely to use virtual clinics [13,14]. A large study analyzing 231,596 visits across 1652 primary and specialty care practices in the United States found that patient sex is not associated with differences in the use of video visits, whereas the type of practice and clinician specialty are the main drivers of variation in telemedicine usage [15].

The inconsistency among these preliminary findings suggests that further research is needed to better identify factors that potentially impact the utilization of virtual health care. In addition, factors beyond patient demographics may also play a role in the utilization of virtual clinics, including disease chronicity, physician characteristics, the physician's attitude toward virtual clinics, and the type of subspecialty. Understanding the impact of these factors may enable policymakers and health care providers to increase patients' receptiveness to virtual health care and to expand the utilization of virtual clinics in the postpandemic era. This study aims to identify factors associated with the utilization of virtual clinics in the experience of virtual clinic service implementation in Taiwan during the COVID-19 pandemic. The findings may help direct future policy for promoting and expanding the use of virtual clinics.

Methods

Background Information and Study Design

Prior to the COVID-19 pandemic, telemedicine regulations in Taiwan were restrictive. Before 2018, only residents of outlying islands or distant mountainous areas with insufficient medical resources were allowed to use telemedicine consultations in disease diagnosis and treatment. After 2018, a few specific patients, such as overseas patients or those admitted to a family physician integrated care plan, were added to the telemedicine project. Therefore, most physicians and patients in Taiwan had never used virtual consultations prior to the COVID-19 pandemic. The Cheng Hsin General Hospital is an 800-bed hospital located in Taipei City, Taiwan. In the pre–COVID-19 pandemic time, the average volume of our outpatient clinic was approximately 100,000 visits per month. Due to restrictive prepandemic telemedicine regulations, the number of telemedicine consultations was limited at approximately 10 visits per month. During the COVID-19 pandemic, the Taiwan government relaxed telemedicine regulations, and virtual clinic visits became reimbursed under the National Health Insurance (NHI) for all patients beginning May 16, 2021. Our institution rapidly responded to the change in policy and initiated virtual outpatient clinics conducted via an integrated user-friendly smartphone application platform beginning May 19, 2021. Both virtual and in-person clinic services were made available to all our outpatients, and patients could easily book an appointment for either type of visit through the smartphone application. This special background gave us a good opportunity to test the acceptance and demand for virtual outpatient clinics by the general public and physicians of various specialties. Thus, we conducted this retrospective, cross-sectional study to determine factors associated with virtual clinic utilization.

Data Source

Data were collected from 2 sources: (1) data from the hospital information system (HIS) and the electronic medical record (EMR) system of the Cheng Hsin General Hospital and (2) results of a physician survey. The data extracted from our HIS and EMR systems contained patient age, sex, and address of residence; visit date; visit type (in-person visit or virtual visit); insurance type (whether reimbursed by the NHI); principal diagnosis codes; the in-charge physician; and the age, sex, and medical specialty/department of the physician. The design and procedures of the physician survey are described in a separate section later.

Study Sample

We collected data of all outpatient visits of the Cheng Hsin General Hospital between May 19 and July 31, 2021. This study period was chosen because it was just after implementation of our virtual clinic platform and was the peak period of virtual visits during that year. To compare the usage of virtual clinics among various specialties, including pediatrics and geriatrics, we included patients of all ages in our study. Data from a total of 197,534 outpatient visits during the study period were collected. We excluded all visits from the Department of Emergency (n=3099, 1.6%), the Department of Health Examination (n=1623, 0.8%), and visits for COVID-19 vaccination (n=5070, 2.6%) because virtual clinic services were not available in those departments. After applying the exclusion criteria, the final data set included 187,742 visits, with 176,815 (94.2%) in-person visits and 10,927 (5.8%) virtual visits completed in the outpatient department of 30 subspecialties during the study period. For in-person visits, patients came to the hospital as usual because there were no lockdown restrictions in place in Taiwan during that period. For virtual visits, physicians conducted video calls with patients at the scheduled appointment time using the integrated platform. All virtual visits were booked by patients themselves and conducted by in-charge physicians using the same smartphone application platform. Audio-only visits occurred under the condition of insufficient internet bandwidth or poor Wi-Fi signals, which resulted in a video call without screen images and only audio signals being transmitted. Since the determination of video or audio visits mainly depended on the internet condition, further subgroup analysis between video and audio visits was not performed.

Physician Survey Design

To understand physicians' opinions on the implementation of virtual clinics and to evaluate the performance of our newly introduced virtual clinic platform to identify areas for future improvement, all full-time physicians who provide outpatient clinic services at our hospital were invited to complete an online service survey. The survey was conducted between September 9 and October 6, 2021.

In the current absence of a widely validated physician survey of telemedicine that met our purpose, we designed a service-specific questionnaire modified from previously published questionnaires [6,16-18] and followed recommendations on the use of telemedicine research surveys [19]. The survey evaluated the following elements: general attitude; reliability; confidence in diagnostic and therapeutic assessment; technique-specific elements, such as audio and video quality; platform-specific elements, such as function and design; efficiency; and satisfaction (see Multimedia Appendix 1). Physicians were asked to provide answers using a 5-point Likert scale (1=strongly disagree to 5=strongly agree). Finally, we allowed for comments and suggestions. A single question regarding each physician's general attitude toward virtual clinics was recorded, and the response was analyzed for the study.

Ethical Considerations

As this was a formal service evaluation, ethical approval was not required for this study. Nonetheless, all invited physicians were fully informed verbally of the aims of this survey and understood that their responses would be analyzed for the purpose of publication. Participation was voluntary, and consent from physicians was implied by participation in the survey. This study was approved by the Institutional Review Board of the Cheng Hsin General Hospital (#(916)110-62). The need for informed written consent was waived by the board, and approval was granted for informed verbal consent prior to data collection.

Study Measures

The outcome of interest in our study was visit type (in-person visit vs virtual visit). For patient-level variables, we included patient characteristics that have been previously demonstrated to influence telemedicine use, including age, sex, and distance between the patient's residence and the hospital. Sociodemographic variables, such as the marital status, highest education level, and income of the patient, were not included, because updated data of this type were not available in our HIS database. In addition, no data were available in our HIS database regarding the patients' race/ethnicity or spoken language. The distance between a patient's place of residence and the hospital was estimated using the patient address' zip code and then grouped as a categorical variable (<20 km, 20-50 km, >50 km, outlying islands, and traveled >5 hours to reach the hospital). We also collected visit-level variables, including insurance/reimbursement type and the coding of principal diagnosis based on the *International Classification of Diseases 10th Revision* (ICD-10) from our HIS and EMR databases.

To investigate the association between disease type and the usage of virtual clinics and to examine the association of major chronic diseases with virtual clinic service, we used the classifications of chronic diseases defined by the Ministry of Health and Welfare of Taiwan and the chronic condition indicator for ICD-10 developed by the Agency for Healthcare Research and Quality of America [20] to categorize the diseases. As defined by the Ministry of Health and Welfare of Taiwan, prescription refills are allowed for 101 chronic diseases in 16 categories (Multimedia Appendix 2). First, we removed certain disease groups that may involve complicated disease conditions, diverse prognoses, and various purposes for visits, including malignant neoplasm, brain tumor, polyneuropathy, nerve root and plexus disorders, trigeminal neuralgia, spinal cord injury, peptic ulcer, colitis, cholangitis, nephritis, arthritis, dermatomyositis, osteomyelitis, osteoporosis, autoimmune disease, ocular disease, skin diseases, ear diseases, blood diseases, prostate and urination diseases, infectious diseases, congenital malformations, hemorrhoids, follow-up after organ transplantation, and menopause syndrome. Next, we removed certain diseases that had few cases in our study cohort, including endometriosis, leprosy, blackfoot disease, and polychlorinated biphenyl intoxication. Finally, we defined 10 types of major chronic and stable diseases (Table 1) to compare with other diseases. The data of the dictionary of the specific ICD-10 codes used to categorize these diseases are provided in Multimedia Appendix 3. For physician-level variables, the physician's age and sex were obtained from the HIS database and the physician's general attitude toward virtual clinics was obtained from the physician survey, as described before. Department-level variables, such as the medical specialty/department of the visit, were collected from our HIS database as well.

Table 1. Major chronic diseases according to ICD-10[a].

Disease group	Visits (N=187,742), n (%)
1. Diabetes mellitus	30,680 (16.3)
2. Coronary artery disease	19,517 (10.4)
3. Hypertension	14,034 (7.5)
4. Chronic cardiac and arterial disease	11,387 (6.1)
5. Psychiatric disease and sleep disorder	9887 (5.3)
6. Cerebrovascular disease and other chronic neurologic diseases	7477 (4.0)
7. Chronic respiratory disease	5196 (2.8)
8. Chronic liver disease	3533 (1.9)
9. Thyroid and endocrine diseases	3408 (1.8)
10. Hyperlipidemia	2791 (1.5)

[a] ICD-10: International Classification of Diseases 10th Revision.

Statistical Analysis

Descriptive statistics were used to assess trends in the use of virtual visits and in-person visits. Continuous variables were described as the mean (SD). Categorical variables were described using frequencies and percentages.

Group comparisons (virtual visit vs in-person visit) were tested for differences using the Student t-test for continuous variables and the chi-square test for categorical variables. To determine the independent factors associated with the utilization of virtual clinics, all variables exhibiting a *P* value of <.01 on univariate analysis were entered into a multivariate binary logistic regression and a multilevel analysis. Multilevel analysis was conducted by using 3-level structure hierarchical linear modeling to incorporate variables at the patient/visit level, physician level, and department level in a statistically correct way.

All statistical analyses were carried out using commercially available software (IBM SPSS Statistics for Windows, version 28.0; IBM Corporation). The multilevel analysis was carried out using HLM version 8.2 (Scientific Software International). A two-sided *P* value of <.01 was considered statistically significant for all analyses.

Results

Visit Characteristics

Characteristics of all visits recorded during the study period are summarized in Table 2. Of 187,742 total visits, 10,927 (5.8%) were virtual visits during the study period. The mean age of patients in all visits was 61.48 (SD 16.86) years, and 96,884 (51.6%) visits were of female patients. In terms of the distance of the patients' residence from the hospital, 168,846 (89.9%), 12,623 (6.7%), and 5284 (2.8%) visits were by patients who lived <20 km, 20-50 km, and >50 km from the hospital, respectively. In addition, 449 (0.3%) visits were by patients who needed to travel extensively to reach the hospital, including 167 (37.2%) visits by patients who lived in the outlying islands of Taiwan and 332 (62.8%) visits by patients who traveled >5 hours to reach the hospital. Nearly all (n=175,881, 93.7%) of the visits were reimbursed by the NHI. For disease type, 107,910 (57.5%) visits were related to a major chronic disease. Regarding specialty, 129,504 (69%) visits were conducted by the heart center, the psychiatry department, or the internal medicine department.

Table 2. Characteristics of all visits.

Characteristics	Total visits (N=187,742)	In-person visits (n=176,815)	Virtual visits (n=10,927)	Virtual visit rate (%)= virtual visits/total visits
Age (years), mean (SD); *P*=.013	61.48 (16.86)	61.50 (16.84)	61.09 (17.22)	N/A[a]
Sex, n (%); *P*<.001				
Male	90,863 (48.4)	85,951(48.6)	4907 (44.9)	5.4
Female	96,884 (51.6)	90,864 (51.4)	6020 (55.1)	6.2
Distance, n (%); *P*<.001				
<20 km	168,846 (89.9)	159,787 (90.4)	9059 (82.9)	5.4
20-50 km	12,623 (6.7)	11,552 (6.5)	1071 (9.8)	8.5
>50 km	5284 (2.8)	4607 (2.6)	677 (6.2)	12.8
Outlying islands	167 (0.1)	133 (0.1)	34 (0.3)	20.4
Traveled >5 hours to reach the hospital	332 (0.2)	263 (0.1)	69 (0.6)	26.2
Unknown	490 (0.3)	473 (0.3)	17 (0.2)	3.5
Insurance type, n (%); *P*<.001				
Reimbursed by the NHI[b]	175,881 (93.7)	165,021 (93.3)	10,860 (99.4)	6.2
Nonreimbursed by the NHI	11,861 (6.3)	11,794 (6.7)	67 (0.6)	0.6
Disease type, n (%); *P*<.001				
Major chronic diseases	107,910 (57.5)	99,624 (56.3)	8286 (75.8)	7.7
Other diseases	79,832 (42.5)	77,191 (43.7)	2641 (24.2)	3.3
Department, n (%); *P*<.001				
Heart center, psychiatry department, and internal medicine department	129,504 (69.0)	120,088 (67.9)	9416 (86.2)	7.3
Other departments	58,238 (31.0)	56,727 (32.1)	1511 (13.8)	2.6

[a] N/A: not applicable.

[b] NHI: National Health Insurance.

Physician Characteristics

Of 174 invited physicians, 165 (94.8%) responded to the survey, accounting for 179,857 (95.8%) of 187,742 outpatient visits during the study period. These physicians were from 30 subspecialties of 13 departments of our hospital. The characteristics of the 165 physicians who responded to the virtual clinic service survey are summarized in Table 3. Their mean age was 55.58 (SD 11.89) years, 25 (15.2%) physicians were female, and 115 (69.7%) physicians expressed a positive attitude toward virtual clinics by agreeing or strongly agreeing (Likert scale score≥4) that virtual clinics are practical and that they are willing to conduct virtual visits.

Table 3. Physician characteristics.

Characteristics	Participants (N=165)
Sex (female), n (%)	25 (15.2)
Age (years), mean (SD)	55.58 (11.89)
Attitude (Likert scale score≥4), n (%)	115 (69.7)
Department, n (%)	
Internal medicine	44 (26.7)
Surgery	37 (22.4)
Heart center	27 (16.4)
Obstetrics and gynecology	9 (5.5)
Oncology and radiotherapy	8 (4.9)
Psychiatry	8 (4.9)
Rehabilitation	6 (3.6)
Otorhinolaryngology	6 (3.6)
Ophthalmology	6 (3.6)
Dentistry	4 (2.4)
Dermatology	4 (2.4)
Pediatrics	4 (2.4)
Traditional Chinese medicine	2 (1.2)

Factors Associated With Virtual Clinic Utilization

In univariate analysis, the percentage of virtual clinic use was higher in female patients (6020/96,884, 6.2%, female patients vs 4907/90863, 5.4%, male patients, *P*<.001), in visits for major chronic diseases (8286/107,910, 7.7%, major chronic diseases vs 2641/79,832, 3.3%, nonmajor chronic diseases, *P*<.001), in visits reimbursed by the NHI (10,860/175,881, 6.2%, reimbursed visits vs 67/11,861, 0.6%, nonreimbursed visits, *P*<.001), and in visits performed by the heart center, the psychiatry department, or the internal medicine department (9416/129,504, 7.3%, visits in these departments vs 1511/58,238, 2.6%, visits not in these departments, *P*<.001). Patients who lived farther away from the hospital were more likely to use virtual clinic, with 69 (26.2%) of 332 visits by patients who needed more than 5 hours of travel time to reach the hospital, 34 (20.4%) of 167 visits by patients who lived in outlying islands of Taiwan, 677 (12.8%) of 5284 visits by patients who lived more than 50 km from the hospital, 1071 (8.5%) of 12,623 visits by patients who lived 20-50 km from the hospital, and 9059 (5.4%) of 168,846 visits by patients who lived within 20 km from the hospital (*P*<.001). There was no significant difference in the mean patient age between the in-person visit group (mean 61.50 years, SD 16.84 years) and the virtual visit group (mean 61.09 years, SD 17.22 years; *P*=.013).

Results of the 3-level structure multilevel analysis are shown in Table 4. The random part of the model represents the variance at each hierarchical level. Based on the formula of the intraclass correlation coefficient [21], physician-level factors accounted for 11.1% of the total variance and department-level factors accounted for 21.4% of the total variance in utilization of virtual clinics. Patient-/visit-level factors contributed to 67.5% of the total variance.

In multilevel analysis, patient-/visit-level factors associated with the utilization of virtual clinics included female sex (odds ratio [OR] 1.27, 95% CI 1.22-1.33, *P*<.001), distance from the hospital (OR 2.36, 95% CI 2.15-2.58 if distance>50 km, *P*<.001; OR 3.95, 95% CI 3.11-5.02 if extensive travel required, *P*<.001), visit for a major chronic disease (OR 1.33, 95% CI 1.24-1.42, *P*<.001), and visit reimbursed by the NHI (OR 7.29, 95% CI 5.71-9.30, *P*<.001). The physician's age and sex were not associated with the utilization of virtual clinics. The only physician-level factor associated with the utilization of virtual clinics was the physician's positive attitude toward virtual clinics (OR 1.50, 95% CI 1.16-1.93, *P*=.002). Department-level factors as represented by the different specialties were associated with the utilization of virtual clinics. The heart center, psychiatry department, and internal medicine department were more likely to use virtual clinics (OR 2.55, 95% CI 1.46-4.46, *P*=.004).

As shown in Table 4, combining all variables of different levels into a binary logistic regression model revealed that the physician's age and sex were significantly associated with the utilization of virtual clinics. Therefore, in our study, the results may be skewed if binary logistic regression analysis were used, and the role of physician- and department-level factors in the utilization of virtual clinics (ie, factors beyond the patient/visit level) could not be assessed with such a model.

Table 4. Results of the multilevel model.

Variables	Logistic regression			3-level model			
	β	SE	*P* value	β	SE	OR[a] (95% CI)	*P* value
Fixed effect level 3							
Intercept	–5.322	0.144	<.001	–5.323	0.355	0.005 (0.002-0.011)	<.001
Heart center, psychiatry department, and internal medicine department (yes/no)	0.630	0.035	<.001	0.936	0.254	2.550 (1.458-4.460)	.004
Fixed effect level 2							
Physician’s sex, female (reference male)	0.119	0.031	<.001	0.193	0.180	1.214 (0.850-1.730)	.28
Physician’s age	–0.004	0.001	<.001	–0.008	0.0050	0.992 (0.982-1.003)	.15
Physician’s attitude, Likert scale score≥4 (yes/no)	0.234	0.024	<.001	0.404	0.129	1.498 (1.162-1.932)	.002
Fixed effect level 1							
Patient’s sex, female (reference male)	0.228	0.020	<.001	0.242	0.020	1.274 (1.222-1.326)	<.001
Insurance type, reimbursed by the NHI[b] (yes/no)	2.030	0.130	<.001	1.986	0.124	7.288 (5.711-9.300)	<.001
Distance (reference <20 km and 20-50 km)							
>50 km	0.955	0.043	<.001	0.857	0.046	2.355 (2.153-2.577)	<.001
Traveled extensively to reach the hospital	1.505	0.115	<.001	1.373	0.122	3.948 (3.108-5.016)	<.001
Disease type, major chronic diseases (yes/no)	0.445	0.029	<.001	0.282	0.036	1.326 (1.236-1.421)	<.001
Random effect							
Level 3	N/A[c]	N/A	N/A	1.040	1.020	N/A	<.001
Level 2	N/A	N/A	N/A	0.541	0.736	N/A	<.001
Level 1	N/A	N/A	N/A	$\pi^2/3$	N/A	N/A	N/A

[a] OR: odds ratio.

[b] NHI: National Health Insurance.

[c] N/A: not applicable.

Discussion

Principal Findings

In our study, results of the multilevel analysis showed that factors at different levels all contributed to the utilization of virtual clinics. In addition to patient-/visit-level factors (67.5%), physician-level (11.1%) and department-level (21.4%) factors also drove variation in virtual clinic use. Patients who were female and lived farther away from the hospital were more likely to use virtual clinics, whereas the patient’s age did not affect the utilization of virtual clinics in our study. Visit-level variables, including insurance type (reimbursed by the NHI) and disease type (major chronic disease), were positively associated with the utilization of virtual clinics. The physician’s positive attitude toward virtual clinics positively predicted the use of virtual clinics, while the physician’s sex and age were not major predictors. Certain medical departments/specialties (the heart center, the psychiatry department, and the internal medicine department) were positively associated with the utilization of virtual clinics. Even when in-person visits and virtual visits were equally available, the use of virtual visits was relatively low, accounting for 10,927 (5.8%) of 187,742 visits in our study period.

Comparison With Existing Literature

During the COVID-19 pandemic, a massive migration from in-person to virtual clinic visits was observed in many countries [2,4,22-24]. In prior reports from New York City, the epicenter of the pandemic during 2020, Mann et al [4] reported that telemedicine visits in a large academic health care system increased from less than 50 daily to more than 1000 daily, co-occurring with a decline of over 80% in in-person visits [4], and Ramaswamy et al [7] reported an 8729% increase in video visit utilization [7]. Various countries, including Italy, the United Kingdom, and India, reported virtual migration percentages between 60% and 95% of their usual practice [2,22]. However, during our study period, the COVID-19 outbreak in Taiwan was well controlled. No lockdown or shelter-in-place orders were enacted, and usual hospital outpatient services were not impacted. The use of virtual clinics was for individual health demand rather than for COVID-19–suspected diagnosis during our study period. Therefore, our hospital has not experienced a massive migration from in-person to virtual visits. Although there was a more than a 300-fold increase in the number of virtual visits compared to the prepandemic period, the majority (176,815/187,742, 94.2%) visits in our outpatient clinic were still in-person visits during the study period. This finding may imply that when both types of clinics are equally available, most patients still prefer in-person visits.

Earlier studies have reported that virtual clinics are associated with high patient/physician satisfaction [6-8,25,26], and findings have supported the effectiveness of virtual consultations in various practices [27-34]. Although teleconsultations or virtual clinics are not novel concepts, the growth of this type of health care service was slow in Taiwan owing to restrictive regulations prior to the COVID-19 pandemic. Our findings show that the utilization of virtual outpatient clinics was still limited even in the modern, high-income, capital city of Taiwan with a high penetration rate of smartphone use and broadband internet, implying that telemedicine may still have a long way to go to be widely accepted. Therefore, it is of great importance to investigate the factors impacting the utilization of virtual clinics in order to accelerate the implementation of telemedicine, especially in countries in which telemedicine services are currently underused.

Older age has been widely recognized as a barrier to adopting telemedicine in previous research [11,12,35-37]. However, no significant difference in patient age between the virtual visit and in-person visit groups was found in our study. Interestingly, some previous studies have found that age does not have a significant influence on a patient's willingness to conduct telemedicine consultations [38,39], which may explain our finding. There are several additional possible reasons for our finding. First, as is common in Asian cultures, elderly patients are commonly cared for by family members who can provide assistance in using the telemedicine platform. Second, since chronic diseases were shown to be positively associated with the use of virtual clinics in our study, the higher prevalence of chronic diseases in the elderly may have contributed to their seeking of virtual clinic services. Finally, with the advances of smartphone technology, telemedicine is easier to use and more accessible than ever. Prior studies have shown that smartphone device usage is high even in people of older age [37,40]. For elderly people, lack of appropriate equipment and lack of exposure to new technology have been identified as significant barriers to adopting telemedicine [35]; therefore, the ease of use of the smartphone platform used in our study may be beneficial for elderly patients to complete a virtual visit.

Although patient sex was not associated with the utilization of telemedicine in some prior research [6,9,15], a few studies did show that being female is negatively associated with the success rate or satisfaction of a video visit. Eberly et al [13,14] reported that female patients are less likely to complete video visits, and Ramaswamy et al [7] found that female patients have lower satisfaction with video visits compared to male patients. Our study found that female patients are more likely than male patients to use virtual clinics. The same finding has been reported in some previous research that showed female sex as a positive predictor of digital health engagement behaviors and telemedicine consultations [11,36,41].

Prior studies have shown that time- and cost-saving benefits of telemedicine consultations are largely affected by distance from the hospital or clinic [30,42-44]. Cannon et al [45] found that for every 23 miles a patient resides from their clinic, patients are 111% more likely to use telemedicine consultation [45]. Our study showed that patients who live farther away from the hospital are more likely to replace in-person visits with virtual visits, which is consistent with previous research. We also found that patients with chronic and stable diseases are more likely to seek care by using virtual clinics. Since patients with chronic diseases and those who need to travel extensively to access health care are more likely to be vulnerable individuals, the implementation of virtual clinics is particularly meaningful and beneficial for these patient groups. Virtual health care may be even more essential during the COVID-19 pandemic in protecting these patients from disease exposure and decreasing the risk of infection.

Previous studies in the United States have shown that patients who use video visits are more likely to be White, have private health insurance, and have a higher level of income [13,15,36]. However, since the health care system is quite different in Taiwan, having private insurance was not a positive predictor of the utilization of virtual clinics in our study. In Taiwan, almost all citizens are covered by the NHI, and private insurance seldom covers the expense of outpatient visits. Interestingly, we found that patients are less likely to use virtual visits if the expense is paid out of pocket, which suggests that if the medical expense related to a visit is to be self-paid, patients would prefer an in-person visit with the physician.

Provider resistance has been reported by studies from Ethiopia, the United Kingdom, Australia, Iran, and the United States as a barrier to adopting telemedicine [35]. In our study, the physician's age and sex did not influence the utilization of virtual clinics, but the physician's positive attitude toward virtual clinics did (OR 1.50, P=.002), which supported the findings of those prior studies. We also found that in our institution, physicians of specific departments and subspecialties, including the Department of Internal Medicine, the Department of Psychiatry, and the heart center, are more likely to utilize virtual clinics than others. Since physician-level factors accounted for 11.1% and department-level factors accounted for 21.4% of the total variance in the utilization of virtual clinics, an increase in the physicians' acceptance of virtual clinic services, particularly for physicians of certain departments, is important to further expand such service.

Limitations and Strengths

This study had some limitations. First, demographic characteristics, including marital status, income, race/ethnicity, spoken language, and educational level, were not available in our database, so we were unable to capture the influence of these sociodemographic variables. However, although patients' race/ethnicity and spoken language are common factors of patient inequities and disparities in telemedicine adoption, those factors may be not significant in Taiwanese society, as nearly all patients in Taiwan are ethnic Chinese. In addition, some prior studies examining various types of digital health utilization did not observe disparities in race/ethnicity [12,46].

Second, the practice and acceptability of telemedicine and virtual clinics may be strongly associated with the health care system structure and the pervasiveness of technology. The health care system of Taiwan is quite different from that in Western countries. A major difference is the lack of a well-established general practice and referral system in Taiwan. In large hospitals, physicians of various specialties also take on the role of primary care providers and have large volumes of outpatient visits for chronic diseases, such as diabetes, hypertension, hyperlipidemia, and coronary artery disease. The results of our study reflect this situation, with up to 107,910 (57.5%) of 187,742 visits during our study period having a principal diagnosis of a major chronic disease. However, our results may serve as a valuable reference for countries having a similar health care system structure, especially in East Asia.

Third, the study data were collected from a single hospital, limiting generalizability to different types of practices and organizations. However, to increase the relevance of study findings outside of the current institutional setting and practice, a multilevel analytic methodology was emphasized, which can be applied to different study cohorts and practice settings.

Despite these limitations, our study is unique in the following ways: First, unlike most of the virtual clinic studies conducted during the COVID-19 pandemic with lockdown or shelter-in-place orders and restrictions of regular medical services [6,7,12-15,36,47], our study was conducted with the background that virtual clinics and in-person clinics were both equally and easily accessible and the utilization of virtual clinics was not largely affected by the pandemic. Second, unlike most prior studies that focused only on patient-level variables that may affect the use of virtual clinics [12,13,36,37,48], we used multilevel analyses to show that insurance type, disease type, physician's attitude, and specialties are associated with the utilization of virtual clinics independent of patient demographic characteristics. Third, many prior telemedicine studies were small and focused on specific patient populations of 1 specific medical specialty [6,9,10,13,25]. Our study was performed across various specialties and departments, and the patient volume of each virtual clinic was controlled by individual physicians; therefore, physician-level variables could be investigated. Finally, all our virtual visits were conducted using a user-friendly smartphone application platform; thus, we could greatly reduce the impact of technology-related barriers that may strongly impact the utilization of virtual clinics.

Conclusion

In conclusion, our study demonstrates that factors at the patient/visit level, physician level, and department level all contribute to the utilization of virtual clinics. Female sex; residing at a greater distance away from the hospital; reimbursement by the NHI; seeking care for a major chronic disease; the physician's positive attitude toward virtual clinics; and visits within certain departments, including the heart center, psychiatry, and internal medicine, were positively associated with the utilization of virtual clinics in our study. The findings may help direct future policy for expanding the use of virtual clinics, especially in countries struggling with the development and promotion of telemedicine virtual clinic services. Further studies should be conducted to evaluate the trends in and utilization of virtual clinics in the postpandemic period in different countries and health care systems, to examine the effectiveness and acceptability of telemedicine as a routine alternative practice to in-person clinics, and to investigate the economic impact of telemedicine from the perspective of the provider, the health care system, and society.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire of the physician survey of virtual clinic service.
[DOCX File , 21 KB - jmir_v24i8e40288_app1.docx]

Multimedia Appendix 2
Classifications of chronic diseases defined by the Ministry of Health and Welfare of Taiwan.
[DOCX File , 29 KB - jmir_v24i8e40288_app2.docx]

Multimedia Appendix 3
Dictionary of specific ICD-10 codes used to categorize the major chronic diseases in this study.
[XLSX File (Microsoft Excel File), 85 KB - jmir_v24i8e40288_app3.xlsx]

References

1. Bhaskar S, Bradley S, Chattu VK, Adisesh A, Nurtazina A, Kyrykbayeva S, et al. Telemedicine across the globe: position paper from the COVID-19 Pandemic Health System Resilience PROGRAM (REPROGRAM) International Consortium (part 1). Front Public Health 2020;8:556720 [FREE Full text] [doi: 10.3389/fpubh.2020.556720] [Medline: 33178656]
2. Hincapié MA, Gallego JC, Gempeler A, Piñeros JA, Nasner D, Escobar MF. Implementation and usefulness of telemedicine during the COVID-19 pandemic: a scoping review. J Prim Care Community Health 2020;11:2150132720980612 [FREE Full text] [doi: 10.1177/2150132720980612] [Medline: 33300414]
3. Haider Z, Aweid B, Subramanian P, Iranpour F. Telemedicine in orthopaedics during COVID-19 and beyond: a systematic review. J Telemed Telecare 2022 Jul 06;28(6):391-403 [FREE Full text] [doi: 10.1177/1357633X20938241] [Medline: 32762270]

4. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
5. Fu R, Sutradhar R, Li Q, Eskander A. Virtual and in-person visits by Ontario physicians in the COVID-19 era. J Telemed Telecare 2022 Mar 16:1357633X2210864. [doi: 10.1177/1357633x221086447]
6. Hobson S, Aleem IS, Bice MJ, Butt BB, Bydon M, Elder BD, et al. A multicenter evaluation of the feasibility, patient/provider satisfaction, and value of virtual spine consultation during the COVID-19 pandemic. World Neurosurg 2021 Oct;154:e781-e789 [FREE Full text] [doi: 10.1016/j.wneu.2021.08.004] [Medline: 34389525]
7. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. J Med Internet Res 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: 10.2196/20786] [Medline: 32810841]
8. Gillman-Wells CC, Sankar TK, Vadodaria S. COVID-19 reducing the risks: telemedicine is the new norm for surgical consultations and communications. Aesthetic Plast Surg 2021 Feb;45(1):343-348 [FREE Full text] [doi: 10.1007/s00266-020-01907-8] [Medline: 32885319]
9. Fassas S, Cummings E, Sykes KJ, Bur AM, Shnayder Y, Kakarala K. Telemedicine for head and neck cancer surveillance in the COVID-19 era: promise and pitfalls. Head Neck 2021 Jun;43(6):1872-1880 [FREE Full text] [doi: 10.1002/hed.26659] [Medline: 33660409]
10. Adams L, Lester S, Hoon E, van der Haak H, Proudman C, Hall C, et al. Patient satisfaction and acceptability with telehealth at specialist medical outpatient clinics during the COVID-19 pandemic in Australia. Intern Med J 2021 Jul;51(7):1028-1037 [FREE Full text] [doi: 10.1111/imj.15205] [Medline: 34213046]
11. Ratcliff CL, Krakow M, Greenberg-Worisek A, Hesse BW. Digital health engagement in the US population: insights from the 2018 Health Information National Trends Survey. Am J Public Health 2021 Jul;111(7):1348-1351. [doi: 10.2105/AJPH.2021.306282] [Medline: 34014759]
12. Nguyen OT, Watson AK, Motwani K, Warpinski C, McDilda K, Leon C, et al. Patient-level factors associated with utilization of telemedicine services from a free clinic during COVID-19. Telemed J E Health 2022 Apr;28(4):526-534. [doi: 10.1089/tmj.2021.0102] [Medline: 34255572]
13. Eberly LA, Khatana SAM, Nathan AS, Snider C, Julien HM, Deleener ME, et al. Telemedicine outpatient cardiovascular care during the COVID-19 pandemic: bridging or opening the digital divide? Circulation 2020 Aug 04;142(5):510-512 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.120.048185] [Medline: 32510987]
14. 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]
15. Rodriguez JA, Betancourt JR, Sequist TD, Ganguli I. Differences in the use of telephone and video telemedicine visits during the COVID-19 pandemic. Am J Manag Care 2021 Jan;27(1):21-26 [FREE Full text] [doi: 10.37765/ajmc.2021.88573] [Medline: 33471458]
16. Zhu C, Williamson J, Lin A, Bush K, Hakim A, Upadhyaya K, et al. Implications for telemedicine for surgery patients after COVID-19: survey of patient and provider experiences. Am Surg 2020 Aug 17;86(8):907-915. [doi: 10.1177/0003134820945196] [Medline: 32805123]
17. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the Telemedicine Satisfaction Questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]
18. Parmanto B, Lewis AN, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
19. Langbecker D, Caffery LJ, Gillespie N, Smith AC. Using survey methods in telehealth research: a practical guide. J Telemed Telecare 2017 Oct;23(9):770-779. [doi: 10.1177/1357633X17721814] [Medline: 28728502]
20. Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP): Chronic Condition Indicators for ICD-10-CM (beta version). 2020. URL: http://www.hcup-us.ahrq.gov/toolssoftware/chronic_icd10/chronic_icd10.jsp [accessed 2022-08-01]
21. Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health 2006 Apr;60(4):290-297 [FREE Full text] [doi: 10.1136/jech.2004.029454] [Medline: 16537344]
22. 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;71(704):e166-e177 [FREE Full text] [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
23. Blue R, Yang AI, Zhou C, De Ravin E, Teng CW, Arguelles GR, et al. Telemedicine in the era of coronavirus disease 2019 (COVID-19): a neurosurgical perspective. World Neurosurg 2020 Jul;139:549-557 [FREE Full text] [doi: 10.1016/j.wneu.2020.05.066] [Medline: 32426065]

24. Hemingway JF, Singh N, Starnes BW. Emerging practice patterns in vascular surgery during the COVID-19 pandemic. J Vasc Surg 2020 Aug;72(2):396-402 [FREE Full text] [doi: 10.1016/j.jvs.2020.04.492] [Medline: 32361072]
25. Melian C, Frampton C, Wyatt MC, Kieser D. Teleconsultation in the management of elective orthopedic and spinal conditions during the COVID-19 pandemic: prospective cohort study of patient experiences. JMIR Form Res 2021 Jun 15;5(6):e28140 [FREE Full text] [doi: 10.2196/28140] [Medline: 34048355]
26. Orrange S, Patel A, Mack WJ, Cassetta J. Patient satisfaction and trust in telemedicine during the COVID-19 pandemic: retrospective observational study. JMIR Hum Factors 2021 Apr 22;8(2):e28589 [FREE Full text] [doi: 10.2196/28589] [Medline: 33822736]
27. Lin JC, Mclaughlin D, Zurawski D, Kennedy N, Kabbani L. Comparison of virtual visit versus traditional clinic for management of varicose veins. J Telemed Telecare 2020 Sep 20;26(1-2):100-104. [doi: 10.1177/1357633X18797181] [Medline: 30235968]
28. Zholudev V, Safir IJ, Painter MN, Petros JA, Filson CP, Issa MM. Comparative cost analysis: teleurology vs conventional face-to-face clinics. Urology 2018 Mar;113:40-44. [doi: 10.1016/j.urology.2017.07.034] [Medline: 28780298]
29. De Guzman KR, Snoswell CL, Caffery LJ, Smith AC. Economic evaluations of videoconference and telephone consultations in primary care: a systematic review. J Telemed Telecare 2021 Oct 07:1357633X211043380. [doi: 10.1177/1357633X211043380] [Medline: 34617819]
30. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of telemedicine in remote orthopedic consultations: randomized controlled trial. J Med Internet Res 2019 Feb 19;21(2):e11330 [FREE Full text] [doi: 10.2196/11330] [Medline: 30777845]
31. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. Lancet 2017 Sep 02;390(10098):959-968. [doi: 10.1016/S0140-6736(17)31327-2] [Medline: 28716313]
32. Fortney JC, Pyne JM, Kimbrell TA, Hudson TJ, Robinson DE, Schneider R, et al. Telemedicine-based collaborative care for posttraumatic stress disorder: a randomized clinical trial. JAMA Psychiatry 2015 Jan;72(1):58-67. [doi: 10.1001/jamapsychiatry.2014.1575] [Medline: 25409287]
33. 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]
34. Viers BR, Lightner DJ, Rivera ME, Tollefson MK, Boorjian SA, Karnes RJ, et al. Efficiency, satisfaction, and costs for remote video visits following radical prostatectomy: a randomized controlled trial. Eur Urol 2015 Oct;68(4):729-735. [doi: 10.1016/j.eururo.2015.04.002] [Medline: 25900782]
35. 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]
36. Luo J, Tong L, Crotty BH, Somai M, Taylor B, Osinski K, et al. Telemedicine adoption during the COVID-19 pandemic: gaps and inequalities. Appl Clin Inform 2021 Aug;12(4):836-844. [doi: 10.1055/s-0041-1733848] [Medline: 34496419]
37. Scherer J, Osterhoff G, Kaufmann E, Estel K, Neuhaus V, Willy C, et al. What is the acceptance of video consultations among orthopedic and trauma outpatients? A multi-center survey in 780 outpatients. Injury 2021 Nov;52(11):3304-3308 [FREE Full text] [doi: 10.1016/j.injury.2021.02.023] [Medline: 33648741]
38. Almathami HKY, 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]
39. Roettl J, Bidmon S, Terlutter R. What predicts patients' willingness to undergo online treatment and pay for online treatment? Results from a web-based survey to investigate the changing patient-physician relationship. J Med Internet Res 2016 Feb 04;18(2):e32 [FREE Full text] [doi: 10.2196/jmir.5244] [Medline: 26846162]
40. Pozza ED, D'Souza GF, DeLeonibus A, Fabiani B, Gharb BB, Zins JE. Patient satisfaction with an early smartphone-based cosmetic surgery postoperative follow-up. Aesthet Surg J 2017 Dec 13;38(1):101-109. [doi: 10.1093/asj/sjx079] [Medline: 29117293]
41. 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]
42. Paquette S, Lin JC. Outpatient telemedicine program in vascular surgery reduces patient travel time, cost, and environmental pollutant emissions. Ann Vasc Surg 2019 Aug;59:167-172. [doi: 10.1016/j.avsg.2019.01.021] [Medline: 31077768]
43. Philips R, Seim N, Matrka L, Locklear B, Moberly AC, Inman M, et al. Cost savings associated with an outpatient otolaryngology telemedicine clinic. Laryngoscope Investig Otolaryngol 2019 Apr 14;4(2):234-240 [FREE Full text] [doi: 10.1002/lio2.244] [Medline: 31024993]
44. Dullet NW, Geraghty EM, Kaufman T, Kissee JL, King J, Dharmar M, et al. Impact of a university-based outpatient telemedicine program on time savings, travel costs, and environmental pollutants. Value Health 2017 Apr;20(4):542-546 [FREE Full text] [doi: 10.1016/j.jval.2017.01.014] [Medline: 28407995]

45. Canon S, Shera A, Patel A, Zamilpa I, Paddack J, Fisher PL, et al. A pilot study of telemedicine for post-operative urological care in children. J Telemed Telecare 2014 Dec;20(8):427-430. [doi: 10.1177/1357633X14555610] [Medline: 25316038]
46. Lyerly MJ, Wu T, Mullen MT, Albright KC, Wolff C, Boehme AK, et al. The effects of telemedicine on racial and ethnic disparities in access to acute stroke care. J Telemed Telecare 2016 Mar;22(2):114-120 [FREE Full text] [doi: 10.1177/1357633X15589534] [Medline: 26116854]
47. Sallam A, Shang M, Vallabhajosyula I, Mori M, Chinian R, Assi R, et al. Telemedicine in the era of coronavirus 19: implications for postoperative care in cardiac surgery. J Card Surg 2021 Oct;36(10):3731-3737 [FREE Full text] [doi: 10.1111/jocs.15875] [Medline: 34338360]
48. Sankaran J, Menachery SM, Bradshaw RD. Patient interest in video integration for after-hours telemedicine. J Am Board Fam Med 2020 Sep 28;33(5):765-773 [FREE Full text] [doi: 10.3122/jabfm.2020.05.190362] [Medline: 32989071]

Abbreviations

EMR: electronic medical record
HIS: hospital information system
ICD-10: International Classification of Diseases 10th Revision
NHI: National Health Insurance
OR: odds ratio

Edited by T Leung; submitted 14.06.22; peer-reviewed by L Tong; comments to author 06.07.22; revised version received 25.07.22; accepted 31.07.22; published 12.08.22.

Please cite as:
Tzeng YH, Yin WH, Lin KC, Wei J, Liou HR, Sung HJ, Lang HC
Factors Associated With the Utilization of Outpatient Virtual Clinics: Retrospective Observational Study Using Multilevel Analysis
J Med Internet Res 2022;24(8):e40288
URL: https://www.jmir.org/2022/8/e40288
doi:10.2196/40288
PMID:35917486

Original Paper

Barriers and Facilitators to the Implementation of Family-Centered Technology in Complex Care: Feasibility Study

Jody L Lin[1], MD, MS; Bernd Huber[2], PhD; Ofra Amir[3], PhD; Sebastian Gehrmann[2], PhD; Kimberly S Ramirez[4], BS; Kimberly M Ochoa[4], BA; Steven M Asch[5,6], MD, MPH; Krzysztof Z Gajos[2], PhD; Barbara J Grosz[2], PhD; Lee M Sanders[4], MD, MPH

[1]Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, United States
[2]John A Paulson School of Engineering and Applied Sciences, Harvard University, Allston, MA, United States
[3]Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa, Israel
[4]Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA, United States
[5]Center for Innovation to Implementation, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, United States
[6]Primary Care and Population Health, Stanford University School of Medicine, Palo Alto, CA, United States

Corresponding Author:
Jody L Lin, MD, MS
Department of Pediatrics
University of Utah School of Medicine
100 Mario Capecchi Drive
Salt Lake City, UT, 84114
United States
Phone: 1 801 662 3645
Email: jody.lin@hsc.utah.edu

Abstract

Background: Care coordination is challenging but crucial for children with medical complexity (CMC). Technology-based solutions are increasingly prevalent but little is known about how to successfully deploy them in the care of CMC.

Objective: The aim of this study was to assess the feasibility and acceptability of GoalKeeper (GK), an internet-based system for eliciting and monitoring family-centered goals for CMC, and to identify barriers and facilitators to implementation.

Methods: We used the Consolidated Framework for Implementation Research (CFIR) to explore the barriers and facilitators to the implementation of GK as part of a clinical trial of GK in ambulatory clinics at a children's hospital (NCT03620071). The study was conducted in 3 phases: preimplementation, implementation (trial), and postimplementation. For the trial, we recruited providers at participating clinics and English-speaking parents of CMC<12 years of age with home internet access. All participants used GK during an initial clinic visit and for 3 months after. We conducted preimplementation focus groups and postimplementation semistructured exit interviews using the CFIR interview guide. Participant exit surveys assessed GK feasibility and acceptability on a 5-point Likert scale. For each interview, 3 independent coders used content analysis and serial coding reviews based on the CFIR qualitative analytic plan and assigned quantitative ratings to each CFIR construct (–2 strong barrier to +2 strong facilitator).

Results: Preimplementation focus groups included 2 parents (1 male participant and 1 female participant) and 3 providers (1 in complex care, 1 in clinical informatics, and 1 in neurology). From focus groups, we developed 3 implementation strategies: education (parents: 5-minute demo; providers: 30-minute tutorial and 5-minute video on use in a clinic visit; both: instructional manual), tech support (in-person, virtual), and automated email reminders for parents. For implementation (April 1, 2019, to December 21, 2020), we enrolled 11 providers (7 female participants, 5 in complex care) and 35 parents (mean age 38.3, SD 7.8 years; n=28, 80% female; n=17, 49% Caucasian; n=16, 46% Hispanic; and n=30, 86% at least some college). One parent-provider pair did not use GK in the clinic visit, and few used GK after the visit. In 18 parent and 9 provider exit interviews, the key facilitators were shared goal setting, GK's internet accessibility and email reminders (parents), and GK's ability to set long-term goals and use at the end of visits (providers). A key barrier was GK's lack of integration into the electronic health record or patient portal. Most parents (13/19) and providers (6/9) would recommend GK to their peers.

Conclusions: Family-centered technologies like GK are feasible and acceptable for the care of CMC, but sustained use depends on integration into electronic health records.

Trial Registration: ClinicalTrials.gov NCT03620071; https://clinicaltrials.gov/ct2/show/NCT03620071

(J Med Internet Res 2022;24(8):e30902) doi:10.2196/30902

KEYWORDS

care coordination; implementation science; chronic illness; pediatric; family medicine; barrier; complex care; children; families; parents; care providers; chronic disease; coordination; implementation; improvement; technology; feasibility; acceptability; monitoring

Introduction

Defined by high service needs, high resource use, and functional disability, children with medical complexity (CMC) represent a disproportionately high share of pediatric care use but receive poor quality care when compared to their noncomplex counterparts [1-6]. CMC often require care coordination across multiple health care systems with a large care team that includes professional care providers, adult caregivers, and community agencies. Care coordination through multidisciplinary care teams centered around a patient-centered and family-centered medical home may improve outcomes for CMC but can be resource-intensive [7,8]. Moreover, many CMC access care across multiple health systems, receive care in resource-limited settings, and do not live adjacent to a tertiary pediatric center where many of these clinics are based, making scalability of these innovative teams difficult. Health care that is centered around shared goal-setting is a commonly proposed approach to coordinate care for CMC to improve clinical decision-making, family engagement, and health outcomes [9,10]. Although prior studies have deployed multidisciplinary teams to create shared care plans, few studies exist for effective and scalable tools to facilitate shared goal setting [8,11]. For children with noncomplex chronic conditions (eg, asthma, type 1 diabetes), mobile health technologies may provide efficacious ways to manage chronic medical conditions for children. These positive outcomes may translate to the care of CMC but to do so may also need to overcome additional challenges such as team hierarchies, loosely coupled teams, and asynchronous time scales among providers [12-14]. Many of these challenges affect the implementation of mobile health tools, which is essential for even the most efficacious tools.

In this study, we evaluated the implementation of an internet-based shared goal-setting tool (GoalKeeper) into the care of CMC. GoalKeeper is an internet-based tool developed by the study team to improve shared goal setting between parents and providers of CMC, and designed through interviews and iterative prototyping with this population. GoalKeeper consists of 2 modules: goal elicitation and tracking. The goal elicitation module is meant to be used jointly by parents and providers during a clinic visit to set family-centered goals and is shown in Figure S1 of Multimedia Appendix 1. During goal elicitation, parents and providers are prompted verbally and visually to share the screen and use verbal prompts on the screen to set goals. The first set of prompts asks for the parent's wishes/worries/concerns for their child's health care and the second set of prompts helps the parents and providers set specific, measurable, and timebound goals based on the wishes/worries/concerns. A third subsection of this module provides sample goals as inspiration. The tracking module includes customizable templates that providers could assign to parents to use to track their child's symptoms and daily progress relevant to the goals they set. After setting goals with their patient's parents, the providers could assign tracking templates relevant to these goals for parents to use in longitudinal symptom tracking during the trial. GoalKeeper was designed to be outside the electronic health record (EHR) to allow for rapid design modifications and to enable thorough assessment of effectiveness before potential future integration into the EHR. To facilitate the integration of the entered data into the EHR, the final screen of the goal-setting module also presents the data (ie, the wishes/worries/concerns and the goals) as a block text with a button to copy the text for easy pasting into the EHR. Parents and providers had distinct interfaces where they could view and input data. Providers could create new goals and tracking forms, while parents could view set goals and input data into the tracking templates. Additional details about GoalKeeper can be found in our forthcoming companion manuscripts (B Huber et al, unpublished data, 2022, and J Lin et al, unpublished data, 2021).

To illustrate how GoalKeeper works, we will consider a sample patient, Alex, a 7-year-old child with medical complexity who arrives at the clinic with his parents. In response to the first verbal prompt, his parents state they worry about Alex not attending school and not sleeping enough. At the next prompt, Alex's parents struggle to identify a specific goal; therefore, they turn to the sample goals for inspiration. After viewing sample goals focused on child development and discussion with their provider, they set a goal that, "Alex could be more awake during school based on adjustment of seizure medicines in the next two months." Alex's provider assigns a tracking template to measure school attendance and quality of sleep.

Nested within a larger effectiveness trial of the tool, this study aims to assess the barriers and facilitators of implementation of the GoalKeeper tool by using the Consolidated Framework for Implementation Research (CFIR) implementation framework. The CFIR is widely used to identify barriers and facilitators of implementation, including in health communication and adult and child chronic illness [15-18]. The CFIR contains 5 domains that interact to influence implementation effectiveness: inner setting, outer setting, characteristics of the individuals involved, intervention characteristics, and implementation process, with multiple constructs nested in each domain [15].

Methods

Study Design

This study is a prospective study of the implementation of a novel internet-based family-centered care plan called GoalKeeper nested under a prospective, stepped-wedge trial of GoalKeeper at a tertiary children's health system, Lucile Packard Children's Hospital Stanford. Details about the intervention design and results from the main trial are published in upcoming companion manuscripts (B Huber et al, unpublished data, 2022, and J Lin et al, unpublished data, 2021). Information about GoalKeeper is available through data-sharing requests directed to lsanders@stanford.edu. This study was conducted in 3 phases: preimplementation, implementation, and postimplementation. We selected a 3-phase approach, as the application of implementation science throughout intervention development is associated with increased success of implementation [18]. We selected the CFIR framework owing to its flexibility in assessing both the process of implementation and the barriers and facilitators to implementation, its use in formative evaluations at the preimplementation phase, and owing to the lack of effectiveness data of the novel tool used in the trial, as proven effectiveness is a key element of other implementation frameworks, whereas our trial evaluated the effectiveness of the tool with a secondary focus on implementation [16,19].

In the preimplementation phase, we conducted user testing in 3 stages of tool development: (1) early: parent and provider focus groups; (2) mid: individual role-play sessions and interviews using screen by screen feedback, hands-on, and think-aloud; and (3) late: pilot testing at Complex Primary Care Clinic (CPCC) with parents using GoalKeeper for a month after their clinic visit, instructing them to use GoalKeeper at least 3 times a week, followed by an exit interview. Focus groups and interviews used the CFIR interview guide questions to explore barriers and facilitators to the implementation of GoalKeeper. In the implementation phase, we implemented GoalKeeper at CPCC and pediatric neurology clinics by using implementation strategies informed by preimplementation focus groups. We recruited parent participants from the clinic of each enrolled provider for 3 weeks. Each parent was asked to use GoalKeeper with their provider at their enrollment clinic visit and for 3 months after the initial visit. At the end of the study, participants completed a postimplementation semistructured interview and survey.

Ethical Considerations

This study was approved by the Stanford University's Single Institutional Review Board (Protocol # 32161) and is registered on ClinicalTrials.gov (NCT03620071).

Study Population

For preimplementation focus groups and interviews, we recruited parents of CMC seen at Stanford by using a convenience sample of parents at CPCC. We selected providers from clinics planned for trial recruitment, CPCC and pediatric neurology, the hospital medicine team with specialization in caring for CMC, and a provider with expertise in clinical informatics for feedback on workflow integration. For the implementation phase, we recruited medical providers (physicians, nurse practitioners, physician assistants) from 2 clinical services that see the highest proportion of CMC: CPCC and pediatric neurology (2 clinics that historically care for many CMC). All providers at these clinics were eligible for recruitment. Based on prior work, goal setting was not routine practice in these clinics [14]. From the patients seen by enrolled providers, we recruited a convenience sample of primary caregivers (eg, parents). Parents were eligible if they were aged ≥18 years, English-speaking, and with a child with medical complexity <12 years presenting for a routine (not sick) visit. Parents who did not have home access to the internet were excluded. We excluded older children, who may have the capacity to participate in decision-making, since the tool was not designed for interaction with children [20]. We defined medical complexity as meeting all of the following criteria in the past 12 months: ambulatory visits with at least 2 subspecialty providers and functional impairment due to a chronic condition [21]. Recruitment occurred solely in-person, but due to the COVID-19 pandemic, recruitment was paused between March 13 and July 20, 2020. We did not pursue remote recruitment owing to concerns of intervention fidelity because GoalKeeper was designed to be used jointly by the provider and parents during a clinical encounter.

Measures and Outcomes

We focused on 3 of the 5 CFIR domains: intervention characteristics (GoalKeeper), inner setting (CPCC and pediatric neurology clinics), and characteristics of individuals involved (parents and providers of CMC). We used an adapted version of the CFIR interview guide focus groups and interviews to inform our approach for implementation of GoalKeeper [15,22]. The CFIR interview guide contains open-ended interview questions organized by the CFIR domain and construct. To limit interview length, the entire study team reviewed the interview guide together and selected questions based on constructs that we felt were the most relevant to our study. All focus groups and interviews were audio recorded and transcribed for subsequent review.

In the preimplementation phase, we conducted focus groups in person with study team members as observers. Focus groups included a project overview, a demonstration of the current intervention prototype, open feedback about the intervention, and semistructured questions. In the implementation phase, we collected user data from all participants, including number of goals set, types of goals set, and number of data entries after the encounter. We assessed feasibility based on the proportion of the intervention group who used GoalKeeper during the clinical encounter. In the postimplementation phase, we conducted individual exit interviews with parent and provider participants. We assessed acceptability by using a 5-point Likert scale to determine whether GoalKeeper was useful and fit into the clinic workflow.

Analysis

For preimplementation focus groups, all study team members collectively synthesized the key facilitators, barriers, and design considerations immediately after each focus group and interview and at weekly team meetings after reviewing the transcripts. Postimplementation exit interviews were analyzed using the CFIR qualitative analytic approach that starts with deductive coding to apply the CFIR as a coding framework and then applies inductive methods by using open, axial, and selective coding to create new codes and ultimately themes that arose from the data [23]. Three independent coders (authors JLL, KSR, and KMO) analyzed each transcript. After coding each transcript, the coders independently rated each represented CFIR construct on a scale of –2 (strong barrier to implementation) to

+2 (strong facilitator to implementation) based on the CFIR analytic approach and calculated the overall ratings for all transcripts based on the median rating. At serial coding reviews, we reviewed codes and CFIR ratings to reach consensus and generated themes based off the transcripts. We calculated means and SDs for the feasibility and acceptability measures and logged data. We compared participant characteristics between parents who were interviewed and those who were not by using chi-square tests for categorical variables and 2-sided Student *t* test for continuous variables with a significance level of $P \leq .10$.

Results

Preimplementation Results

For the preimplementation phase, in the early design stage, we conducted 1 provider focus group and 1 parent focus group. The provider focus group consisted of 3 physicians with specialties in general pediatrics, pediatric neurology, and clinical informatics. The parent focus group consisted of 2 parents of CMC. In the middesign stage, we conducted role-play sessions with 2 providers (1 complex care, 1 hospitalist) and interviewed 3 parents and 1 neurologist. In the late design stage, 2 parents pilot-tested GoalKeeper. From these focus groups and interviews, we devised 3 implementation strategies based on those previously tested by other studies: educational materials, individual technical assistance, and automated email reminders to parent participants [22].

Implementation Results

During the implementation phase, participants received educational materials, including paper and electronic copies of an instructional manual on GoalKeeper, with parent and provider versions. Parents also received a 5-minute in-person overview of GoalKeeper, while providers received a 30-minute overview, including a 5-minute educational video on using GoalKeeper during a clinic visit. A member of the study team was available for in-person individual technical assistance in clinic and reachable by email or phone outside of clinic but did not attend the clinic visit with intervention participants. Parent participants received automated weekly email reminders to log into GoalKeeper and track their progress on the goals that were set. Email frequency was set at 1 week by default, but the reminder frequency could be modified by parents, including turning reminders off.

We enrolled 11 providers and 35 parents (15 from complex care and 20 from neurology) in the intervention arm of the trial. The providers were mostly physicians (9/11, 82%), mostly female (7/11, 64%), and pediatric neurologists (6/11, 54%). Parents had a mean age of 38.3 (SD 7.8) years, were mostly female (28/35, 80%), primarily identified as White (17/35, 49%), almost half identified as of Hispanic ethnicity (16/35, 46%), with a mean household size of 4.1 (SD 1.0), were married (23/35, 66%), and had at least some college education (86%). Children of participants had a mean age of 5.9 (SD 3.8) years, 18 (51%) identified as White, 16 (46%) as Hispanic, 25 (71%) were followed by a neurologist, 20 (57%) had technology dependence, and 20 (57%) had neurodevelopmental delay. A total of 16 parents were lost to follow-up: 7 at the 1-month follow-up and 9 at the 3-month follow-up; 9 providers and 19 parents completed exit surveys, and 9 providers and 18 parents (7 CPCC, 11 neurology) completed exit interviews. Parent, provider, and child characteristics can be found in Tables 1-3, respectively.

During the initial clinic visit at the start of the study, 34 parent-provider dyads completed goal setting with GoalKeeper. For the 1 parent-provider dyad who did not complete goal setting, the provider decided in the visit that the parent was not a good fit to participate in the trial as the patient was revealed during the clinic visit to have an acute medical issue that needed to be the focus of the entire visit.

Table 1. Parent participant characteristics.

Characteristics	Total (N=36)	Interviewed (n=19)	Not interviewed (n=17)
Age (years), mean (SD)	38.7 (7.8)	40.4 (8.4)	36.9 (7.0)
Sex (female), n (%)	28 (78)	15 (79)	13 (77)
Race/ethnicity, n (%)			
Caucasian	17 (47)	11 (58)	6 (35)
African American	2 (6)	0 (0)	2 (12)
American Indian	1 (3)	1 (5)	0 (0)
Asian	3 (8)	2 (11)	1 (6)
Hispanic	14 (39)	4 (21)	10 (59)
Other	1 (2.8)	1 (5)	0 (0)
Insurance, n (%)			
Medicaid	13 (36)	4 (21)	9 (53)
Private	18 (50)	11 (58)	7 (41)
State Children’s Health Insurance Program	1 (3)	0 (0)	1 (6)
Medicare	5 (14)	3 (16)	2 (12)
Other	7 (19)	6 (32)	1 (6)
Don’t know	1 (3)	0 (0)	1 (6)
Decline to answer	1 (3)	0 (0)	1 (6)
Household size, mean (SD)	4.12 (1.0)	4.0 (1.1)	4.3 (0.9)
Marital status, n (%)			
Single	9 (25)	3 (16)	6 (35)
Married	23 (63)	13 (68)	10 (59)
Living with partner	1 (3)	1 (5)	0 (0)
Single, divorced	1 (3)	1 (5)	0 (0)
Education level, n (%)			
>9th grade	1 (3)	0 (0)	1 (6)
Some high school	1 (3)	1 (5)	0 (0)
High school diploma	4 (11)	1 (5)	3 (18)
Some college	10 (28)	6 (32)	4 (24)
College degree	10 (28)	6 (32)	4 (24)
Advanced degree	7 (19)	4 (21)	3 (18)
Home internet access (select all), n (%)			
Laptop	24 (67)	12 (63)	12 (71)
Tablet/e-reader	15 (42)	7 (37)	8 (47)
Smartphone	26 (72)	14 (74)	12 (71)
Mobile phone	7 (19)	1 (5)	6 (35)

Table 2. Provider characteristics (n=11).

Characteristics	Value, n (%)
Sex (female)	7 (64)
Specialty	
General pediatrics	5 (46)
Pediatric neurology	6 (54)
Degree	
Doctor of medicine	9 (82)
Doctor of osteopathic medicine	1 (9)
Nurse practitioner	1 (9)

Table 3. Child characteristics.

Characteristics	Total (N=36)	Interviewed (n=19)	Not interviewed (n=17)
Age (years), mean (SD)	5.8 (3.7)	5.3 (3.9)	6.3 (3.7)
Sex (female), n (%)	16 (44)	8 (42)	8 (47)
Race/ethnicity, n (%)			
Caucasian	18 (50)	12 (63)	6 (35)
African American	1 (3)	0 (0)	1 (6)
Asian	5 (14)	3 (16)	2 (12)
American Indian	1 (3)	1 (5)	0 (0)
Hispanic	16 (44)	7 (37)	9 (53)
Other	2 (6)	1 (5)	1 (6)
Children with special health care needs screener, n (%)			
Needs or uses prescription medicines	30 (83)	14 (74)	16 (94)
Needs or uses more medical care than usual	31 (86)	16 (84)	15 (88)
Functional limitations more than usual	29 (81)	15 (79)	14 (82)
Needs or uses special therapies	31 (86)	17 (90)	14 (82)
Needs or uses treatment for emotional/developmental/behavioral issues	20 (56)	12 (63)	8 (47)
Subspecialists, n (%)			
Cardiology	12 (33)	6 (32)	6 (35)
Neurology	27 (75)	14 (74)	13 (77)
Pulmonology	18 (50)	10 (53)	8 (47)
Development	20 (56)	12 (63)	8 (47)
Gastroenterology	22 (61)	10 (53)	12 (71)
Occupational therapy	25 (69)	14 (74)	11 (65)
Speech therapy	15 (42)	9 (48)	6 (35)
Physical therapy	25 (69)	14 (74)	11 (65)
Other	13 (36)	8 (42)	5 (29)
Technology dependence, n (%)			
Ventriculoperitoneal shunt	4 (11)	2 (11)	2 (12)
Gastrostomy tube	14 (39)	7 (37)	7 (41)
Tracheostomy	2 (6)	0 (0)	2 (12)
Other: vagal nerve stimulator[a]	3 (8)	3 (16)	0 (0)
None	15 (42)	7 (37)	8 (47)
Neurodevelopmental delay, n (%)			
Intellectual disability	18 (50)	9 (47)	9 (53)
Cerebral palsy	8 (22)	4 (21)	4 (24)
Visual impairment	7 (19)	4 (21)	3 (18)
Hearing deficit	2 (6)	0 (0)	2 (12)
None	15 (42)	7 (37)	8 (47)

[a] Participants who were interviewed were more likely to have a child with a vagal nerve stimulator (*P*=.08).

Postimplementation Results: Feasibility and Acceptability

Of the 19 parents who completed the exit survey, 13 (68%) parents responded that they would recommend GoalKeeper to other parents. Only 1 respondent would not, while others were undecided; 6 (67%) providers would recommend GoalKeeper to other providers. Providers commented that they would have wanted GoalKeeper integrated into the EHR. Parents would have wanted more options for reminders for use of the tool, 12 (86%) parents felt the tool was easy to use, 6 (75%) providers found GoalKeeper useful, and 4 (50%) providers felt GoalKeeper fit into their workflow. More details on the survey results are summarized in Table S2 of Multimedia Appendix 1.

From data use logs of participant use of GoalKeeper, parents and providers set a median of 2.5 (range 0-4) goals per initial clinic visit, and providers assigned a median of 3 (range 0-5) tracking templates to each parent participant. Each provider saw a median of 2.5 (range 0-10) parent participants. After the initial clinic visit, parents input information into GoalKeeper a median of 0 (range 0-19) times and providers viewed tracked data a median of 0.5 (range 0-2) times throughout the study period. The patterns of tool use based on the number of tracking templates created are summarized in Figure S3 of Multimedia Appendix 1. We conducted a post hoc analysis of correlation between the number of tracking templates used and the number of times parents entered data into the tracking templates and found a Pearson correlation coefficient of 0.41 (*P*=.01); 19 parents (8 CPCC and 11 neurology) did not enter any data into the tracking templates, while 13 parents (4 CPCC and 9 neurology) entered data 1-10 times, and 3 parents all from CPCC entered data over 10 times. Of the parents who were interviewed, 6 (1 CPCC and 5 neurology) did not enter any data, while 8 (2 CPCC and 6 neurology) entered data 1-10 times, and 3 (all CPCC) entered data over 10 times.

Postimplementation Results: Barriers and Facilitators

Participant exit interviews covered CFIR domains of intervention characteristics, inner setting, and characteristics of individuals involved and their related constructs. From these interviews, we categorized each barrier and facilitator under a CFIR domain and construct based on topic and rated each CFIR construct that was represented in the interviews. Participant knowledge and beliefs about goal setting were facilitators to implementation with a rating of +1 whereas adaptability and compatibility were barriers, with each receiving a rating of –1. The complete ratings are given in Table 4. Limited quotes are provided in the text with additional quotes found in Table 5. There were no statistically significant differences in parent or child characteristics between those who were interviewed and those who were not except for participants who were interviewed were more likely to have a child with a vagal nerve stimulator (*P*=.08).

Table 4. Consolidated Framework for Implementation Research ratings by construct and participant type.

Domain mean ratings (scale, –2: strong barrier; +2: strong facilitator)	Parents (n=19)	Providers (n=9)
Inner setting		
Tension for change: the degree to which stakeholders perceive the current situation as intolerable or needing change	1	0
Compatibility: the degree of tangible fit between meaning and values attached to the intervention by involved individuals, how those align with individuals' own norms, values, and perceived risks and needs, and how the intervention fits with existing workflows and systems	–1	–1
Characteristics of individuals		
Knowledge and beliefs: individual's attitudes toward and value placed on intervention as well as familiarity with facts, truths, and principles related to the intervention	1	1
Individual state of change: characterization of the phase an individual is in, as he or she progresses toward skilled, enthusiastic, and sustained use of the intervention	1	0
Intervention characteristics		
Relative advantage: stakeholder's perception of the advantage of implementing the intervention versus an alternative solution	0	0
Complexity: perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, and intricacy and number of steps required to implement	1	0
Adaptability: the degree to which an intervention can be adapted, tailored, refined, or reinvented to meet local needs	–1	–1

Table 5. Sample quotes and main themes for implementation.

	Parents	Providers
Facilitators		
Parents and providers		
Goal-centered care is important to the care of children with medical complexity (tension for change)	"...I thought, 'This is interesting. I think this is going to be good.' Setting up the goals and giving me something on my side to kind of think about and work toward." (Parent 1227, age 57 years, some college)	"...I think it's important because sometimes the physician's goals are not the same as the family's goals. And so shared decision making doesn't always happen." (Provider 39, neurology)
Reminders are helpful, and more flexibility around reminder frequency would have improved use; providers wanted limited reminders (individual state of change)	"...Mostly when I get emails from you and stuff and it also reminds me too that I need to check her goals and everything I have on there." (Parent 1494, age 29 years, high school diploma)	"...And kind of the structured way of kind of connecting, whether it's weekly, or daily, or monthly, or whatever it is. To keep on something that is important for the family. I think it is important. So that might be useful." (Provider 33, complex primary care clinic)
The tool helped facilitate goal-centered care (knowledge and beliefs)	"...There are so many things I need to take care of. But then, when it comes to zeroing down to the main thing that matters, in that way, Goalkeeper was very helpful for me." (Parent 1510, age unknown, college degree)	"...And so in theory, this is a really great and I totally think that this is the way as a physician you should be thinking about it and trying to see what your family's goals are so that you can understand what it is that they want you to help with." (Provider 69, neurology)
Parents only		
The tool could be accessed from anywhere, but 2 parents still used other tools (relative advantage)	"...I always used to bring it in my own binder to the doctor and it's just a lot to carry rather than when I just have my cell phone all the time." (Parent 1476, age 38 years, some college)	N/A[a]
Providers only		
The tool should be used at end of visit to fit into clinic workflow (compatibility)	N/A	"...I mean sometimes if it worked out that I had to do it in the beginning of the visit because you were in the room to help me out and things like that, I found that less desirable than if I was able to time it toward the end of the visit." (Provider 50, complex primary care clinic)
The tool opens doors in patient care that providers otherwise would not explore (knowledge and beliefs)	N/A	"...So what the patient or their family values the most. So what they find most important to them, often times that comes into quality of life decisions, for example. So in that discussion about eating, they might value being able to eat some or having their child being able to eat some food more than a 50% reduction in their seizure frequency, or even being seizure free. Whereas by default, normally I would-- we're very focused on seeing if we can get to zero seizures. And knowing that that's the relative value that the family's putting on things is very, very important. Sometimes that would mean that maybe there is a reason why that is, or just as also just means important that there'll be something to work on because there can be trade-offs between those two things. For example, higher doses of antiseizure medications might cause more drooling or less ability to be awake to eat. Does that make sense? So both things are valuable, but which one is more valuable to that family? And so that you can-- it helps you prioritize between those two goals or two potential actions." (Provider 63, neurology)
Barriers		

	Parents	Providers
Parents and providers		
The tool would have been easier to access if it was embedded in the electronic health record/MyChart (compatibility)	"...I think the disadvantages is having another thing to log into. As is evidenced by the fact that I can't seem to get in right now I can barely keep all of my logins straight at the moment but MyChart is sticky. I have to use MyChart to keep track of-- my son has a complex medical condition, right? So between the insurance billings and keeping the appointments straight and dealing with all the different providers at Stanford, I'm tethered to MyChart." (Parent 1150, age 45 years, advanced degree)	"...The actual use of the application is a little awkward because it requires a separate login to the website. In terms of feedback, it would be nice if that could be embedded in the chart." (Provider 63, neurology)
Lack of feedback and closed loop communication hindered frequent use of the tool due to feelings that the other party was not using the tool (relative advantage)	"...So right now, being able to communicate with the doctor and having them understand that there is a communication between the two of you rather than on GoalKeeper it's only a one-way email and then they'd have to call you because there was no option for them to email." (Parent 1476, age 38 years, some college)	"...So I wouldn't want to use it if that kind of-- for some reason we're just not getting reliable feedback from families. If there's not a loop there." (Provider 50, complex primary care clinic)
Parents only		
Only providers could create and change goals, hindering parent engagement with the tool when goals became irrelevant (adaptability)	"...in the beginning when we first started, it was nice. But at the same time, as time goes by, it becomes a little bit more repetitive, especially when you're not able to change things unless you go to the provider's office or speak to the provider." (Parent 1490, age 33 years, some college).	N/A
Limited or unreliable internet access prevented constant use of the tool (adaptability)	"...I have limited internet data on my phone. The first time when I was able to access the first surveys, they were easy. The alert came. I was able to connect, and I was able to answer the questions with no problem. Again, in my case is not having access to technology when make this difficult for me." (Parent 1668, age 53 years, some college)	N/A
Providers only		
When providers perceived they were already practicing goal-centered care, they felt the tool was redundant even if they adopted aspects of the tool into their practice (relative advantage)	N/A	"...I think it is something that I do regularly as part of our visit. So I don't know that it's going to change my practice or very much in terms of goal-setting." (Provider 42, neurology)

[a] N/A: not applicable.

Intervention Characteristics

When compared to existing tools, parents and providers felt that GoalKeeper was a better way to start goals-of-care conversations. One provider shared:

> *I think, personally, I always phrased it in my own clinical practice, just as, "Tell me something that is most important to you or something that is a priority for us to work on for your child right now." Again, just trying to not actually always use the word "goal" itself. Or if you did ask what the parents' goals were, then providing just a little bit more of an explanation. So I did think that, again, the wording of "wishes and worries" I liked. It was something new for me that was not a wording choice I had used before* [Provider 57, CPCC]

Parents and providers also felt that GoalKeeper facilitated teamwork during and after clinic visits. Parents also liked that GoalKeeper could be accessed from anywhere they had internet access.

All providers and most parents felt a barrier to use of the intervention was difficulty accessing the intervention because the URL link was not user-friendly, and they had to access it through a previously received email from the study team. Home internet connectivity issues also prevented some parents from using GoalKeeper in the follow-up period. Two parents perceived that the goal-setting portal being provider-driven hindered their use of the intervention because their provider was unengaged about updating the goals after the initial visit.

Inner Setting

Almost all parents and providers felt there was a need for more shared goal setting in their current clinical care, which they felt promoted improved parent-provider engagement particularly for patients with more active complex medical issues. Providers overwhelmingly felt that parents of children who would need the provider's long-term care to be the ideal audiences for the intervention with some providers saying that the intervention allowed them to focus on parents' long-term goals. Parents also felt the intervention could fit parents of younger children who had developmental goals such as those with prematurity or other children with special health care needs. GoalKeeper was felt by both providers and parents to be incompatible with existing workflows because it was not integrated into the EHR and patient portal. Providers felt using the intervention during the middle or end of their visit fit better into their workflow, but that clinic time constraints make it hard to squeeze the intervention into the visit. Providers also wanted a Spanish-language form of the platform to target a population they felt is most in need of help setting goals for their child. When using the tracking module, parents overall desired provider feedback and communication about the information parents entered into the intervention with 1 parenting summarizing, “if there is no two-way communication or if there is no template or anything set up, then the value goes down” (Parent 1898, age 47 years, advanced degree). Parents admitted that at times, competing priorities for their child's health superseded the use of GoalKeeper such as when a child became hospitalized for an issue that was not captured in the goals they set in GoalKeeper.

Characteristics of the Individuals Involved

All participants felt that conversations centered around goal-setting were beneficial because they switched medical discussions to long-term and in the context of what is important to the family. Parents felt that the intervention shifted their mindsets by focusing them on the main concerns for their child. Parents felt that using the intervention made them more confident to articulate their concerns to their providers during the clinic visit and helped them identify to-do lists to achieve their goals for their child, with 1 parent remarking:

> *So I think that that's kind of the gift of motivation because it's like, “Oh, yes. We have some ability.” I think sometimes you look at a child who has a lot of needs and as a parent, you can get discouraged and then think, “Okay. We'll just do whatever the doctor say.”...so I guess that's a part of knowing that you can do some things to improve your child's life and to reduce their sort of future medical interventions is so helpful as a lot of parents can feel helpless with these kind of situations* [Parent 1368, age 37 years, college degree]

Parents also remarked that GoalKeeper helped them prioritize their child's short- and long-term goals. Parents felt supported and hopeful about the well-being of their child. However, most parents and providers did not adopt the intervention in its entirety with most participants only using the intervention during the initial clinical encounter. One provider shared that “like every habit, I think it might take many repetitions to start to want to use it regularly” (Provider 33, CPCC). Furthermore, providers were more motivated to use the intervention if they felt it opened new doors to insights about the patient that were not captured in their typical clinical practice. One provider remarked:

> *Well, I think there are some issues that came up and some questions I just hadn't asked before and a lot of it focused around happiness and joy and things like that that led to conversations with families I hadn't had before. So I think sometimes it'll open doors for me and the preset questions might even be more so. It'd opened doors for me that I might not have opened without Goalkeeper or without asking* [Provider 50, CPCC]

Two providers felt GoalKeeper was redundant to their practice even though they integrated part of GoalKeeper into their future practice.

Implementation Strategies

The 3 implementation strategies used during the study, that is, educational materials, individual technical assistance, and automated email reminders to parent participants, had varying success. The video tutorial provided to providers was found to be useful during the initial training, but few viewed the video after the training. No participant used the paper manual. Few participants used the individual technical assistant either virtually or in person, but for those who encountered technical issues during the visit, they found it useful with 1 provider commenting:

> *I always forgot the step where I was supposed to copy the goal and then go back to the page to add the template. So I just felt like each time you enrolled a patient, I was always turning back toward you to ask if I was clicking through it correctly to actually enter goals* [Provider 57, CPCC]

Parents overwhelmingly felt that the automated email reminders helped sustain their use of the intervention by reminding them of the goals they set for their child and to log into the system to use the tracking module with 1 parent remarking:

> *The email reminders were really great at making it easy to just log on and track it. And then it really was not time-consuming* [Parent 1352, age 28, some college]

Discussion

In this pilot clinical trial of a novel internet-based goal-setting tool, we successfully implemented the tool for use during ambulatory clinic visits to facilitate goal elicitation. However, we did not succeed in the sustained use of the tool after the initial clinic visit. A key facilitator to implementation was participant value for the intervention's stated aim: family-centered goal setting. A key barrier to implementation was inadequate integration with the EHR and patient portals. Provider use of the tool was also influenced by whether they felt the tool opened new doors to insights about the patient and their family that they were unable to get in their usual practice such as the relative priority of seizure control compared to appetite or wakefulness. Although automated email reminders were important to sustain tool use, they were insufficient. Parents and providers wanted a more dynamic way to communicate about goals through the intervention with most participants sharing that a feedback loop from the other party would have encouraged continued tool use. One aspect of the intervention design that is critically important is that the English-language and internet-based platform may exacerbate difference in care quality for minority and less-resourced populations.

Although our findings that automated email reminders and training are helpful implementation strategies are consistent with the implementation of digital health tools in adult chronic illness, we provide new insights for pediatric chronic illness care [24]. For example, we found that parents and providers both desired a feedback loop to sustain, which adds the caregiver perspective to prior studies of patient-generated health information where sustained use hinged on a tangible immediate benefit [25]. These findings also contribute practical considerations to the implementation of multidisciplinary care and family-centered care plans, particularly with regard to scalability and replicability across multiple institutions [11].

Our findings also contribute an implementation science perspective to practice transformation in digital health. We found that a lack of workflow integration hindered the compatibility of our intervention with the existing workflows of the providers. Future research should explore the use of clinical staff as mediators between providers and patients or caregivers for patient-reported information to facilitate intervention adoption and alleviate potential burdens that providers face with the introduction of a new intervention [26]. Furthermore, integration of our intervention into the workflow of clinical staff could address the needed feedback loop that our study participants desired but felt was lacking in the current form of the intervention. Having a tool that is well-integrated into clinic workflows would allow providers to revise and update goals as they evolve, the absence of which prevented parents from using the tool for a longer term in our study. A lack of workflow integration hindered other interventions across broad populations, including in advanced care planning for adults, pain management, and surgical safety [27-29]. Our intervention also was outside the electronic health portal, resulting in similar issues of workflow integration that have been observed in other digital interventions for populations with chronic conditions [30].

Although we purposefully built the intervention outside of the EHR to allow for rapid design improvements based on study results, future research should also explore integration of goal-setting tools within the EHR and patient portals to improve adoption. Recent efforts by payers to endorse the use of open application programming interfaces may help accelerate the integration of patient-generated health data into EHRs [31,32]. Such integration efforts should consider barriers found in our study, including how to represent patient-generated data and how to integrate goal-setting actions into the provider workflow [33,34]. As patients and parents gain access to provider notes, this research may guide efforts to improve patient and parent understanding, improve communication, and increase their empowerment [35,36].

Finally, language-related disparities must be addressed in any digital health intervention. Our participants emphasized the potential for an English-only internet-based intervention to exacerbate disparities in high-quality care coordination. Populations with low health literacy report lower use of digital health tools, which may widen the gaps in care quality they already experience owing to low health literacy [37,38]. Although 90% of the adults in the United States use the internet, fewer than 2 in 3 adults who identify as Hispanic or African American have broadband access at home with similar patterns based on lower income and education level [39]. Moreover, preference for languages other than English is associated with a decreased use of digital health tools for patient-provider communication [40]. Therefore, future work in implementation science of digital health tools should aim to understand the modifiable factors that influence the adoption by patients and families with preferences for languages other than English.

Our study should be interpreted in the context of a few limitations. The preimplementation phase used feedback from a small sample of parents and providers, which may have biased the selection of implementation strategies. Given the small number of providers practicing at each clinic, we limited feedback to a few providers to preserve an adequate provider sample for the trial. This study was conducted in a single academic medical center, which may not be reflective of the practice at other institutions caring for CMC. Not all parent participants participated in the exit interview, which may have introduced selection bias, but aside from the presence of the vagal nerve stimulator, there were no statistically significant differences in these 2 populations. Overall, our study period was quite short. Thus, the long-term use of our intervention, particularly on a repeat clinic visit, was not observed. The success of long-term implementation should be assessed with future studies.

Conclusion

Family-centered technologies like our intervention can be successfully implemented into ambulatory primary and subspecialty care. However, long-term adoption rests on integration into the EHR and patient portal as well as adaptation of tools for users who prefer languages other than English.

Acknowledgments

The authors have no financial relationships relevant to this article to disclose. This work was supported by the National Institutes of Health, National Cancer Institute (NIH R01CA204585). The funding sources had no involvement in the research activities related to this work. The authors thank all the parents and providers who participated in this clinical trial and those who helped inform the development of the tool.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental files.
[DOCX File , 427 KB - jmir_v24i8e30902_app1.docx]

References

1. Cohen E, Kuo DZ, Agrawal R, Berry JG, Bhagat SKM, Simon TD, et al. Children with medical complexity: an emerging population for clinical and research initiatives. Pediatrics 2011 Mar;127(3):529-538 [FREE Full text] [doi: 10.1542/peds.2010-0910] [Medline: 21339266]
2. Neff JM, Sharp VL, Muldoon J, Graham J, Myers K. Profile of medical charges for children by health status group and severity level in a Washington State Health Plan. Health Serv Res 2004 Feb;39(1):73-89 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00216.x] [Medline: 14965078]
3. Simon TD, Berry J, Feudtner C, Stone BL, Sheng X, Bratton SL, et al. Children with complex chronic conditions in inpatient hospital settings in the United States. Pediatrics 2010 Oct;126(4):647-655 [FREE Full text] [doi: 10.1542/peds.2009-3266] [Medline: 20855394]
4. Cohen E, Berry JG, Camacho X, Anderson G, Wodchis W, Guttmann A. Patterns and costs of health care use of children with medical complexity. Pediatrics 2012 Dec;130(6):e1463-e1470 [FREE Full text] [doi: 10.1542/peds.2012-0175] [Medline: 23184117]
5. Strickland BB, Jones JR, Newacheck PW, Bethell CD, Blumberg SJ, Kogan MD. Assessing systems quality in a changing health care environment: the 2009-10 national survey of children with special health care needs. Matern Child Health J 2015 Feb;19(2):353-361 [FREE Full text] [doi: 10.1007/s10995-014-1517-9] [Medline: 24912943]
6. Kuo DZ, Goudie A, Cohen E, Houtrow A, Agrawal R, Carle AC, et al. Inequities in health care needs for children with medical complexity. Health Aff (Millwood) 2014 Dec;33(12):2190-2198 [FREE Full text] [doi: 10.1377/hlthaff.2014.0273] [Medline: 25489038]
7. Mosquera RA, Avritscher EBC, Samuels CL, Harris TS, Pedroza C, Evans P, et al. Effect of an enhanced medical home on serious illness and cost of care among high-risk children with chronic illness: a randomized clinical trial. JAMA 2014;312(24):2640-2648. [doi: 10.1001/jama.2014.16419] [Medline: 25536255]
8. Kuo DZ, Berry JG, Glader L, Morin MJ, Johaningsmeir S, Gordon J. Health services and health care needs fulfilled by structured clinical programs for children with medical complexity. J Pediatr 2016 Feb;169:291-6.e1 [FREE Full text] [doi: 10.1016/j.jpeds.2015.10.010] [Medline: 26526361]
9. Council on Children with Disabilities and Medical Home Implementation Project Advisory Committee. Patient- and family-centered care coordination: a framework for integrating care for children and youth across multiple systems. Pediatrics 2014 May;133(5):e1451-e1460. [doi: 10.1542/peds.2014-0318] [Medline: 24777209]
10. McAllister JW. Achieving a shared plan of care with children and youth with special health care needs. Lucille Packard Foundation for Children's Health. 2014 May. URL: https://www.lpfch.org/publication/achieving-shared-plan-care-children-and-youth-special-health-care-needs [accessed 2022-08-08]
11. McAllister JW, Keehn RM, Rodgers R, Lock TM. Care coordination using a shared plan of care approach: from model to practice. J Pediatr Nurs 2018;43:88-96. [doi: 10.1016/j.pedn.2018.09.009] [Medline: 30473161]
12. Prahalad P, Tanenbaum M, Hood K, Maahs DM. Diabetes technology: improving care, improving patient-reported outcomes and preventing complications in young people with Type 1 diabetes. Diabet Med 2018 Apr;35(4):419-429. [doi: 10.1111/dme.13588] [Medline: 29356074]
13. Nkoy FL, Fassl BA, Wilkins VL, Johnson J, Unsicker EH, Koopmeiners KJ, et al. Ambulatory management of childhood asthma using a novel self-management application. Pediatrics 2019 Jun;143(6):e20181711. [doi: 10.1542/peds.2018-1711] [Medline: 31097465]
14. Amir O, Grosz BJ, Gajos KF, Swenson SM, Sanders LM. From care plans to care coordination: opportunities for computer support of teamwork in complex healthcare. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15; April 18-23, 2015; Seoul, Republic of Korea p. 1419-1428. [doi: 10.1145/2702123.2702320]
15. 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]
16. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]

17. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]
18. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the Consolidated Framework for Implementation Research. Implement Sci 2016 May 17;11:72 [FREE Full text] [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]
19. Landes SJ, McBain SA, Curran GM. An introduction to effectiveness-implementation hybrid designs. Psychiatry Res 2019 Oct;280:112513 [FREE Full text] [doi: 10.1016/j.psychres.2019.112513] [Medline: 31434011]
20. Garth B, Murphy GC, Reddihough DS. Perceptions of participation: child patients with a disability in the doctor-parent-child partnership. Patient Educ Couns 2009 Jan;74(1):45-52. [doi: 10.1016/j.pec.2008.07.031] [Medline: 18760558]
21. Lin JL, Clark CL, Halpern-Felsher B, Bennett PN, Assis-Hassid S, Amir O, et al. Parent perspectives in shared decision-making for children with medical complexity. Acad Pediatr 2020;20(8):1101-1108 [FREE Full text] [doi: 10.1016/j.acap.2020.06.008] [Medline: 32540424]
22. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
23. Damschroder LJ, Lowery JC. Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). Implement Sci 2013 May 10;8:51 [FREE Full text] [doi: 10.1186/1748-5908-8-51] [Medline: 23663819]
24. Varsi C, Solberg Nes L, Kristjansdottir OB, Kelders SM, Stenberg U, Zangi HA, et al. Implementation strategies to enhance the implementation of eHealth programs for patients with chronic illnesses: realist systematic review. J Med Internet Res 2019 Sep 27;21(9):e14255 [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
25. Reading MJ, Merrill JA. Converging and diverging needs between patients and providers who are collecting and using patient-generated health data: an integrative review. J Am Med Inform Assoc 2018 Jun 01;25(6):759-771 [FREE Full text] [doi: 10.1093/jamia/ocy006] [Medline: 29471330]
26. Buis L. Implementation: the next giant hurdle to clinical transformation with digital health. J Med Internet Res 2019 Nov 20;21(11):e16259 [FREE Full text] [doi: 10.2196/16259] [Medline: 31746763]
27. Haverhals LM, Gilman C, Manheim C, Bauers C, Kononowech J, Levy C. Implementation of VA's Life-Sustaining Treatment Decisions Initiative: facilitators and barriers to early implementation across seven VA medical centers. J Pain Symptom Manage 2021 Jul;62(1):125-133.e2. [doi: 10.1016/j.jpainsymman.2020.10.034] [Medline: 33157178]
28. Sarkar U, Lee JE, Nguyen KH, Lisker S, Lyles CR. Barriers and facilitators to the implementation of virtual reality as a pain management modality in academic, community, and safety-net settings: qualitative analysis. J Med Internet Res 2021 Sep 22;23(9):e26623 [FREE Full text] [doi: 10.2196/26623] [Medline: 34550074]
29. Bergs J, Lambrechts F, Simons P, Vlayen A, Marneffe W, Hellings J, et al. Barriers and facilitators related to the implementation of surgical safety checklists: a systematic review of the qualitative evidence. BMJ Qual Saf 2015 Dec;24(12):776-786. [doi: 10.1136/bmjqs-2015-004021] [Medline: 26199428]
30. Steele Gray C, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
31. Medicare and Medicaid programs; Electronic Health Record Incentive Program-Stage 3 and Modifications to Meaningful Use in 2015 Through 2017. Fed Regist 2015 Oct;80(200):62761-62955.
32. Public law 114–255. Library of Congress. URL: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf [accessed 2022-08-08]
33. Adler-Milstein J, Nong P. Early experiences with patient generated health data: health system and patient perspectives. J Am Med Inform Assoc 2019 Oct 01;26(10):952-959 [FREE Full text] [doi: 10.1093/jamia/ocz045] [Medline: 31329886]
34. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
35. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
36. Smith CA, Coller RJ, Dean SM, Sklansky D, Hoonakker PLT, Smith W, et al. Parent perspectives on pediatric inpatient OpenNotes. AMIA Annu Symp Proc 2019;2019:812-819 [FREE Full text] [Medline: 32308877]
37. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]

38. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
39. Internet/broadband fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2021-05-06]
40. Khoong EC, Rivadeneira NA, Hiatt RA, Sarkar U. The use of technology for communicating with clinicians or seeking health information in a multilingual urban cohort: cross-sectional survey. J Med Internet Res 2020 Apr 06;22(4):e16951 [FREE Full text] [doi: 10.2196/16951] [Medline: 32250280]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
CMC: children with medical complexity
CPCC: Complex Primary Care Clinic
EHR: electronic health record

Edited by G Eysenbach; submitted 02.06.21; peer-reviewed by M Nomali, C Steele Gray, R Keith, K Blondon; comments to author 04.02.22; revised version received 03.06.22; accepted 14.07.22; published 23.08.22.

Please cite as:
Lin JL, Huber B, Amir O, Gehrmann S, Ramirez KS, Ochoa KM, Asch SM, Gajos KZ, Grosz BJ, Sanders LM
Barriers and Facilitators to the Implementation of Family-Centered Technology in Complex Care: Feasibility Study
J Med Internet Res 2022;24(8):e30902
URL: https://www.jmir.org/2022/8/e30902
doi:10.2196/30902
PMID:35998021

Original Paper

The Technology Acceptance of Video Consultations for Type 2 Diabetes Care in General Practice: Cross-sectional Survey of Danish General Practitioners

Daniel Cæsar Torp[1], BMedSci; Annelli Sandbæk[1], MD, PhD; Thim Prætorius[1], PhD

[1] Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark

Corresponding Author:
Thim Prætorius, PhD
Steno Diabetes Center Aarhus
Aarhus University Hospital
Palle Juul-Jensens Boulevard 11
Aarhus, 8200
Denmark
Phone: 45 20539002
Email: thipra@rm.dk

Abstract

Background: During the COVID-19 pandemic, video consultations became a common method of delivering care in general practice. To date, research has mostly studied acute or subacute care, thereby leaving a knowledge gap regarding the potential of using video consultations to manage chronic diseases.

Objective: This study aimed to examine general practitioners' technology acceptance of video consultations for the purpose of managing type 2 diabetes in general practice.

Methods: A web-based survey based on the technology acceptance model measuring 4 dimensions—perceived usefulness, perceived ease of use, attitude, and behavioral intention to use—was sent to all general practices (N=1678) in Denmark to elicit user perspectives. The data were analyzed using structural equation modeling.

Results: The survey sample comprised 425 general practitioners who were representative of the population. Structural equation modeling showed that 4 of the 5 hypotheses in the final research model were statistically significant ($P<.001$). Perceived ease of use had a positive influence on perceived usefulness and attitude. Attitude was positively influenced by perceived usefulness. Attitude had a positive influence on behavioral intention to use, although perceived usefulness did not. Goodness-of-fit indices showed acceptable fits for the structural equation modeling estimation.

Conclusions: Perceived usefulness was the primary driver of general practitioners' positive attitude toward video consultations for type 2 diabetes care. The study suggests that to improve attitude and technology use, decision-makers should focus on improving usefulness, that is, how it can improve treatment and make it more effective and easier.

(J Med Internet Res 2022;24(8):e37223) doi:10.2196/37223

KEYWORDS

video consultations; telemedicine; diabetes; chronic diseases; general practice; technology acceptance; technology acceptance model

Introduction

Background

Technological change and the use of new technologies in health care are driven by objectives to increase access to health care, reduce care costs, coordinate health care, and facilitate chronic disease prevention and management [1]. The COVID-19 pandemic, caused by SARS-CoV-2 infection, has spurred health care systems to rapidly change from delivering in-person care to using different types of web-based care [2-4] such as video consultations [5]. Within the primary care sector, the uptake of video consultations has increased [6], and general practitioners' use of the technology has internationally moved from being used in pilot projects to wider-scale use [7-9]. The care potential of using video consultations in general practice is considered high [10,11], and this technology holds the potential to disrupt how health care is delivered in the primary care sector [12].

The recent uptake of video consultations in general practice is intriguing as the use of new health care technology and its implementation typically takes years [5,13]. This is because digital-first approaches to primary care could increase general practice workload [14] or threaten professional autonomy [15]. Similar to the hospital sector [16,17], knowledge about the impact of video consultations on general practice is in its infancy, and the literature is particularly short on quantitative studies [18]. The nascent literature finds that offering video consultations constitutes a significant change in how health care professionals deliver and patients receive care [19]. Research into factors that influence the implementation of video consultations in routine practice finds that, for instance, training is an important facilitator [20], and hesitance to change is an equally important barrier [21]. Research suggests that general practitioner characteristics (eg, age and sex) do not influence use, although working in larger practices makes it more likely [22,23]. Interaction and communication between patients and general practitioners during video consultations are usually effective [24,25]. However, patients and practitioners report mixed user experiences but with the important point that user ratings depend on the context in which video consultations are used [26-31]. Younger patients were found to be more likely to request or be offered a web-based visit [32].

However, research has not systematically elicited general practitioners' attitudes toward video consultations or their perceptions of the ease of use or usefulness in general practice. This research gap is unfortunate as it is well established in IT literature that attitude and perception influence physicians' use of other types of health care technology such as electronic patient records or telemedicine [33-35]. The technology acceptance model (TAM) has proven to be a robust model through rigorous empirical testing within and beyond health care [36,37]. TAM is capable of studying user attitudes and perceptions and has good predictive power of health technology use [38]. Central to the original TAM [39] and later extensions [40] is that the behavioral intention (BI) to use technology is influenced by users' ratings of perceived usefulness (PU), perceived ease of use (PEOU), and attitude toward the technology. Importantly, BI to use predicts actual user behavior [41,42].

Using the insight that chronic disease prevention and management are key drivers of technological change, this paper studies the potential of using video consultations in general practice to manage type 2 diabetes for 3 reasons. First, type 2 diabetes is a chronic disease for which video consultation appears promising in general practice [43-45]. Second, previous research on the use of video consultations in general practice has mostly studied acute or subacute or out-of-hours care and, to a much lesser extent, the management of chronic care taking place during regular hours [17,25,31]. Third, it is important to find care models capable of delivering high-quality and efficient type 2 diabetes care in general practice [46,47] as the disease prevalence is increasing [48] and people living with type 2 diabetes are at higher risk of developing complications [49].

The aim of this paper is to use TAM to study general practitioners' technology acceptance of video consultations to manage type 2 diabetes in general practice. The hypotheses were that higher levels of attitude, PU, and PEOU positively affect general practitioners' BI to use video consultations to manage type 2 diabetes. Bringing to bear TAM on video consultations in general practice allows exploring the potential of using the technology for a type of chronic care where health care systems need to find new ways of increasing health care access and cutting care costs.

Research Model and Hypotheses

The research model (Figure 1) builds on TAM [39] and posits that general practitioners' perception of the degree to which video consultations used to manage type 2 diabetes are easy to use affects both perceptions of usefulness and attitudes toward using the technology. General practitioners' attitudes are also influenced by their perception of how useful the technology is. Ultimately, general practitioners' intention to use video consultations to manage type 2 diabetes can be explained by their attitude toward the technology and PU. The following develops 5 hypotheses by combining research insights on TAM, general practitioners, and the primary health care domain.

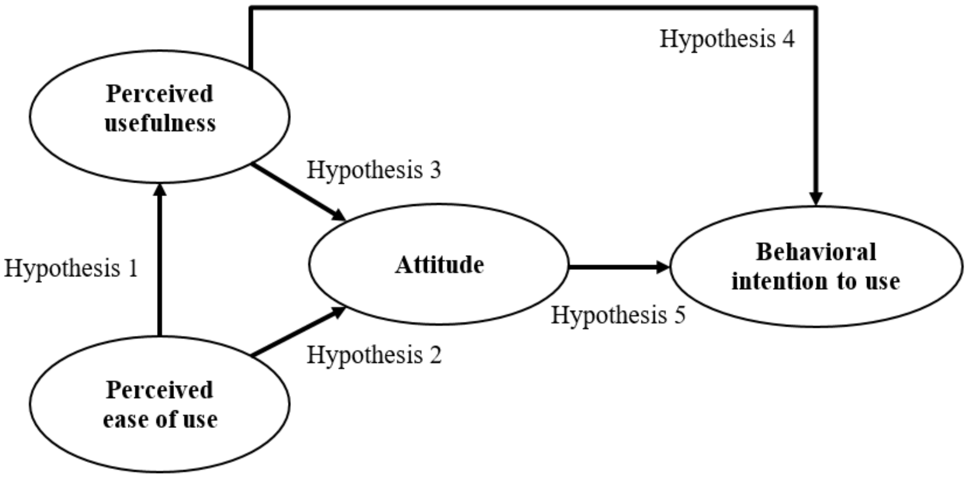
PEOU influences BI to use indirectly through both attitude and PU. A high PEOU represents the belief that using the technology will require little to no effort [39]. PU concerns the extent to which a user believes that the technology can improve or make their work more effective and easier and how it will be advantageous over the current practice. The relationship between PEOU and PU is expected to be positive as health care studies find that a higher level of PEOU leads to higher ratings of P [50-52]. Moreover, studies have shown that when a technology is perceived as easy to use, the attitude toward the technology is more positive [40,52]. The attitudinal component of the model measures an individual's affective response to adopting a new technology. Attitude concerns the extent to which a user finds that using the technology is a good idea, beneficial, or unpleasant for the way they work [39]. PU is considered particularly important in general practice [53,54], and research using TAM finds that physicians' PU influences attitudes toward health care technology [55,56]. Thus, 3 hypotheses about PEOU, PU, and attitude were formed:

- Hypothesis 1: PEOU has a positive impact on the PU of video consultations for type 2 diabetes care.
- Hypothesis 2: PEOU has a positive impact on attitudes toward video consultations for type 2 diabetes care.
- Hypothesis 3: PU has a positive impact on attitude toward video consultations for type 2 diabetes care.

The BI to use represents an individual's intention to use a new technology [41]. BI to use is an important component as it is a proxy capable of predicting subsequent actual user behavior in health care and beyond [33,41,42]. According to TAM, the extent to which users perceive a technology to be useful is directly influenced by their ratings of BI to use [38]. In the context of general practice, research has found a positive relationship between PU and BI to use [35,57-59]. Similarly, TAM suggests that the attitude of a user manifests itself as a positive or negative view of the BI to use technology. Research in the domain of primary health care finds that attitude influences the BI to use health care technology [23,60,61]. Thus, 2 hypotheses about PU, attitude, and BI to use were formulated:

- Hypothesis 4: PU has a positive impact on the BI to use video consultations for type 2 diabetes care.
- Hypothesis 5: Attitude toward video consultations for type 2 diabetes care has a positive impact on the BI to use the technology.

Figure 1. Research model based on the technology acceptance model.



Methods

Research Design and Setting

Data were collected through a cross-sectional web-based survey distributed to all general practitioners in Denmark (n=3326). The Danish health care system is mostly tax financed, and citizens can receive care from general practice free of per service charge. Danish general practitioners are self-employed but work on contracts for the public funder. Most general practitioners work in partnership practices, and their income is generated as a combination of fee for service and capitation [62]. The incentive for Danish general practitioners to use video consultations increased during the COVID-19 pandemic because of an agreement between the General Practitioners' Organization (negotiating on behalf of Danish general practitioners) and the Danish Regions (responsible for procuring health services), which agreed on a fee for service to general practitioners to provide video consultations to patients.

Survey Measures

The main measures (13 items) central to our hypotheses originated from TAM [39] and health care studies [55] to ensure the validity of the measures. The measures were adapted to the specific context of general practice and video consultations, translated into Danish, and repeatedly examined to ensure consistency. PU, attitude, and BI to use were measured using 3 items each, and PEOU was measured using 4 items (Textbox 1). An item each in the attitude and BI to use dimensions was negatively worded to reduce the risk of agreement bias [55]. All items were measured on 5-point Likert scales, with scores ranging from 1 (strongly disagree) to 5 (strongly agree). For PEOU, the items were worded according to the user status of the respondent (user vs nonuser of video consultations) to make the formulation relevant to the respondent. Respondents were able to skip questions or choose *do not know* (the latter being treated as missing data in subsequent analyses). Demographic measures (12 items) such as age and sex were collected to analyze the representativeness of the study sample in comparison with the total population of general practitioners. Before distribution and to test face validity, the survey was evaluated and revised according to inputs from 5 general practitioners working in each of the 5 Danish Regions.

Textbox 1. Items used in the research model.

- Perceived usefulness (PU)
  - PU1: can *improve* my treatment
  - PU2: can make my treatment more *effective*
  - PU3: can make my treatment *easier*
- Perceived ease of use (PEOU; worded differently for nonusers of video consultations as illustrated in brackets)
  - PEOU1: *learning* to use was (would be) easy
  - PEOU2: *(would be) easy to get* software to do what I need
  - PEOU3: (would be) easy to *master*
  - PEOU4: (would be) easy to *use*
- Attitude (ATT)
  - ATT1: using is a *good idea*
  - ATT2: using is *unpleasant*
  - ATT3: using is *beneficial*
- Behavioral intention (BI)
  - BI1: intend to use as *often as possible*
  - BI2: even when possible, *do not intend* to use
  - BI3: would *use to the extent possible*

Recruitment and Data Collection

The survey was administered using SurveyXact (Rambøll Management) [63]. To identify general practices, a list of all 1718 general practices in Denmark was obtained from MedCom (a provider of Danish public health care systems) [64] in January 2021. Of these 1718 practices, 44 (2.56%) general practices were excluded as they were managed by parties outside the target group of our study (eg, by Danish Regions). In total, 1674 general practices, representing 3326 general practitioners, were available for distribution [65].

The survey was distributed to general practices as an electronic letter on January 7, 2021, via the Danish public electronic mailbox system (e-Boks Business) using publicly available data from MedCom. The letter contained information about the study and a survey link. Participants were informed about data protection measures, anonymity of participation, and the option to be paid—DKK 276.72 (US $44) based on a General Practitioners'' Organization tariff—for the 20 minutes it maximally takes to complete the survey. The letter was addressed to the clinic, and all trained general practitioners were encouraged to participate. Unfortunately, it was not possible to contact each general practitioner directly as this information was not publicly available. The survey link was open and only available in a letter to ensure anonymity and availability for all general practitioners in a clinic. Data entry for payments was conducted in a separate survey to preserve anonymity. Two reminders were sent on January 21, 2021, and February 2, 2021. The data collection ended on February 7, 2021.

The Committee of Multipractice Studies in General Practice (journal number 25-2020) evaluated the study and recommended that general practitioners participate in the survey. This study was reported to the Danish Data Protection Agency (journal number 1-16-02-343-20).

Ethics Approval

The Research Ethics Committees for Central Denmark Region (1-10-72-181-20) concluded that the study could be conducted without approval from the committee as “According to the Consolidation Act on Research Ethics Review of Health Research Projects, Consolidation Act number 1083 of 15 September 2017, section 14(2) notification of questionnaire surveys or medical database research projects to the research ethics committee system is only required if the project involves human biological material.”

Data Analysis

Data were analyzed using Stata (version 17.0; StataCorp) [66]. To compare sample demographics with the population of general practitioners, we analyzed the latter using registry data made available by the Danish Health Data Authority [67]. The measures used in TAM were analyzed for normality distribution, internal consistency, convergent validity, and discriminant validity. Normality was examined by calculating skewness, kurtosis, and the Mardia multivariate kurtosis test. Internal consistency was assessed using Cronbach α with an acceptable threshold of .70 [68]. Confirmatory factor analysis was performed to determine model validity. Factor loadings of ≥0.7 were deemed acceptable [69]. Subsequently, we explored the research model using structured equation modeling [70], which is standard in the data analysis of TAM [37]. We used quasi-maximum likelihood as the estimator, with Satorra-Bentler adjustments because of our findings of nonnormality for some of the measures [71]. *P*<.05 was set as the threshold for statistical significance.

We report the unstandardized and standardized path coefficients from structured equation modeling. The unstandardized path coefficients reflect the expected (linear) change in the dependent variable with each unit change in the independent variable, given the other variables in the model. The standardized path coefficients express relationships in the same unit; that is, SDs. The interpretation is that when an independent variable (eg, PU) changes by 1 SD, then the dependent variable (eg, BI to use) changes by an SD as well. By placing all coefficients in the same unit, the SDs for different variables measured in different metrics become interpretationally equivalent.

Results

Demographic Characteristics

A total of 457 general practitioners answered the survey, from which 32 (7%) incomplete responses were excluded, resulting in 425 (93%) respondents. The sample represented 12.78% (425/3326) of all Danish general practitioners. The sample represented 18.82% (315/1674) of Danish general practices. Compared with the population of general practitioners, Pearson chi-square tests showed that the individual characteristics of the study sample (ie, sex and age groups) were representative of the population not participating (Table 1). The sample differed with regard to general practice characteristics (ie, clinic and municipality type) as general practitioners from more partnership practices participated than from solo practices, and a larger share of general practitioners working in practices in the capital area participated. The incomplete responses had similar demographics to the complete responses, with most (23/32, 72%) dropping out during or directly after the demographic items.

Table 1. Overview of respondents in sample and comparison with the remaining population.

Characteristics[a]	Survey sample (n=425), n (%)	Population not in the sample (n=2901), n (%)	Pearson chi-square (*df*)
Sex (female)[b]	226 (53.1)	1659 (57.1)	0.2 (1)
Age group (years)[b]			0.8 (6)
30-39	26 (6.3)	205 (7.1)	
40-44	75 (18.1)	577 (20)	
45-49	100 (24.2)	614 (21.2)	
50-54	59 (14.3)	416 (14.4)	
55-59	64 (15.5)	433 (15)	
60-64	57 (13.8)	387 (13.4)	
≥65	33 (8)	260 (9)	
Municipality type where general practitioners work[c,d]			0.0 (4)
Capital area	133 (31.3)	789 (25.5)	
Large city	63 (14.8)	392 (12.7)	
Province city	88 (20.7)	754 (24.4)	
Suburban	70 (16.5)	507 (16.4)	
County	71 (16.7)	654 (21.1)	
Clinic type[c]			<0.001 (2)
Solo clinic	105 (25.1)	447 (35.7)	
Cooperation clinic	52 (12.4)	145 (11.6)	
Partnership clinic	419 (98.5)	659 (52.7)	

[a] Missing data in the population not in the sample and in the survey sample means that sums do not add to the population of general practitioners (N=3326), general practices (N=1674), and study sample (N=425).

[b] Population data from General Practitioners' Organization [65].

[c] Population calculated from data by the Danish Health Data Authority [67].

[d] Municipality types based on the definition by Statistics Denmark [72].

Measurements Based on the TAM

Table 2 presents the mean values (SD) of the 4 dimensions and the items from TAM. On a 5-point Likert scale, the highest mean value was PEOU 3.76 (SD 0.86) and ATT 3.48 (SD 0.92), thus indicating that respondents were confident that they, for instance, can use video consultations to manage type 2 diabetes and that the technology was a good idea. The mean values for PU 2.99 (SD 0.96) and BI to use 3.06 (SD 1.04) were similar, and the answers averaged around neither agreeing nor disagreeing. Across the studied dimensions and items, the data variability around the mean of the study sample was approximately 1 point on a 5-point Likert scale.

Table 2. Means and internal consistency of items in the research model (N=425).

Item	Participants, n (%)	Values, mean (SD)	Cronbach α
PU[a]			
PU1: can *improve* my treatment	389 (91.5)	2.70 (0.97)	.86
PU2: can make my treatment more *effective*	397 (93.4)	3.01 (1.07)	.78
PU3: can make my treatment *easier*	396 (93.2)	3.24 (1.13)	.85
PU: all usability items	379 (89.2)	2.99 (0.96)	.88
PEOU[b]			
PEOU1: *learning* to use was (would be) easy	417 (98.1)	3.99 (0.95)	.85
PEOU2: (would be) *easy to get* software to do what I need	401 (94.4)	3.81 (0.98)	.84
PEOU3: (would be) easy to *master*	412 (96.9)	3.91 (0.91)	.83
PEOU4: (would be) easy to *use*	372 (87.5)	3.28 (1.1)	.92
PEOU: all ease of use items	359 (84.5)	3.76 (0.86)	.89
ATT[c]			
ATT1: using is a *good idea*	409 (96.2)	3.29 (1.15)	.63
ATT2: using is *unpleasant*	398 (93.6)	2.04 (0.96)	.92
ATT3: using is *beneficial*	397 (93.4)	3.13 (1.09)	.68
ATT: all attitude items[d]	380 (89.4)	3.48 (0.92)	.83
ATT1+3: ATT excluding ATT2	393 (92.5)	3.21 (1.08)	.92
BI[e] **to use**			
BI1: intend to use as *often as possible*	403 (94.8)	2.66 (1.12)	.82
BI2: even when possible, *do not intend* to use	404 (95.1)	2.61 (1.2)	.88
BI3: would *use to the extent possible*	402 (94.6)	3.12 (1.12)	.78
BI: all intention items[f]	383 (90.1)	3.06 (1.04)	.88

[a] PU: perceived usefulness.

[b] PEOU: perceived ease of use.

[c] ATT: attitude.

[d] The mean represents all ATT variables with ATT2 reversed because of its negative wording.

[e] BI: behavioral intention.

[f] The mean represents all BI variables with BI2 reversed because of its negative wording.

The internal consistency of the items that comprise the 4 dimensions in TAM had Cronbach α >.8 (Table 2). Cronbach α values of ≥.7 indicate acceptable internal consistency. Although the internal consistency of attitude was .83, this value should be interpreted with caution. The right-hand column of Table 2 shows the effect of removing 1 of the 3 items on Cronbach α; that is, for the attitude dimension, the Cronbach α drops to .63 and .68 when removing items 1 and 2 and increases to .92 when removing item 3. In addition to attributing this change in internal consistency to this analytical finding, free-text remarks by some respondents indicated that the negative wording of item 3 could be confusing and challenging to answer. On the basis of logical reasoning [73] and to reflect the attitude dimension more accurately, we excluded item 2 from the subsequent analysis.

To determine the correct structural equation modeling estimation method, we calculated the skewness and kurtosis of all the measures to examine normality. The results showed a mild degree of skewness (ranging from −0.971 to 0.232) with moderate kurtosis (ranging from 2.134 to 3.841). Normality was further evaluated using the Mardia multivariate kurtosis test, in which all dimensions failed except attitude, thereby indicating nonnormally distributed measures (PU 20.4, χ^2_1=90.9, P<.001; PEOU 43.3, χ^2_1=694.6, P<.001; attitude 8.22, χ^2_1=0.3, P=.57; BI 17.9, χ^2_1=26.0, P<.001). As nonnormality invalidates the assumption for the maximum likelihood method of structural equation modeling estimation, we used Satorra-Bentler adjustments to relax the assumption of normality. The measures in TAM were also assessed for convergent validity and discriminant validity (Table 3).

The measures were further validated using a confirmatory factor analysis that showed factor loadings >0.7, except for the item PEOU4—*easy to use* (0.63). PEOU4 was also an outlier in terms of missing data, with 12.7% (53/425) of missing responses, leading to the suspicion that the data were not missing at random. We excluded PEOU4 from the analysis and ran a new confirmatory factor analysis, which had factor loadings ranging from 0.77 to 0.92, thereby confirming that the latent variables of TAM were explained by the observed variables. Goodness-of-fit indices confirmed that the confirmatory factor analysis was a good fit for the data (χ^2_{38}=51.5, χ^2/df=1.4; P=.07; root mean squared error of approximation 0.033 [recommended value <0.05]; standardized root mean square residual 0.024 [recommended value <0.08]; comparative fit index 0.995 [recommended value >0.95]) [74]. The final research model included data from 76.9% (327/425) of respondents.

Table 3. Correlations between dimensions and items in the research model.

Item	PU[a]	PEOU[b]	ATT[c]	BI[d]
PU				
PU1	0.731	0.213	0.702	0.640
PU2	0.824	0.335	0.761	0.700
PU3	0.747	0.328	0.785	0.701
PEOU				
PEOU1	0.204	0.803	0.250	0.378
PEOU2	0.181	0.826	0.265	0.359
PEOU3	0.224	0.853	0.301	0.410
PEOU4	0.477	0.607	0.553	0.551
ATT				
ATT1	0.800	0.419	0.844	0.789
ATT3	0.801	0.369	0.844	0.765
BI				
BI1	0.703	0.454	0.754	0.813
BI2	0.613	0.441	0.668	0.711
BI3	0.709	0.426	0.750	0.773

[a] PU: perceived usefulness.

[b] PEOU: perceived ease of use.

[c] ATT: attitude.

[d] BI: behavioral intention.

Hypothesis Testing

We used structural equation modeling to analyze our hypotheses and the final research model. The goodness-of-fit indices model showed an acceptable fit (Table 4).

Analysis of the research model using unstandardized coefficients (Figure 2; Table 5) showed that the original paths of the model were significant (P<.005), except for the path from PU to BI to use (P=.84). PEOU had a positive influence on PU (β=.26, 95% CI 0.14-0.38) and attitude (β=.16, 95% CI 0.08-0.24). PU had a positive influence on attitude (β=1.22, 95% CI 1.09-1.36). The influence of attitude and PU on BI to use was also positive (β=.82, 95% CI 0.52-1.12; β=.04, −0.38 to 0.47); however, the latter was statistically insignificant. The calculated R^2 values (Figure 2) showed that 82% of the variance in BI to use was explained by attitude and PEOU, with attitude having the strongest influence. Standardized coefficients showed similar results (Figure 2; Table 6) and indicated that the strongest relationship existed between PU and attitude and between attitude and BI.

Table 4. Fit indices for structural equation modeling estimation.

Fit index	Structural equation modeling model with Satorra-Bentler	Recommended value [74,75]
Chi-square (*df*)	63.59 (39)	N/A[a]
Chi-square/*df*	1.63	<3.0
P value>chi-square (*df*)	0.008	>0.05
Root mean squared error of approximation	0.044	<0.05
Comparative fit index	0.991	>0.95
Tucker-Lewis index	0.987	>0.95
Standardized root mean square residual	0.036	<0.08

[a] N/A: not applicable (the literature on structural equation modeling does not recommend a value).

Figure 2. Results of structural equation modeling, unstandardized (and standardized) coefficients. **P*<.001.

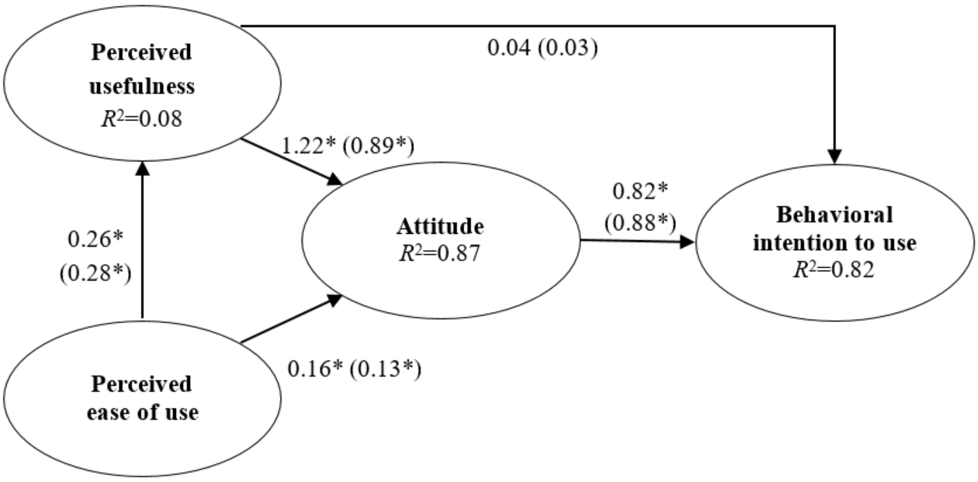


Table 5. Structural equation modeling estimation, unstandardized coefficients[a].

Path	β coefficient	*z* value	*P* value	95% CI
PEOU[b]→PU[c]	.26	4.26	<.001	0.14 to 0.38
PU→attitude	1.22	17.44	<.001	1.09 to 1.36
PEOU→attitude	.16	4.01	<.001	0.08 to 0.24
PU→BI[d]	.04	0.20	.84	−0.38 to 0.47
Attitude→BI	.82	5.35	<.001	0.52 to 1.12

[a] Satorra-Bentler adjusted; unstandardized coefficients.

[b] PEOU: perceived ease of use.

[c] PU: perceived usefulness.

[d] BI: behavioral intention.

Table 6. Structural equation modeling estimation, standardized coefficients[a].

Path	β coefficient	*z* value	*P* value	95% CI
PEOU[b]→PU[c]	.28	4.09	<.001	0.15 to 0.42
PU→attitude	.89	38.19	<.001	0.84 to 0.94
PEOU→attitude	.13	4.09	<.001	0.07 to 0.19
PU→BI[d]	.03	0.19	.85	−0.31 to 0.37
Attitude→BI	.88	5.54	<.001	0.57 to 1.19

[a] Satorra-Bentler adjusted; standardized coefficients.

[b] PEOU: perceived ease of use.

[c] PU: perceived usefulness.

[d] BI: behavioral intention.

Discussion

Principal Findings and Comparison With Prior Work

To explore the potential of using video consultations to provide type 2 diabetes care in general practice, we used insights from technology adoption [36-40] to systematically elicit the technology acceptance of general practitioners. From our survey of Danish general practitioners, we found support for 4 of the 5 research hypotheses (standardized and unstandardized path coefficients).

First, our findings suggest that PU is the primary driver of a positive attitude toward using video consultations to provide type 2 diabetes in general practice (hypothesis 3 accepted: unstandardized β=1.22, 95% CI 1.09-1.36). Similarly, earlier research in general practice found that this relationship appeared to be highly important [53,54]. The unstandardized path coefficient indicates that increasing the PU of the technology by 1 unit will increase the attitude by 1.22 units, given the other variables in the model. The standardized coefficient (β=.89, 95% 0.84-0.94) shows that a change of 1 SD in PU leads to an increase by 0.89 SDs in attitude. Second, attitude toward the technology is positively influenced by general practitioners' PEOU (hypothesis 2 accepted: unstandardized β=.16, 95% CI 0.08-0.24); however, the impact is lower than that for PU (β=1.22 vs β=.16). This finding mirrors previous studies that found that PU, not PEOU, is the primary driver of users' attitudes toward health care technology. A reason is that ease of use is not necessarily a sufficiently large benefit to offset the difficulties of integrating new technology into established work routines [76]. Another reason is that the importance of a technology that is easy to use tends to decrease with general technology use [38,55,56].

Third, our analysis confirmed the expectation that general practitioners' PU of video consultations would be positively influenced by their ratings of PEOU (hypothesis 1 accepted: unstandardized β=.26,95% CI 0.14-0.38). This mirrors findings from studies of other types of health care technology [50-52]. The relatively small impact of PEOU may be attributed to the high education level of Danish general practitioners who use IT technologies daily to deliver care, such as electronic patient records, and thus have a basic level of IT skills that could be speculated to give them confidence in learning new technologies. Fourth, the BI to use video consultations to provide type 2 diabetes was positively influenced by the attitude toward the technology (hypothesis 5 accepted: unstandardized β=.82, 95% CI 0.52-1.12). This particular relationship has also been found in other studies in the domain of primary health care [23,60,61]. Attitude is a central driver that corresponds to other influential theories of behavior change, such as the theory of planned behavior [77]. Fifth, our research model links PU to BI to use; however, the positive influence was statistically insignificant (hypothesis 4 rejected: unstandardized β=.04, −0.38 to 0.47). Compared with the impact of attitude, the influence of the PU of video consultations was also less influential (β=.82 vs β=.04). Studies from general practice generally report that PU has a positive influence on BI to use [35,57-59]. However, these studies do not include the attitude dimension from the original model [39] in their research models and, thus, do not address the relative importance. Our findings indicate that the BI to use video consultations for type 2 diabetes care is primarily the result of the positive impact PU has on attitude.

By studying chronic care in our context—type 2 diabetes—our research findings contribute to an emerging literature on video consultations in general practice that has hitherto mostly studied acute or subacute or out-of-hours care [17,25,31]. A major strength of the study is that the findings build on TAM, which is a robust model [36,37] with good predictive power for health technology use [38]. The findings are also supported by goodness-of-fit tests, showing that the research model has an acceptable fit for structural equation modeling estimation. A strength of our analysis is that it did not rely on the assumption that the measures were normally distributed as we used the Satorra-Bentler adjustments in the structural equation modeling.

Practical Implications

The potential of using video consultations in general practice to deliver chronic disease management is promising [1,10,11] and could fundamentally change how the primary care sector delivers care [12,19]. Type 2 diabetes is a chronic disease for which video consultations in general practice are particularly relevant [43-45] because, as a new care model, it can deliver high-quality, efficient care [46,47] at a time when the prevalence of diabetes is increasing [48]. Our findings (standardized and unstandardized path coefficients in the research model) indicate that the strongest positive relationships are between PU and attitude and between attitude and BI to use. This suggests that if a policy maker wants to increase general practitioners' use of video consultations to provide type 2 diabetes care, they must ensure that the technology is useful in general practice as it will have a positive influence on their attitude, which, in turn, will positively affect their intention to use the technology. Policy makers interested in scaling up video consultations could benefit from looking into the items of the dimensions that constitute the research model. For example, to improve PU, policy makers should find solutions to three questions: how can it be ensured that video consultations (1) improve treatment, (2) make treatment more effective, and (3) make treatment easier?

Relatedly, our findings provide suggestions for mitigating change hesitance, which remains a barrier to implementing video consultations in routine practice [21]. As research shows that working in larger practices—but not individual characteristics such as age or sex—increases the likelihood that a general practitioner uses video consultation [22,23], it appears relevant to explore the perceptions of small and large practices separately. Using the example of PU, small and large practices may differ in the ways in which video consultations can improve and make treatment easier. These insights are important as data from, for example, the Danish Health Authority show a decrease in the use of video consultations in general practice from 2020 to 2021 [78], which suggests that general practitioners use the technology but also that it is not yet a regular work routine in general practice. Moreover, continuous improvement of the technology and its use in practice is central as there is a risk that this new care model increases general practitioner workload, and there may be a need to allocate more resources to implement digital-first pathways [14]. To the latter end, research finds that training facilitates the implementation of video consultations in routine practice [20].

Limitations

Two modifications were made to the original TAM, underlining the final research model. First, an item (attitude item 2) was removed as it decreased the Cronbach α of the attitude dimension. Another item (PEOU4) was dropped because of the low factor loading from the confirmatory factor analysis. To assess the extent to which removing these items changed the findings, a structural equation modeling estimation, including these items, was performed, which showed path coefficients very similar to our final model, thereby supporting the accuracy of the final structural equation model. Second, structural equation modeling estimations were not performed with all respondents as those skipping questions were omitted. Running a structural equation modeling estimation that included respondents with missing answers resulted in similar path coefficients but had poorer goodness of fit. The final research model met the recommended indices of the goodness of fit but failed the chi-square test. Failing the chi-square test is a known issue with structural equation modeling, which, similar to our study, has a high number of respondents and survey answers that are not normally distributed [75]. The issue of nonnormality was addressed using Satorra-Bentler adjustments.

With the widespread research validation of TAM in combination with acceptable goodness-of-fit indices, the final research model is considered valid. However, as this study surveyed general practitioners from a tax-financed health care system, the findings may be most generalizable to countries with similar health care systems such as the English National Health System. Some authors also raise the concern that the original TAM and later extensions lack precision in health care because of their inability to consider the influence of external variables and barriers to technology acceptance [36] such as psychological ownership of IT [79] or social norms [55]. Nevertheless, for the purposes of this study, the research model was kept simple for 2 main reasons. First, findings from health care that extend TAM only result in a relatively modest increase in explanatory power [55]. Second, getting general practitioners to answer surveys is difficult [43], and including other variables to increase the precision a little would likely come at the expense of a lower response rate. More questions also increased the risk of respondent fatigue and missing answers.

The relatively low response rate of 12.8% of all 3326 Danish general practitioners increased the risk of selection bias. Nevertheless, it improved confidence in the findings that the individual characteristics of the sample of general practitioners were comparable with the population, and the share of respondents in the sample who used video consultations was similar to that of other sources [78]. This finding supports the generalizability of our results. The difficulty in getting Danish general practitioners to participate in survey research is an explanation as they operate as for-profit firms and are often on a tight schedule [62]. The survey was also distributed during the COVID-19 pandemic when other surveys of general practitioners had similar low response rates [22,43,80]. It could be speculated that general practitioners with the strongest positive or negative attitudes toward technology were more likely to participate. Univariate normality tests of the items in the attitude dimension, as mentioned previously, showed that the respondents' attitudes were relatively normally distributed and did not only represent the most negative or positive attitudes toward video consultations used for diabetes care.

The study design was cross-sectional and, thus, only capable of capturing the views of general practitioners at the time of data collection. Although the cross-sectional design is standard in most studies on TAM [37,38], longitudinal studies are generally recommended to assess changes over time to make study findings more robust. Collecting data on the variables in TAM from the same source (ie, general practitioners) makes common method bias [81] a potential risk in the study. However, common method bias is of modest importance here as the research model asks about the intention to use rather than actual use.

Conclusions

This study explored the potential of using video consultations to provide type 2 diabetes care in general practice by eliciting the technology acceptance of a representative survey sample of Danish general practitioners. On the basis of TAM, our study suggests 2 main drivers: PU positively affects attitude toward using video consultations for diabetes care, and attitude positively affects the BI to use the technology. For policy makers interested in scaling up general practitioners' use of video consultations to provide diabetes care, our findings indicate that they should emphasize how the technology can improve treatment and make it more effective and easier. To this end, policy makers may need to explore what these aspects of usefulness mean to general practitioners working in different organizational contexts.

Acknowledgments

The authors are grateful for the insights provided by the reference group of 5 general practitioners from 5 Danish regions: Rasmus Dahl-Larsen, Michel Kjeldsen, Gitte Krogh Madsen, Janus Nikolaj Laust Thomsen, and Mogens Vestergaard. The authors thank Henrik Støvring, Lasse Bjerg Hansen, Daniel Witte, and Kasper Norman for their methodological comments.

The authors would also like to thank the Quality and Training Committees of North Denmark Region, Central Denmark Region, Region of Southern Denmark, Region Zealand, the Capital Region, and the Lundbeck Foundation Scholarship for General Practice. Support was provided by the Steno Diabetes Center Aarhus, Aarhus University Hospital, which is partially funded by an unrestricted donation from the Novo Nordisk Foundation.

Authors' Contributions

TP and AS conceptualized and designed the study. TP and DCT collected the data, conducted the statistical analyses, and wrote the first draft of the manuscript. All the authors critically revised the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

References

1. Litwin AS. Technological Change in Health Care Delivery: Its Drivers and Consequences for Work and Workers. UC Berkeley Labor Center. 2020 Jun. URL: https://laborcenter.berkeley.edu/wp-content/uploads/2020/07/Technological-Change-in-Health-Care-Delivery.pdf [accessed 2022-06-01]
2. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
3. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
4. An MH, You SC, Park RW, Lee S. Using an extended technology acceptance model to understand the factors influencing telehealth utilization after flattening the COVID-19 curve in South Korea: cross-sectional survey study. JMIR Med Inform 2021 Jan 08;9(1):e25435 [FREE Full text] [doi: 10.2196/25435] [Medline: 33395397]
5. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
6. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the nasss framework. J Med Internet Res 2021 Jan 26;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
7. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Serv Deliv Res 2018 Jun;6(21):1-136. [doi: 10.3310/hsdr06210] [Medline: 29939519]
8. Tönnies J, Hartmann M, Wensing M, Szecsenyi J, Peters-Klimm F, Brinster R, et al. Mental health specialist video consultations versus treatment-as-usual for patients with depression or anxiety disorders in primary care: randomized controlled feasibility trial. JMIR Ment Health 2021 Mar 12;8(3):e22569 [FREE Full text] [doi: 10.2196/22569] [Medline: 33709931]
9. 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]
10. Mold F, Cooke D, Ip A, Roy P, Denton S, Armes J. COVID-19 and beyond: virtual consultations in primary care-reflecting on the evidence base for implementation and ensuring reach: commentary article. BMJ Health Care Inform 2021 Jan;28(1):e100256 [FREE Full text] [doi: 10.1136/bmjhci-2020-100256] [Medline: 33436372]
11. Car J, Koh GC, Foong PS, Wang CJ. Video consultations in primary and specialist care during the covid-19 pandemic and beyond. BMJ 2020 Oct 20;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
12. Salisbury C, Quigley A, Hex N, Aznar C. Private video consultation services and the future of primary care. J Med Internet Res 2020 Oct 01;22(10):e19415 [FREE Full text] [doi: 10.2196/19415] [Medline: 32812887]
13. Paré G, Raymond L, de Guinea AO, Poba-Nzaou P, Trudel MC, Marsan J, et al. Barriers to organizational adoption of EMR systems in family physician practices: a mixed-methods study in Canada. Int J Med Inform 2014 Aug;83(8):548-558. [doi: 10.1016/j.ijmedinf.2014.06.003] [Medline: 24969270]

14. Salisbury C, Murphy M, Duncan P. The impact of digital-first consultations on workload in general practice: modeling study. J Med Internet Res 2020 Jun 16;22(6):e18203 [FREE Full text] [doi: 10.2196/18203] [Medline: 32543441]
15. Walter Z, Lopez MS. Physician acceptance of information technologies: role of perceived threat to professional autonomy. Decis Support Syst 2008 Dec;46(1):206-215. [doi: 10.1016/j.dss.2008.06.004]
16. 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]
17. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: a review of reviews and applications in line with guidelines and recommendations. Digit Health 2019 Apr 23;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
18. Randhawa RS, Chandan JS, Thomas T, Singh S. An exploration of the attitudes and views of general practitioners on the use of video consultations in a primary healthcare setting: a qualitative pilot study. Prim Health Care Res Dev 2019 Jan;20:e5 [FREE Full text] [doi: 10.1017/S1463423618000361] [Medline: 29909798]
19. Wherton J, Shaw S, Papoutsi C, Seuren L, Greenhalgh T. Guidance on the introduction and use of video consultations during COVID-19: important lessons from qualitative research. BMJ Leader 2020 May 18;4(3):120-123. [doi: 10.1136/leader-2020-000262]
20. Hoffmann M, Wensing M, Peters-Klimm F, Szecsenyi J, Hartmann M, Friederich H, et al. Perspectives of psychotherapists and psychiatrists on mental health care integration within primary care via video consultations: qualitative preimplementation study. J Med Internet Res 2020 Jun 18;22(6):e17569 [FREE Full text] [doi: 10.2196/17569] [Medline: 32554369]
21. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
22. 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]
23. Jiwa M, Meng X. Video consultation use by Australian general practitioners: video vignette study. J Med Internet Res 2013 Jun 19;15(6):e117 [FREE Full text] [doi: 10.2196/jmir.2638] [Medline: 23782753]
24. Shaw SE, Seuren LM, Wherton J, Cameron D, A'Court C, Vijayaraghavan S, et al. Video consultations between patients and clinicians in diabetes, cancer, and heart failure services: linguistic ethnographic study of video-mediated interaction. J Med Internet Res 2020 May 11;22(5):e18378 [FREE Full text] [doi: 10.2196/18378] [Medline: 32391799]
25. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Sep;69(686):e595-e604 [FREE Full text] [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
26. 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]
27. 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 May 1;4(1):bjgpopen20X101020 [FREE Full text] [doi: 10.3399/bjgpopen20X101020] [Medline: 32184212]
28. Mueller M, Knop M, Niehaves B, Adarkwah CC. Investigating the acceptance of video consultation by patients in rural primary care: empirical comparison of preusers and actual users. JMIR Med Inform 2020 Oct 22;8(10):e20813 [FREE Full text] [doi: 10.2196/20813] [Medline: 32969339]
29. Bleyel C, Hoffmann M, Wensing M, Hartmann M, Friederich H, Haun MW. Patients' perspective on mental health specialist video consultations in primary care: qualitative preimplementation study of anticipated benefits and barriers. J Med Internet Res 2020 Apr 20;22(4):e17330 [FREE Full text] [doi: 10.2196/17330] [Medline: 32310139]
30. Chudner I, Drach-Zahavy A, Karkabi K. Choosing video instead of in-clinic consultations in primary care in Israel: discrete choice experiment among key stakeholders-patients, primary care physicians, and policy makers. Value Health 2019 Oct;22(10):1187-1196 [FREE Full text] [doi: 10.1016/j.jval.2019.05.001] [Medline: 31563262]
31. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
32. 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]
33. Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F. Theories informing eHealth implementation: systematic review and typology classification. J Med Internet Res 2021 May 31;23(5):e18500 [FREE Full text] [doi: 10.2196/18500] [Medline: 34057427]
34. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]

35. Saigí-Rubió F, Vidal-Alaball J, Torrent-Sellens J, Jiménez-Zarco A, López Segui F, Carrasco Hernandez M, et al. Determinants of Catalan public primary care professionals' intention to use digital clinical consultations (eConsulta) in the post-COVID-19 context: optical illusion or permanent transformation? J Med Internet Res 2021 May 31;23(6):e28944 [FREE Full text] [doi: 10.2196/28944] [Medline: 34097638]
36. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. Med Care Res Rev 2007 Dec;64(6):650-672. [doi: 10.1177/1077558707305942] [Medline: 17717378]
37. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
38. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
39. 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]
40. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
41. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
42. Sheppard BH, Hartwick J, Warshaw PR. The theory of reasoned action: a meta-analysis of past research with recommendations for modifications and future research. J Consum Res 1988 Dec;15(3):325-343. [doi: 10.1086/209170]
43. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
44. Verhoeven F, van Gemert-Pijnen L, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007 Dec 14;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]
45. Fatehi F, Menon A, Bird D. Diabetes care in the digital era: a synoptic overview. Curr Diab Rep 2018 May 10;18(7):38. [doi: 10.1007/s11892-018-1013-5] [Medline: 29748905]
46. Murphy ME, Byrne M, Galvin R, Boland F, Fahey T, Smith SM. Improving risk factor management for patients with poorly controlled type 2 diabetes: a systematic review of healthcare interventions in primary care and community settings. BMJ Open 2017 Aug 04;7(8):e015135 [FREE Full text] [doi: 10.1136/bmjopen-2016-015135] [Medline: 28780542]
47. Siegel KR, Ali MK, Zhou X, Ng BP, Jawanda S, Proia K, et al. Cost-effectiveness of interventions to manage diabetes: has the evidence changed since 2008? Diabetes Care 2020 Jul;43(7):1557-1592. [doi: 10.2337/dci20-0017] [Medline: 33534729]
48. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016 Apr 09;387(10027):1513-1530 [FREE Full text] [doi: 10.1016/S0140-6736(16)00618-8] [Medline: 27061677]
49. Larsson SC, Wallin A, Håkansson N, Stackelberg O, Bäck M, Wolk A. Type 1 and type 2 diabetes mellitus and incidence of seven cardiovascular diseases. Int J Cardiol 2018 Jul 01;262:66-70 [FREE Full text] [doi: 10.1016/j.ijcard.2018.03.099] [Medline: 29605469]
50. Venkatesh V, Zhang X, Sykes TA. “Doctors do too little technology”: a longitudinal field study of an electronic healthcare system implementation. Inf Syst Res 2011 Sep;22(3):523-546. [doi: 10.1287/isre.1110.0383]
51. Raymond L, Paré G, Ortiz de Guinea A, Poba-Nzaou P, Trudel MC, Marsan J, et al. Improving performance in medical practices through the extended use of electronic medical record systems: a survey of Canadian family physicians. BMC Med Inform Decis Mak 2015 Apr 14;15:27 [FREE Full text] [doi: 10.1186/s12911-015-0152-8] [Medline: 25888991]
52. Chaudhuri A, Saddikutti V, Prætorius T. iKure Techsoft: providing technology enabled affordable health care in rural India. Asian Case Res J 2018;22(02):385-411. [doi: 10.1142/s0218927518500165]
53. Krog MD, Nielsen MG, Le JV, Bro F, Christensen KS, Mygind A. Barriers and facilitators to using a Web-based tool for diagnosis and monitoring of patients with depression: a qualitative study among Danish general practitioners. BMC Health Serv Res 2018 Jun 27;18(1):503 [FREE Full text] [doi: 10.1186/s12913-018-3309-1] [Medline: 29945613]
54. Kivekäs E, Enlund H, Borycki E, Saranto K. General practitioners' attitudes towards electronic prescribing and the use of the national prescription centre. J Eval Clin Pract 2016 Oct;22(5):816-825. [doi: 10.1111/jep.12548] [Medline: 27121026]
55. Chau PY, Hu PJ. Investigating healthcare professionals’ decisions to accept telemedicine technology: an empirical test of competing theories. Inf Manag 2002 Jan;39(4):297-311. [doi: 10.1016/S0378-7206(01)00098-2]
56. Hu PJ, Chau PY, Sheng OR, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. J Manag Inf Syst 1999;16(2):91-112. [doi: 10.1080/07421222.1999.11518247]
57. Asua J, Orruño E, Reviriego E, Gagnon MP. Healthcare professional acceptance of telemonitoring for chronic care patients in primary care. BMC Med Inform Decis Mak 2012 Nov 30;12:139 [FREE Full text] [doi: 10.1186/1472-6947-12-139] [Medline: 23194420]

58. Saigi-Rubió F, Jiménez-Zarco A, Torrent-Sellens J. Determinants of the intention to use telemedicine: evidence from primary care physicians. Int J Technol Assess Health Care 2016 Jan;32(1-2):29-36. [doi: 10.1017/S0266462316000015] [Medline: 27472158]
59. Heselmans A, Aertgeerts B, Donceel P, Geens S, Van de Velde S, Ramaekers D. Family physicians' perceptions and use of electronic clinical decision support during the first year of implementation. J Med Syst 2012 Dec;36(6):3677-3684. [doi: 10.1007/s10916-012-9841-3] [Medline: 22402980]
60. Pikkemaat M, Thulesius H, Milos Nymberg V. Swedish primary care physicians' intentions to use telemedicine: a survey using a new questionnaire - physician attitudes and intentions to use telemedicine (PAIT). Int J Gen Med 2021 Jul 15;14:3445-3455 [FREE Full text] [doi: 10.2147/IJGM.S319497] [Medline: 34295177]
61. Hung SY, Tsai JC, Chuang CC. Investigating primary health care nurses' intention to use information technology: an empirical study in Taiwan. Decis Support Syst 2014 Jan;57:331-342. [doi: 10.1016/j.dss.2013.09.016]
62. Pedersen KM, Andersen JS, Søndergaard J. General practice and primary health care in Denmark. J Am Board Fam Med 2012 Mar;25 Suppl 1:S34-S38 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110216] [Medline: 22403249]
63. SurveyXact. 2022. URL: https://www.surveyxact.com/ [accessed 2022-02-01]
64. Koder, tabeller, ydere. MedCom. 2021. URL: https://www.medcom.dk/opslag/koder-tabeller-ydere [accessed 2021-07-22]
65. Læge- og praksispopulationen, 1977-2020: Nøgletal fra medlemsregisteret. Praktiserende Lægers Organisation. 2020. URL: https://www.laeger.dk/sites/default/files/laege-_og_praksispopulationen_2020_004.pdf [accessed 2021-07-22]
66. Stata 17.0. StataCorp. 2022. URL: https://www.stata.com [accessed 2022-02-01]
67. Læger og tilmeldte patienter i almen praksis 2014-2020. Sundhedsdatastyrelsen. 2020. URL: https://sundhedsdatastyrelsen.dk/da/tal-og-analyser/analyser-og-rapporter/almen-praksis-og-kommuner/almen-praksis [accessed 2021-07-22]
68. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2017 Jun 7;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
69. Knekta E, Runyon C, Eddy S. One size doesn't fit all: using factor analysis to gather validity evidence when using surveys in your research. CBE Life Sci Educ 2019 Mar;18(1):rm1 [FREE Full text] [doi: 10.1187/cbe.18-04-0064] [Medline: 30821600]
70. Beran TN, Violato C. Structural equation modeling in medical research: a primer. BMC Res Notes 2010 Oct 22;3:267 [FREE Full text] [doi: 10.1186/1756-0500-3-267] [Medline: 20969789]
71. Tomarken AJ, Waller NG. Structural equation modeling: strengths, limitations, and misconceptions. Annu Rev Clin Psychol 2005;1:31-65. [doi: 10.1146/annurev.clinpsy.1.102803.144239] [Medline: 17716081]
72. Municipality Groups v1:2018. Statistics Denmark. 2018. URL: https://www.dst.dk/en/Statistik/dokumentation/nomenklaturer/kommunegrupper [accessed 2021-06-01]
73. Kopalle PK, Lehmann DR. Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. Organ Behav Hum Decis Process 1997 Jun;70(3):189-197. [doi: 10.1006/obhd.1997.2702]
74. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
75. Shi D, Lee T, Maydeu-Olivares A. Understanding the model size effect on SEM fit indices. Educ Psychol Meas 2019 Apr;79(2):310-334 [FREE Full text] [doi: 10.1177/0013164418783530] [Medline: 30911195]
76. Anderson JG. Clearing the way for physicians' use of clinical information systems. Commun ACM 1997 Aug;40(8):83-90. [doi: 10.1145/257874.257895]
77. 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]
78. COVID-19: Monitorering af aktivitet i sundhedsvæsenet: Beskrivelse af udviklingen i aktivitet i sundhedsvæsenet under COVID-19 epidemien - 10. rapport. Sundhedsstyrelsen. 2021 Jun. URL: https://www.sst.dk/-/media/Udgivelser/2020/Corona/Monitorering/10_-monitoreringsrapport.ashx [accessed 2021-06-04]
79. Barki H, Pare G, Sicotte C. Linking IT implementation and acceptance via the construct of psychological ownership of information technology. J Inf Technol 2008 Dec 01;23(4):269-280. [doi: 10.1057/jit.2008.12]
80. Saint-Lary O, Gautier S, Le Breton J, Gilberg S, Frappé P, Schuers M, et al. How GPs adapted their practices and organisations at the beginning of COVID-19 outbreak: a French national observational survey. BMJ Open 2020 Dec 02;10(12):e042119 [FREE Full text] [doi: 10.1136/bmjopen-2020-042119] [Medline: 33268433]
81. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]

Abbreviations

BI: behavioral intention
PEOU: perceived ease of use
PU: perceived usefulness
TAM: technology acceptance model

Edited by R Kukafka; submitted 03.03.22; peer-reviewed by M Pikkemaat, H Thulesius; comments to author 14.06.22; revised version received 07.07.22; accepted 18.07.22; published 30.08.22.

Please cite as:
Torp DC, Sandbæk A, Prætorius T
The Technology Acceptance of Video Consultations for Type 2 Diabetes Care in General Practice: Cross-sectional Survey of Danish General Practitioners
J Med Internet Res 2022;24(8):e37223
URL: https://www.jmir.org/2022/8/e37223
doi:10.2196/37223
PMID:36040765

Original Paper

Proactive Electronic Visits for Smoking Cessation and Chronic Obstructive Pulmonary Disease Screening in Primary Care: Randomized Controlled Trial of Feasibility, Acceptability, and Efficacy

Jennifer Dahne[1,2], PhD; Marty S Player[3], MD, MSCR; Charlie Strange[4], MD; Matthew J Carpenter[1,2,5], PhD; Dee W Ford[4], MD, MSCR; Kathryn King[6], MD, MHS; Sarah Miller[7], RN, PhD; Ryan Kruis[8], MSW; Elizabeth Hawes[1,2], BA; Johanna E Hidalgo[1,2], BA; Vanessa A Diaz[3], MD, MSCR

[1]Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, SC, United States
[2]Hollings Cancer Center, Medical University of South Carolina, Charleston, SC, United States
[3]Department of Family Medicine, Medical University of South Carolina, Charleston, SC, United States
[4]Division of Pulmonary and Critical Care Medicine, Medical University of South Carolina, Charleston, SC, United States
[5]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States
[6]Department of Pediatrics, Medical University of South Carolina, Charleston, SC, United States
[7]College of Nursing, Medical University of South Carolina, Charleston, SC, United States
[8]Center for Telehealth, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Jennifer Dahne, PhD
Department of Psychiatry and Behavioral Sciences
Medical University of South Carolina
MSC 955
86 Jonathan Lucas Street
Charleston, SC, 29425
United States
Phone: 1 843 876 2280
Email: dahne@musc.edu

Abstract

Background: Most smokers with chronic obstructive pulmonary disease (COPD) have not yet been diagnosed, a statistic that has remained unchanged for over two decades. A dual-focused telehealth intervention that promotes smoking cessation, while also facilitating COPD screening, could help address national priorities to improve the diagnosis, prevention, treatment, and management of COPD. The purpose of this study was to preliminarily evaluate an integrated asynchronous smoking cessation and COPD screening e-visit (electronic visit) that could be delivered proactively to adult smokers at risk for COPD, who are treated within primary care.

Objective: The aims of this study were (1) to examine e-visit feasibility and acceptability, particularly as compared to in-lab diagnostic pulmonary function testing (PFT), and (2) to examine the efficacy of smoking cessation e-visits relative to treatment as usual (TAU), all within primary care.

Methods: In a randomized clinical trial, 125 primary care patients who smoke were randomized 2:1 to receive either proactive e-visits or TAU. Participants randomized to the e-visit condition were screened for COPD symptoms via the COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk (CAPTURE). Those with scores ≥2 were invited to complete both home spirometry and in-lab PFTs, in addition to two smoking cessation e-visits. Smoking cessation e-visits assessed smoking history and motivation to quit and included completion of an algorithm to determine the best Food and Drug Administration–approved cessation medication to prescribe. Primary outcomes included measures related to (1) e-visit acceptability, feasibility, and treatment metrics; (2) smoking cessation outcomes (cessation medication use, 24-hour quit attempts, smoking reduction ≥50%, self-reported abstinence, and biochemically confirmed abstinence); and (3) COPD screening outcomes.

Results: Of 85 participants assigned to the e-visits, 64 (75.3%) were invited to complete home spirometry and in-lab PFTs based on CAPTURE. Among those eligible for spirometry, 76.6% (49/64) completed home spirometry, and 35.9% (23/64) completed in-lab PFTs. At 1 month, all cessation outcomes favored the e-visit, with a significant effect for cessation medication use (odds ratio [OR]=3.22). At 3 months, all cessation outcomes except for 24-hour quit attempts favored the e-visit, with significant effects for cessation medication use (OR=3.96) and smoking reduction (OR=3.09).

Conclusions: A proactive, asynchronous e-visit for smoking cessation and COPD screening may offer a feasible, efficacious approach for broad interventions within primary care.

Trial Registration: ClinicalTrials.gov NCT04155073; https://clinicaltrials.gov/ct2/show/NCT04155073

(J Med Internet Res 2022;24(8):e38663) doi:10.2196/38663

KEYWORDS

electronic visits; e-visit; COPD; chronic obstructive pulmonary disease; smoking cessation; telehealth; electronic health record; patient portal; EHR; feasibility; efficacy; intervention; screening; diagnosis; prevention; treatment; management; acceptability; pulmonary; function

Introduction

Cigarette smoking remains the leading cause of preventable death globally and is responsible for more than 480,000 deaths each year in the United States [1]. A total of 21% percent of tobacco-related deaths are caused by chronic obstructive pulmonary disease (COPD), a progressive inflammatory lung disease that causes airflow obstruction and breathing-related problems [1,2]. Between 45% and 72% of smokers with COPD have not yet been diagnosed [3,4], a statistic that has remained largely unchanged for over two decades [4]. As noted in a recent viewpoint article by Yawn and Martinez [5], “COPD screening must develop better, more symptom-based tools and appropriate follow-up support.” A dual-focused intervention that simultaneously promotes smoking cessation, while also facilitating COPD screening, could address national priorities to improve the diagnosis, prevention, treatment, and management of COPD [6].

The vast majority (~70%) of adult smokers visit a primary care provider at least once per year, making primary care an ideal environment within which to identify smokers, provide evidence-based smoking cessation treatment, and screen for COPD [7-9]. Within the primary care setting, prior studies demonstrate up to a fourfold increase in COPD diagnosis when using screening tools to identify respiratory symptoms [10-12]. Although not COPD-specific, our team previously developed an asynchronous e-visit (electronic visit) to be delivered within the primary care environment to patients identified as smokers via the electronic health record (EHR) [13]. This e-visit was developed based on best practice guidelines [7] for smoking cessation treatment within primary care—the 5 A’s (ask, advise, assess, assist, and arrange). Results from an initial evaluation of the asynchronous smoking cessation e-visit as compared to treatment as usual (TAU) within primary care indicated high feasibility and acceptability with cessation outcomes that favored the e-visit condition at both 1 (odds ratios [ORs] 2.10-5.39) and 3 months (ORs 1.31-4.67) [13].

The purpose of this study was to preliminarily evaluate an integrated asynchronous smoking cessation and COPD screening e-visit that could be delivered proactively to adult smokers at risk for COPD, treated within primary care. Prior studies have evaluated the feasibility, acceptability, and validity of remote home spirometry and have found high test-retest reliability when compared to in-clinic assessments [14,15], high adherence rates [16], and high patient satisfaction [15,16]. As such, we opted to leverage our existing asynchronous smoking cessation e-visit platform and add to it remote, telehealth-facilitated COPD screening and completion of remote home spirometry for those eligible. The aims of this study were (1) to examine e-visit (for smoking cessation and remote home spirometry) feasibility and acceptability, particularly as compared to in-lab diagnostic pulmonary function testing (PFTs), and (2) to examine the efficacy of smoking cessation e-visits relative to TAU, all within primary care.

Methods

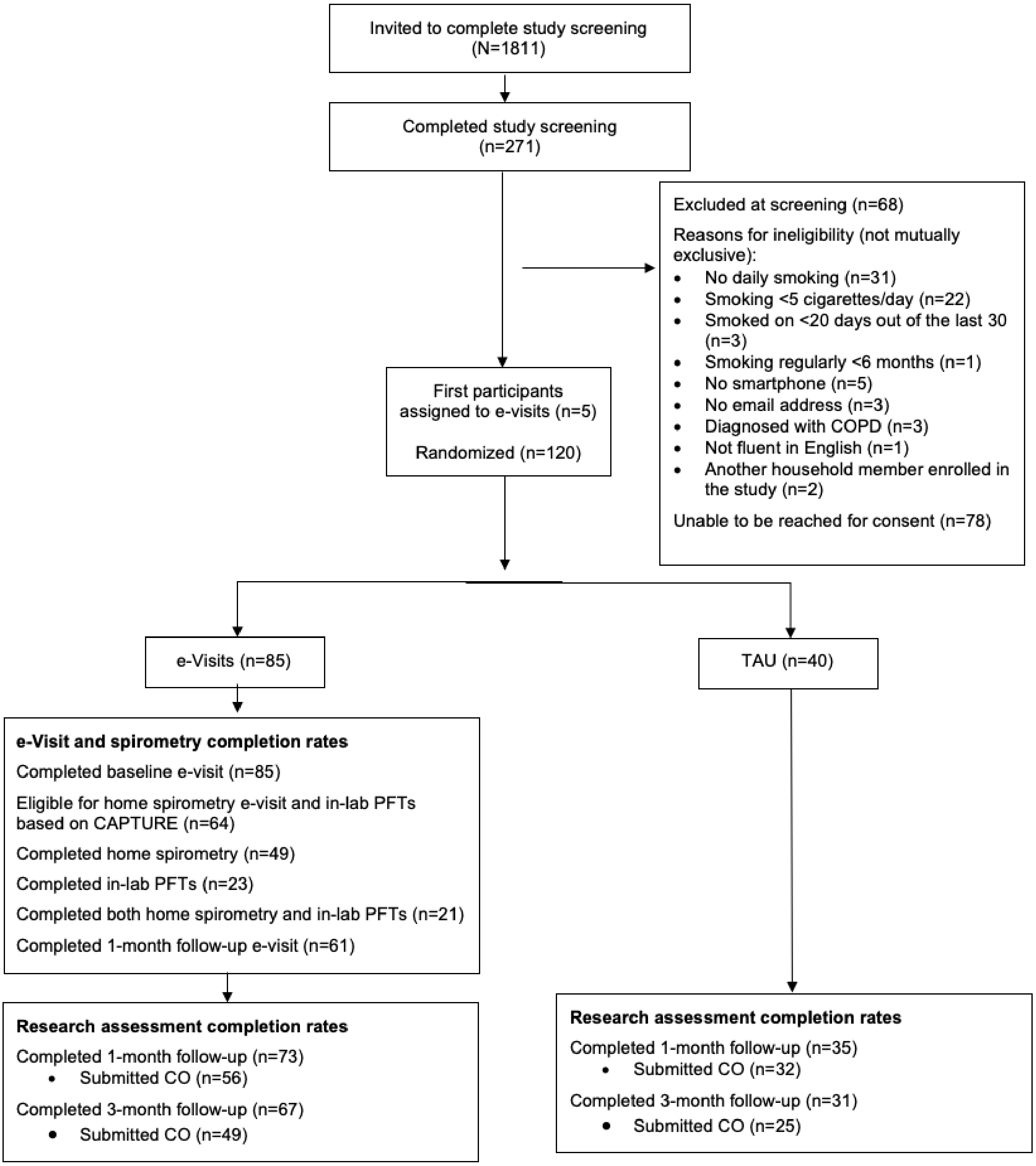
Ethics Approval

All study procedures were approved by the Medical University of South Carolina (MUSC) institutional review board (PRO00086016), and the trial was preregistered with ClinicalTrials.gov (NCT04155073). Eligible patients were scheduled to complete informed consent remotely with a member of the study team. Consent was completed either electronically or via mail, in both cases paired with a discussion with a member of the research team.

Participants

Participants were recruited from 13 primary care practices affiliated with MUSC between December 2019 and January 2021. Within our systemwide EHR, Epic, a study recruitment report was generated for all patients meeting the following criteria: (1) aged >40 years; (2) seen at an MUSC primary care practice in the last year; (3) current smoker; (4) no previous diagnosis of COPD (defined as International Classification of diseases, 10th revision codes J44.9, J44.1, J44.0, J43.9, or Z87.9) associated with any prior visit; and (5) access to MyChart, Epic’s patient portal. Via MyChart, 1811 patients meeting initial eligibility criteria were sent an invitation and link to study screening. Study invitations included introductory text highlighting the importance of quitting smoking and then continued with an invitation to participate in a research study to help change smoking behavior. All invitations noted that the study had been discussed with the patient’s primary care provider, who supported the study invitation. Patients were deemed eligible for the study if during the initial screening they met the following additional criteria: (1) current cigarette smoking, defined as smoking 5 cigarettes per day for at least 20 out of the preceding 30 days, for at least the last 6 months; (2) possess a valid email address, checked daily; (3) owner of an iOS or Android-compatible smartphone; and (4) fluent in English (study e-visits were only available in English, thus English fluency was required). In total, 271 patients completed study screening (ie, 15% of those invited), and 203 were deemed eligible following screening (Figure 1).

Figure 1. CONSORT flow diagram. CAPTURE: COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk; CO: carbon monoxide; COPD: chronic obstructive pulmonary disease; PFT: peak expiratory flow; TAU: treatment as usual.



Study Procedures

Upon consent, enrolled participants completed baseline assessments via REDCap (version 11.2.1; Vanderbilt University) and then were randomized 2:1 to receive either the e-visits or TAU. The first 5 enrolled participants were all assigned to the e-visit condition to ensure that both smoking cessation and COPD e-visit components were functioning properly. All study participants were invited to complete follow-up research assessments at 1 and 3 months following baseline. To biochemically verify smoking status, participants were asked to submit an expired air carbon monoxide (CO) sample remotely via a smartphone-enabled CO monitor (iCO Smokerlyzer) at both follow-ups. iCO monitors were mailed to all enrolled participants following completion of the baseline visit, and mailings included information regarding how to submit CO readings. Participants were compensated up to US $250 total for all study procedures. Participants randomized to the e-visit condition were not separately compensated for completion of the e-visits or for completion of in-lab PFTs.

Interventions

Smoking Cessation and COPD e-Visits

Participants in this condition were automatically linked to initiate an asynchronous e-visit via MyChart. Smoking cessation components of the e-visit were similar to our team's prior work [13] but modified to include an additional focus on COPD screening. e-Visit functionality was similar to that of an electronic questionnaire, with participants answering 1 question per screen and then advancing to the next screen. The initial baseline e-visit gathered information about smoking history and motivation to quit, followed by an algorithm to determine the best Food and Drug Administration (FDA)-approved cessation medication (ie, nicotine replacement therapy [NRT], varenicline, and bupropion) to prescribe. This algorithm was based on prior research [17,18] and evidence-based guidelines [7], using branching logic to prioritize the most efficacious medications (ie, varenicline and combination NRT), while tailoring recommendations based on contraindications and participant preference. A medication recommendation was then displayed to the participant, with a personalized rationale, to which the participant could agree or request a different treatment. e-Visit results were then sent to the provider's electronic in-basket, who reviewed the e-visit, responded to the patient via MyChart with instructions, and e-prescribed (if indicated) medication. All medications were prescribed on label by 3 study physicians and were mailed to the patient at no cost. Responses from providers to participants also included information on the state tobacco Quitline, which participants could contact for additional behavioral support. Participants were subsequently invited to complete a follow-up smoking cessation e-visit 1 month following completion of the baseline session, consistent with the 5th A in the 5 A's guideline to arrange follow-up [7], at which time participants could request a refill of the medication prescribed at baseline, if needed, or could request a new smoking cessation medication. Results were sent to providers and reviewed in the same manner as the baseline e-visit.

In addition to smoking cessation content, the baseline e-visit included completion of the 5-item COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk (CAPTURE) [19]. CAPTURE assesses the presence or absence of COPD symptoms, risk exposures, and recent history of acute respiratory illness. Responses are summed and scores of 2 or higher suggest a need for diagnostic assessment [5]. Thus, e-visit participants with a CAPTURE score ≥2 were subsequently invited to complete both remote spirometry and in-lab PFTs. For remote spirometry testing, participants were mailed a home spirometer (Vitalograph asma-1) and were sent a link to complete an additional e-visit in which they recorded themselves using the home spirometer. At the beginning of the home spirometry e-visit, participants reviewed an educational video, developed by our team, demonstrating how to use the device, while video recording themselves. Participants submitted 3 breath samples with accompanying videos for evaluation and were asked to enter peak expiratory flow (PEF) and forced expiratory volume in one second (FEV1) values into text boxes within the e-visit. Completed home spirometry e-visits and videos were reviewed by a study physician, who coded them for effort (ie, acceptable, unacceptable, and unable to determine), technique (coded similarly), and whether the participant correctly recorded PEF and FEV1 values from the home spirometer (ie, yes, no, or participant did not show values to the camera). For each participant, the percent predicted PEF was calculated from the highest acceptable PEF recorded, factoring in age, gender, and ethnicity, based on equations with standard values [20,21]. Participants with predicted PEF ≤80% were considered abnormal. Similarly, the highest acceptable FEV1 measurement was categorized as normal (≥80%), moderate (≥50%-<80% predicted), severe (≥30%-<50% predicted), or very severe (<30% predicted). After review, the physician sent a message to the patient via the EHR portal with results (normal or abnormal) along with a recommendation to complete in-lab PFTs previously ordered. These messages also included encouragement related to quitting smoking (eg, "we still recommend that you attempt to quit smoking. Quitting smoking now will help to prevent any further lung damage as well as reduce your risk of heart disease and cancer linked to tobacco use. We are here to continue helping you in those efforts"). No intervention for COPD was provided as part of this study.

All participants eligible for home spirometry testing were also referred for PFTs, regardless of the home testing results, to examine the comparative feasibility of home versus in-lab testing. To further remove barriers to completion of in-lab PFTs, all PFT costs were paid for by the study. Once completed, in-lab PFT results were communicated to the patient with recommendations for follow-up with their primary provider. Chart reviews were completed at 3 months following study enrollment for all study participants to determine whether PFTs were completed and whether the participant was subsequently diagnosed with COPD. PFT appointments for these participants were scheduled per usual practice (ie, a referral was placed by study coordinators, and central scheduling contacted participants to schedule testing).

Treatment as Usual (TAU)

TAU was designed to mimic existing standard cessation practices. Research staff provided participants in this condition with information on the state Quitline and a recommendation to contact their primary care provider to schedule a medical visit to discuss quitting smoking. Chart review was also completed at 3 months for these participants to determine whether they completed in-lab PFTs and whether they were diagnosed with COPD.

Measures

All participants at baseline completed a general assessment of demographics and health history. Primary outcomes for this trial include measures related to (1) e-visit acceptability, feasibility, and treatment metrics; (2) smoking cessation outcomes; and (3) COPD screening outcomes. Unless otherwise noted, participants self-input responses to study outcomes assessments in REDCap.

e-Visit Acceptability, Feasibility, and Treatment Metrics

To assess participant perception of the e-visits, during the 1-month research assessment, e-visit participants responded to the following items: (1) I found the e-visit easy to use; (2) I would use an e-visit again in the future; (3) during my e-visit, I felt I could trust my provider with my medical care; (4) I would recommend e-visits to other people; (5) It was as easy for me to state concerns through the e-visit as it would be in an in-person visit; (6) the e-visit was as good as an in-person visit with my doctor; and (7) I have experienced benefits from the e-visit. Response options ranged from strongly disagree to strongly agree.

To examine comparative feasibility of remote versus in-lab PFTs, completion rates for each were captured and compared. Feasibility of remote home spirometry was further assessed through clinician ratings of spirometry effort or technique, as described above. Feasibility of the smoking cessation component of the e-visit was captured via EHR chart reviews, as follows: (1) whether the patient opted for the medication recommended by the e-visit, (2) whether the physician prescribed the medication recommended by the e-visit, (3) whether the participant completed the 1-month follow-up e-visit, and (4) time to complete the 1-month follow-up e-visit.

Smoking Cessation Outcomes

All participants at baseline were queried for the number of cigarettes smoked per smoking day, incidence of quit attempts within the last year, and motivation or confidence to quit (0-10 on the visual analogue scale [22]) in the next month. During the 1- and 3-month follow-ups, all participants self-reported the following: (1) number of cigarettes smoked per day over the last 7 days, (2) incidence of 24-hour quit attempts since the prior assessment, and (3) use of an FDA-approved smoking cessation medication since the last assessment. Past-week smoking data allowed for a computed outcome to assess if participants reduced their smoking by at least 50% since baseline. Participants who reported smoking zero cigarettes over the last 7 days were coded as having self-reported 7-day point prevalence abstinence. Self-reported abstinence was biochemically confirmed via CO, using a cutoff point <6 ppm to define abstinence [23].

COPD Outcomes

COPD diagnostic status was ascertained for all participants via chart review at 3 months following study enrollment. Additionally, at baseline, all participants completed CAPTURE [19].

Statistical Analysis Plan

Chi-square and ANOVA analyses were used to determine baseline group differences in participant demographics as well as retention rates over time. Descriptive statistics were used to examine e-visit (for spirometry and smoking cessation) acceptability, feasibility, and treatment metrics. Binary logistic regressions were used to examine differences in cessation outcomes across treatment group, at both 1- and 3-month time points. For cessation outcomes, an intent-to-treat approach was used such that those who did not complete the assessment were coded as not having modified smoking [24].

Results

Participant Characteristics

In total, 125 participants were enrolled in the trial (e-visit=85 and TAU=40). The first 5 enrolled participants were all assigned to the e-visit condition, and the remaining 120 were randomized 2:1 to either e-visit or TAU. The first 5 enrolled participants did not significantly differ from those randomized, either in baseline characteristics or follow-up outcomes. There were no significant between-group differences in demographics or smoking history at baseline (Table 1), though participants in the TAU condition reported significantly greater COPD symptoms on CAPTURE ($F_{1,123}$=8.11, P=.005).

Study retention was generally high across both 1-month (86.4%) and 3-month (78.4%) follow-ups, with no significant differences in retention between treatment groups. Regarding demographic differences between those who completed follow-up assessments and those who did not, White participants were significantly more likely to complete the 1-month follow-up assessment compared to non-White participants (91% completion versus 75%; $\chi^2_{1,125}$=5.59; P=.02). There were no other significant demographic differences between those who completed follow-up assessments and those who did not, at either 1 or 3 months. Among 1-month respondents, 81.5% (88/108) also provided CO. Among 3-month respondents, 73.5% (72/98) also provided CO.

Table 1. Participant demographics.

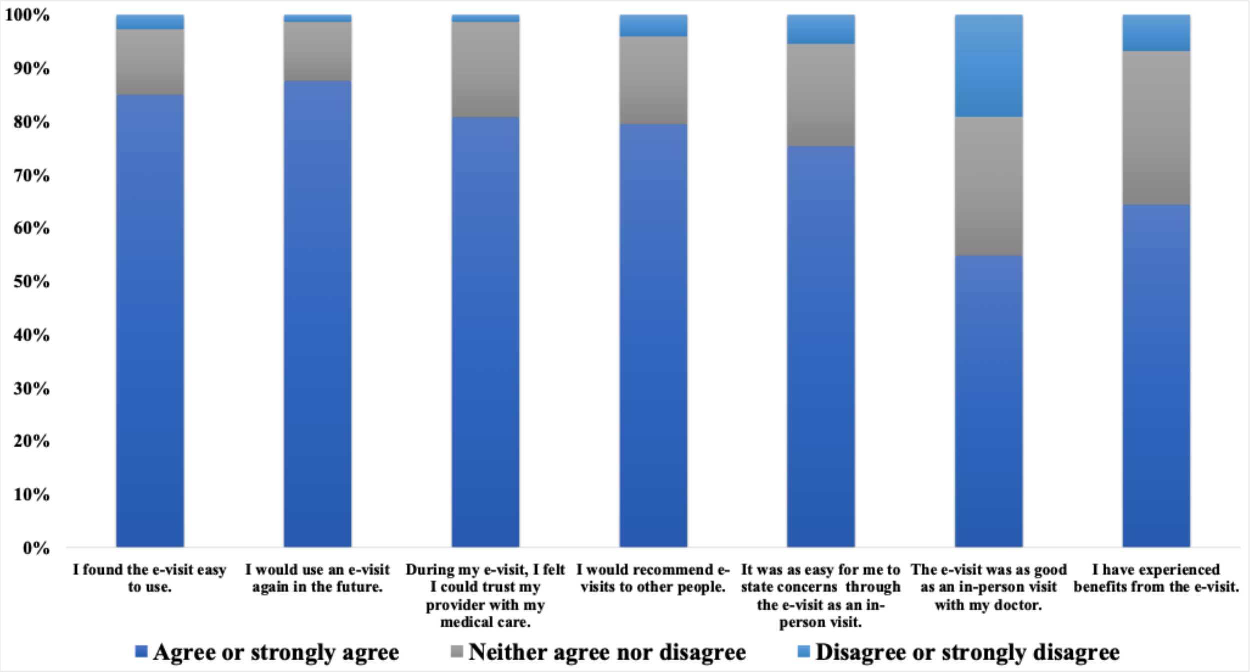
Characteristics	Full sample (N=125)	e-Visit (n=85)	Treatment as usual (n=40)
Age (years), mean (SD)	53.42 (9.40)	54.12 (9.75)	51.93 (8.52)
Age (years), n (%)			
40-64	109 (87.2)	71 (83.5)	38 (95)
>65	16 (12.8)	14 (16.5)	2 (5)
Sex (female), n (%)	75 (59.2)	52 (61.2)	22 (55)
Race, n (%)			
White	89 (71.2)	61 (71.8)	28 (70)
Black	28 (22.4)	18 (21.2)	10 (25)
Other	8 (6.4)	6 (7.2)	2 (5)
Ethnicity (Hispanic/Latinx), n (%)	4 (3.2)	2 (2.4)	2 (5)
Education, n (%)			
<High school diploma	44 (35.2)	27 (31.8)	17 (42.5)
>High school diploma	81 (64.8)	58 (68.3)	23 (57.5)
Annual household income, n (%)			
<US $50K	65 (52)	43 (50.6)	22 (55)
>US $50K	56 (44.8)	39 (45.8)	17 (42.5)
Not sure or refused to answer	4 (3.2)	3 (3.6)	1 (2.5)
Health insurance status, n (%)			
Total number of participants insured	112 (89.6)	79 (92.9)	33 (82.5)
Medicaid	15 (12)	10 (11.8)	5 (12.5)
Medicare	23 (18.4)	17 (20.0)	6 (15)
Employer-provided insurance	53 (42.4)	37 (43.5)	16 (40)
Other	21 (16.8)	15 (17.6)	6 (15)
Baseline cigarettes per day, mean (SD)	18.43 (9.79)	18.09 (8.98)	19.15 (11.43)
Quit attempt in the past year (yes), n (%)	73 (58.4)	48 (56.5)	25 (62.5)
Motivation to quit in the next month, mean (SD)	7.53 (2.50)	7.36 (2.56)	7.88 (2.34)
Confidence in quitting in the next month, mean (SD)	5.76 (2.88)	5.64 (2.87)	6.03 (2.92)
Baseline CAPTURE[a], mean (SD)	3.13 (1.72)	2.84 (1.72)	3.75 (1.56)

[a] CAPTURE: COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk.

e-Visit Feasibility, Acceptability, and Uptake

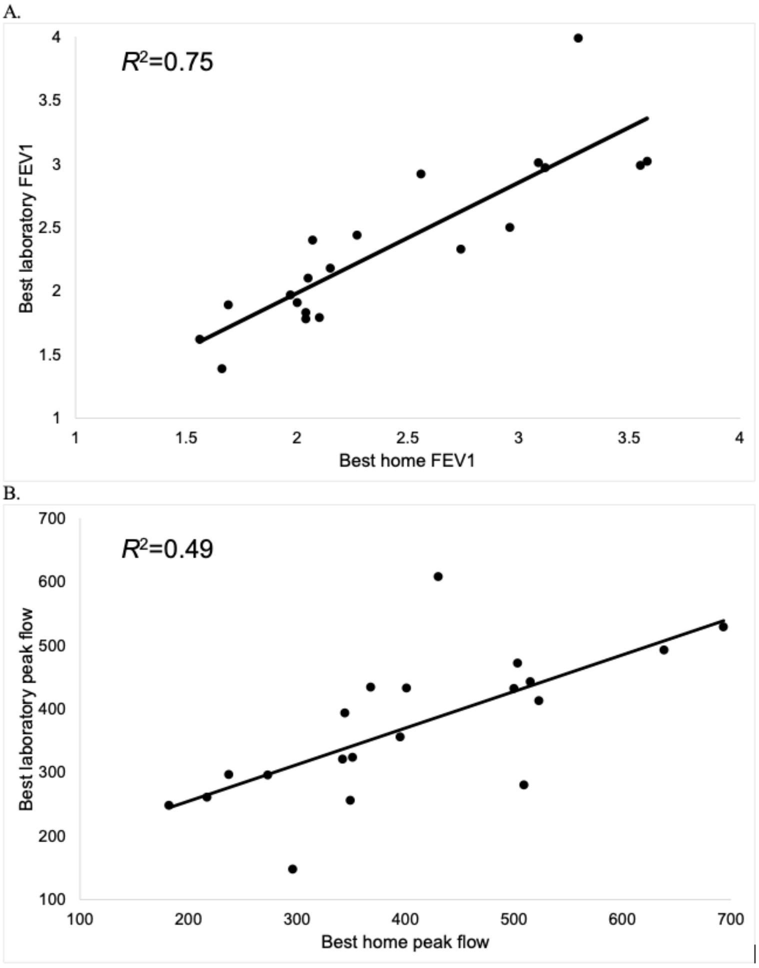
Participant feedback following completion of the baseline e-visit was generally positive (Figure 2). Of the 85 participants assigned to the e-visit condition, 64 (75.3%) were invited to complete home spirometry and in-lab PFTs because of a CAPTURE score of 2 or higher. Mean CAPTURE score among those eligible for spirometry was 3.7 (SD 1.4). Among those eligible for spirometry, 76.6% (49/64) completed home spirometry. Most of those (37/49; 75.5%) who completed home spirometry had acceptable effort; 79.6% (39/49) had acceptable technique, and 87.8% (43/49) correctly recorded values on at least 1 video. Two-thirds (33/49; 67.3%) of participants who completed home spirometry had at least 1 video with acceptable effort and technique.

Figure 2. e-Visit feedback.



Compared to completion rates for home spirometry, fewer participants in the e-visit condition completed in-lab PFTs (23/64, 35.9% of those eligible). Twenty-one participants completed both home and in-lab PFTs. Among these participants, concordance between home and in-lab PFTs was higher for FEV1 (R^2=0.75) compared to PEF (R^2=0.49; Figure 3). Among the sample completing both home spirometry and in-person PFTs, in-person spirometric diagnoses included normal spirometry (4/21, 19%), Global Initiative for Obstructive Lung Disease undifferentiated obstruction with FEV1 <80% but FEV1/forced vital capacity >0.7 (2/21, 9.5%), probable or confirmed restriction (5/21, 23.8%), mild obstruction (5/21, 23.8%) and moderate obstruction (5/21, 23.8%). No participant was found to have severe obstruction. Using a cutoff of home spirometry PEF <80% predicted, 5 participants who completed both home spirometry and in-person PFTs were considered to have abnormal home spirometry readings. Among these participants, in-person PFTs confirmed restriction (1/5, 20%), mild obstruction (1/5, 20%), moderate obstruction (2/5, 40%), and normal spirometry (1/5, 20%). In the TAU group, only 1 participant completed in-lab PFTs and that participant was subsequently diagnosed with COPD.

Figure 3. Home vs. In-Lab FEV1 and Peak Flow Among Participants Who Completed Both.

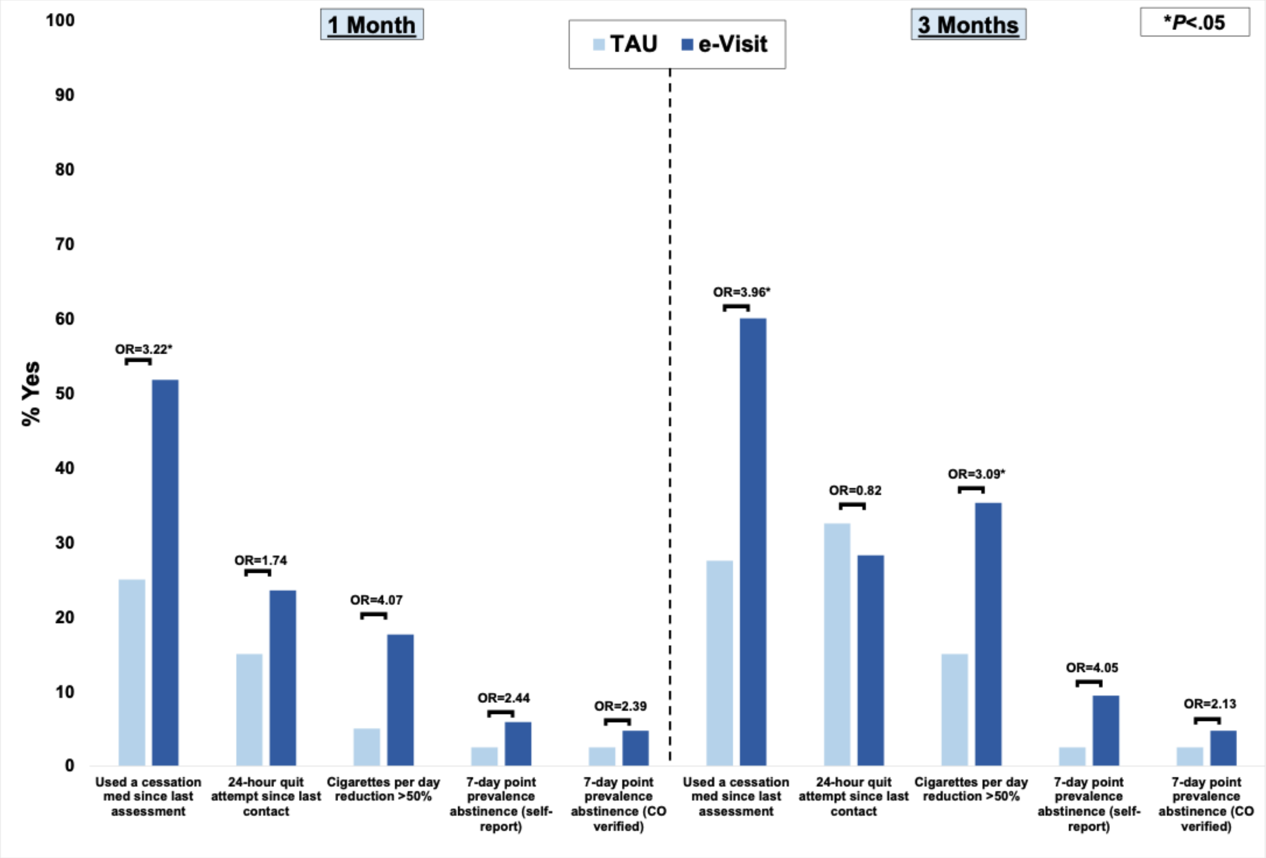


Smoking Cessation, Treatment, and Outcomes

The most common treatment recommendation as a result of the baseline e-visit was varenicline (65/85, 76.5%), followed by NRT patch (8/85, 9.4%), combination NRT (5/85, 5.9%), and NRT lozenge (3/85, 3.53%). Participants (73/85, 86%) and providers (72/85, 85%) agreed with recommendations from the medication algorithm. Three quarters (64/85, 75.3%) of e-visit participants completed the 1-month e-visit, on average within 2.7 (SD 6.9) days after invitation. During this follow-up e-visit, participants most often requested either a prescription for varenicline (20/64, 31.3%), combination NRT (11/64, 17.2%), NRT inhaler (10/64, 15.6%), or NRT gum (6/64, 9.4%), and providers abided by these preferences (63/64, 98.4%).

In general, smoking cessation outcomes favored the e-visit condition at both 1 and 3 months (Figure 4). At 1 month, all cessation outcomes favored the e-visit condition (ORs 1.6-4.1). At 3 months, all cessation outcomes except for 24-hour quit attempts favored the e-visit condition (ORs 1.1-5.8). Regarding significant effects, as compared to TAU, e-visit participants were 3.2 times more likely to have used a cessation medication at 1 month (95% CI 1.4-7.4; *P*=.006), and 4.0 times more likely to have used a cessation medication at 3 months (95% CI 1.7-9.0; *P*<.001). At 3 months, e-visit participants were 3.1 times more likely to have reduced their cigarettes per day by at least 50% (95% CI 1.2-8.2; *P*=.02), with a similar trend toward significance at 1 month (OR 4.1, 95% CI 0.9-18.8; *P*=.07).

Figure 4. Smoking cessation outcomes. CO: carbon monoxide; OR: odds ratio; TAU: treatment as usual.



Discussion

Study results preliminarily indicate feasibility, acceptability, and efficacy of a proactive, asynchronous e-visit for smoking cessation and COPD screening. Metrics of feasibility and acceptability were strong, with 88% (75/85) of participants indicating that they would use an e-visit again in the future. However, only a small majority (47/85, 55%) of e-visit participants reported preferring the e-visit to an in-person visit. Thus, although the e-visit may offer a scalable, feasible method to extend the reach of cessation treatment, it may not be preferred for all patients. Future research should examine which subgroups of patients may be most amenable to receiving cessation treatment via telehealth platforms such as e-visits.

Our study also confirms the feasibility of home spirometry, with promising compliance rates for submission of viable tests. This is consistent with prior research among other populations with obstructive lung diseases such as cystic fibrosis [25,26] and asthma [27]. Thus, home spirometry, completed via an e-visit and reviewed by a provider, may be a useful tool for COPD screening among high-risk smokers. A large body of literature has focused on both the potential utility and drawbacks of population-based screening for COPD, particularly among asymptomatic patients. Current US Preventive Services Task Force guidelines recommend against screening asymptomatic adults for COPD, citing lack of supportive data [28]. However, screening of individuals who self-report unaddressed respiratory symptoms, as implemented in this trial, can increase COPD diagnoses and facilitate treatment initiation [10-12]. Although a tool such as CAPTURE has broad reach and may help to identify those who are symptomatic, pairing CAPTURE with home-based spirometry could help further identify the subset of patients who should be strongly encouraged to complete diagnostic PFTs.

This study was not designed specifically to examine the validity of home spirometry; nevertheless, comparing home versus in-lab results for the small subset of participants who completed both suggests there are opportunities to improve validity. To maximize scalability of the e-visits, we opted to minimize the amount of training provided to participants. However, prior home spirometry trials have had success with incorporating synchronous coaching via video calls [29]. This approach may help improve the validity of home spirometry among adult smokers at risk for COPD, though it would limit scalability. In the future, it will be important to determine the appropriate amount of training needed for participants to submit valid samples and how best to embed this training in primary care.

To our knowledge, this is the first study to evaluate the CAPTURE, which was developed to be used in a broad population of primary care patients, in a population of smokers. We were surprised to see that 82.4% (103/125) of smokers in our full sample had a score of 2 or higher at baseline. Because smokers typically develop COPD after 10 pack years at a prevalence <20% [30], the CAPTURE instrument should be reevaluated among active smokers, as sensitivity may be too high. This is likely due to scoring 1 point for living or working in a place with smoke or secondhand smoke.

Our results further substantiate the potential of e-visits to promote smoking cessation. Results generally echoed those of our prior trial [13], supporting the efficacy of the e-visit approach. Whereas our prior trial did not provide free medications, the current study provided prescribed medications free of charge. Comparing intent-to-treat results across studies, the provision of free medication appears to have slightly increased the use of cessation medications (at 1 month: 44.1% vs 51.8%; at 3 months: 41.2% vs 60%). Thus, where possible, pairing the proactive e-visit with free medication may increase evidence-based cessation treatment uptake.

Results of this study should be interpreted with limitations in mind. The trial was largely conducted in the midst of the COVID-19 pandemic, which may have impacted rates of trial enrollment, engagement with the e-visits, and completion rates for in-lab PFTs. However, it is important to note that at MUSC, pulmonary function testing continued throughout the pandemic, thus all participants eligible for in-lab PFTs had the option to complete them. Future evaluation of proactive e-visits for COPD screening and smoking cessation outside the context of the COVID-19 pandemic will be important to determine whether e-visit acceptability and feasibility change as a function of the pandemic waning. Regarding generalizability, the Vitalograph asma-1 was used for remote home spirometry completion. Feasibility and validity results may not generalize to other remote monitors. Moreover, study inclusion criteria, including smartphone ownership and regular email use, may limit generalizability of results. Proactive study invitations were sent via MyChart, which may also decrease results' generalizability. However, this decision was made because the study e-visits were delivered via the MyChart patient portal. Given the preliminary nature of this trial and focus on feasibility and acceptability, resources to support deployment of both remote spirometry and in-lab PFTs were not comprehensively assessed. However, future cost-effectiveness analyses could help determine whether potential benefits of the e-visit approach are cost-effective or cost-saving at the health care system level. Finally, the completed e-visits were reviewed by study physicians. Implementation of the e-visits within routine clinical practice and with non–study-affiliated providers remains unclear but is an important avenue for future research.

In sum, a proactive, asynchronous e-visit for smoking cessation treatment and COPD screening may offer a feasible, efficacious approach for broad intervention within primary care. If validity of home spirometry can be improved over time, the e-visit platform may help not only promote uptake of evidence-based smoking cessation treatment but also provide an early screening mechanism to identify smokers with COPD or other important lung diseases.

Acknowledgments

JD, MSP, CS, MJC, DWF, KK, SM, and VAD collectively conceived the research questions, which form the focus of this manuscript. JD directed the trial and led data analysis and manuscript preparation. RK, EH, and JEH assisted with conducting the trial. All authors contributed to manuscript preparation, including providing feedback and edits, and have approved this version of the manuscript. Funding for this research was provided by the Health Resources and Services Administration (U66 RH31458), the South Carolina Telehealth Alliance, the National Institute on Drug Abuse (K23 DA045766), and the National Cancer Institute (R21 CA241842). The authors would like to thank the Biomedical Informatics Center at MUSC, including Buck Rogers and Paul Powers, who led technical development of the e-visit.

Conflicts of Interest

JD is co-owner of Behavioral Activation Tech LLC, a small business that develops and evaluates mobile app–based treatments for depression and co-occurring disorders. MJC has received consulting honoraria from Pfizer and from Frutarom Inc. CS has grants related to chronic obstructive pulmonary disease (COPD) paid to the Medical University of South Carolina from AstraZeneca, CSA Medical, National Institutes of Health, Nuvaira, and Takeda within the past 3 years. He has consulted for Bronchus, GlaxoSmithKline, and Pulmanage, for COPD. The funding sources had no role in study design, data collection, data analysis, data interpretation, writing this report, or in the decision to submit this paper for publication.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 348 KB - jmir_v24i8e38663_app1.pdf]

References

1. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014.
2. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. N Engl J Med 2013 Jan 24;368(4):341-350. [doi: 10.1056/NEJMsa1211128] [Medline: 23343063]

3. Pleasants RA, Heidari K, Wheaton AG, Ohar JA, Strange C, Croft JB, et al. Targeting persons with or at high risk for chronic obstructive pulmonary disease by state-based surveillance. COPD 2015;12(6):680-689 [FREE Full text] [doi: 10.3109/15412555.2015.1043424] [Medline: 26367193]
4. Martinez CH, Mannino DM, Jaimes FA, Curtis JL, Han MK, Hansel NN, et al. Undiagnosed obstructive lung disease in the united states. Associated factors and long-term mortality. Ann Am Thorac Soc 2015 Dec;12(12):1788-1795 [FREE Full text] [doi: 10.1513/AnnalsATS.201506-388OC] [Medline: 26524488]
5. Leidy NK, Martinez FJ, Malley KG, Mannino DM, Han MK, Bacci ED, et al. Can CAPTURE be used to identify undiagnosed patients with mild-to-moderate COPD likely to benefit from treatment? Int J Chron Obstruct Pulmon Dis 2018;13:1901-1912 [FREE Full text] [doi: 10.2147/COPD.S152226] [Medline: 29942123]
6. Kiley JP, Gibbons GH. COPD national action plan: addressing a public health need together. Chest 2017 Oct;152(4):698-699. [doi: 10.1016/j.chest.2017.08.1155] [Medline: 28991541]
7. Fiore M, Jaen C, Baker T, Bailey W, Benowitz N. Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: US Department of Health and Human Services, Public Health Service; May 2008.
8. Jamal A, Dube SR, Malarcher AM, Shaw L, Engstrom MC, Centers for Disease Control Prevention (CDC). Tobacco use screening and counseling during physician office visits among adults--National Ambulatory Medical Care Survey and National Health Interview Survey, United States, 2005-2009. MMWR Suppl 2012 Jun 15;61(2):38-45. [Medline: 22695462]
9. Juster HR, Ortega-Peluso CA, Brown EM, Hayes KA, Sneegas K, Gopez G, et al. A media campaign to increase health care provider assistance for patients who smoke cigarettes. Prev Chronic Dis 2019 Oct 17;16:E143 [FREE Full text] [doi: 10.5888/pcd16.180613] [Medline: 31625868]
10. Haroon S, Adab P, Riley RD, Fitzmaurice D, Jordan RE. Predicting risk of undiagnosed COPD: development and validation of the TargetCOPD score. Eur Respir J 2017 Jun 22;49(6):1602191 [FREE Full text] [doi: 10.1183/13993003.02191-2016] [Medline: 28642308]
11. Jordan RE, Adab P, Sitch A, Enocson A, Blissett D, Jowett S, et al. Targeted case finding for chronic obstructive pulmonary disease versus routine practice in primary care (TargetCOPD): a cluster-randomised controlled trial. Lancet Respir Med 2016 Sep;4(9):720-730. [doi: 10.1016/S2213-2600(16)30149-7] [Medline: 27444687]
12. Yawn BP, Martinez FJ. POINT: can screening for COPD improve outcomes? Yes. Chest 2020 Jan;157(1):7-9. [doi: 10.1016/j.chest.2019.05.034] [Medline: 31916966]
13. Dahne J, Player M, Carpenter MJ, Ford DW, Diaz VA. Evaluation of a proactive smoking cessation electronic visit to extend the reach of evidence-based cessation treatment via primary care. Telemed J E Health 2021 Mar;27(3):347-354 [FREE Full text] [doi: 10.1089/tmj.2020.0167] [Medline: 33085578]
14. Huang C, Izmailova ES, Jackson N, Ellis R, Bhatia G, Ruddy M, et al. Remote FEV1 monitoring in asthma patients: a pilot study. Clin Transl Sci 2021 Mar;14(2):529-535 [FREE Full text] [doi: 10.1111/cts.12901] [Medline: 33048470]
15. Moor CC, Wapenaar M, Miedema JR, Geelhoed JJM, Chandoesing PP, Wijsenbeek MS. A home monitoring program including real-time wireless home spirometry in idiopathic pulmonary fibrosis: a pilot study on experiences and barriers. Respir Res 2018 May 29;19(1):105 [FREE Full text] [doi: 10.1186/s12931-018-0810-3] [Medline: 29843728]
16. Moor CC, Mostard RLM, Grutters JC, Bresser P, Aerts JGJV, Chavannes NH, et al. Home monitoring in patients with idiopathic pulmonary fibrosis: a randomized controlled trial. Am J Respir Crit Care Med 2020 Aug 01;202(3):393-401. [doi: 10.1164/rccm.202002-0328OC] [Medline: 32325005]
17. Cropsey KL, Jardin BF, Burkholder GA, Clark CB, Raper JL, Saag MS. An algorithm approach to determining smoking cessation treatment for persons living with HIV/AIDS: results of a pilot trial. J Acquir Immune Defic Syndr 2015 Jul 01;69(3):291-298 [FREE Full text] [doi: 10.1097/QAI.0000000000000579] [Medline: 26181705]
18. Valera P, McClernon FJ, Burkholder G, Mugavero MJ, Willig J, O'Cleirigh C, et al. A pilot trial examining african american and white responses to algorithm-guided smoking cessation medication selection in persons living with HIV. AIDS Behav 2017 Jul;21(7):1975-1984 [FREE Full text] [doi: 10.1007/s10461-016-1634-0] [Medline: 27942999]
19. Martinez FJ, Mannino D, Leidy NK, Malley KG, Bacci ED, Barr RG, High-Risk-COPD Screening Study Group *. A new approach for identifying patients with undiagnosed chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2017 Mar 15;195(6):748-756 [FREE Full text] [doi: 10.1164/rccm.201603-0622OC] [Medline: 27783539]
20. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999 Jan;159(1):179-187. [doi: 10.1164/ajrccm.159.1.9712108] [Medline: 9872837]
21. Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B. Changes in the normal maximal expiratory flow-volume curve with growth and aging. Am Rev Respir Dis 1983 Jun;127(6):725-734. [doi: 10.1164/arrd.1983.127.6.725] [Medline: 6859656]
22. Biener L, Abrams DB. The Contemplation Ladder: validation of a measure of readiness to consider smoking cessation. Health Psychol 1991;10(5):360-365. [doi: 10.1037//0278-6133.10.5.360] [Medline: 1935872]
23. Tuck BM, Karelitz JL, Tomko RL, Dahne J, Cato P, McClure EA. Mobile, remote, and individual focused: comparing breath carbon monoxide readings and abstinence between smartphone-enabled and stand-alone monitors. Nicotine Tob Res 2021 Mar 19;23(4):741-747 [FREE Full text] [doi: 10.1093/ntr/ntaa203] [Medline: 33022057]
24. Lachin JM. Statistical considerations in the intent-to-treat principle. Control Clin Trials 2000 Jun;21(3):167-189. [doi: 10.1016/s0197-2456(00)00046-5] [Medline: 10822117]

25. Shakkottai A, Nasr SZ. The use of home spirometry in pediatric cystic fibrosis patients: results of a feasibility study. Glob Pediatr Health 2017;4:2333794X17690315 [FREE Full text] [doi: 10.1177/2333794X17690315] [Medline: 28229102]
26. Logie K, Welsh L, Ranganathan SC. Telehealth spirometry for children with cystic fibrosis. Arch Dis Child 2020 Dec;105(12):1203-1205. [doi: 10.1136/archdischild-2019-317147] [Medline: 31694805]
27. Kupczyk M, Hofman A, Kołtowski, Kuna P, Łukaszyk M, Buczyłko K, et al. Home self-monitoring in patients with asthma using a mobile spirometry system. J Asthma 2021 Apr;58(4):505-511. [doi: 10.1080/02770903.2019.1709864] [Medline: 31877056]
28. US Preventive Services Task Force (USPSTF), Siu AL, Bibbins-Domingo K, Grossman DC, Davidson KW, Epling JW, et al. Screening for chronic obstructive pulmonary disease: US preventive services task force recommendation statement. JAMA 2016 Apr 05;315(13):1372-1377. [doi: 10.1001/jama.2016.2638] [Medline: 27046365]
29. Thee S, Stahl M, Fischer R, Sutharsan S, Ballmann M, Müller A, et al. A multi-centre, randomized, controlled trial on coaching and telemonitoring in patients with cystic fibrosis: conneCT CF. BMC Pulm Med 2021 Apr 21;21(1):131 [FREE Full text] [doi: 10.1186/s12890-021-01500-y] [Medline: 33882893]
30. Liu Y, Pleasants RA, Croft JB, Wheaton AG, Heidari K, Malarcher AM, et al. Smoking duration, respiratory symptoms, and COPD in adults aged ≥45 years with a smoking history. Int J Chron Obstruct Pulmon Dis 2015;10:1409-1416 [FREE Full text] [doi: 10.2147/COPD.S82259] [Medline: 26229460]

Abbreviations

CAPTURE: COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk
CO: carbon monoxide
COPD: chronic obstructive pulmonary disease
EHR: electronic health record
FDA: Food and Drug Administration
FEV1: forced expiratory volume in one second
MUSC: Medical University of South Carolina
OR: odds ratio
PEF: peak expiratory flow
PFT: Pulmonary Function Test
TAU: treatment as usual

Edited by R Kukafka; submitted 11.04.22; peer-reviewed by J Polo, F Martinez; comments to author 01.05.22; revised version received 08.06.22; accepted 11.07.22; published 30.08.22.

Please cite as:
Dahne J, Player MS, Strange C, Carpenter MJ, Ford DW, King K, Miller S, Kruis R, Hawes E, Hidalgo JE, Diaz VA
Proactive Electronic Visits for Smoking Cessation and Chronic Obstructive Pulmonary Disease Screening in Primary Care: Randomized Controlled Trial of Feasibility, Acceptability, and Efficacy
J Med Internet Res 2022;24(8):e38663
URL: https://www.jmir.org/2022/8/e38663
doi:10.2196/38663
PMID:36040766

Original Paper

Digital Device Exposure and Cognition Levels of Children in Low- and Middle-Income Countries: Cross-sectional Study in Cambodia

Hye Hyeon Kim[1*], PhD; JooHyun Lee[1*], BSN; Ho Heon Kim[1], RN, MS; Sangho Hwang[1], MS; Ilcheong Yi[2], PhD; Sambath Kao[3], MSc, MD; DooRa Kim[4], BS; Hyuk-Sang Sohn[5], PhD; Joohye Kim[6], PhD; Yejin Choi[4], BS; Sangchul Yoon[7], MD, PhD; Yu Rang Park[1], PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
[2]United Nations Research Institute for Social Development, Geneva, Switzerland
[3]Department of Malnutrition and Neurology, National Pediatric Hospital, Phnom Penh, Cambodia
[4]DoBrain Inc, Seoul, Republic of Korea
[5]Graduate School of Public Policy & Civic Engagement, Kyung Hee University, Seoul, Republic of Korea
[6]Department of Special Education, Baekseok University, Cheonan, Republic of Korea
[7]Department of Medical Humanities and Social Sciences, Yonsei University College of Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 10 5240 3434
Email: yurangpark@yuhs.ac

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/12/e44106

Abstract

Background: Policy makers and practitioners in low- and middle-income countries (LMICs) are increasingly focusing on the effectiveness of digital devices in the delivery of medical and educational services to children under resource constraints. It is widely known that digital literacy can be fostered through exposure to and education regarding digital devices, which can improve children's academic performance as well as their search and communication skills in the digital era. However, the correlation between the cognitive function of children and exposure and intensity of the exposure to digital devices has rarely been studied, and the association between digital device exposure and the socioeconomic characteristics and cognitive development of children in LMICs is unknown.

Objective: This study examines the association among exposure to digital devices, socioeconomic status, and cognitive function in children aged 3 to 9 years in Cambodia.

Methods: We used a survey of 232 children that gathered data on familiarity with digital devices, demographic characteristics, and socioeconomic status, as well as a Cambridge Neuropsychological Test Automated Battery test for cognitive function, to examine the association between possible barriers and factors that may influence the cognitive function of children in 2 Cambodian schools from April 22, 2019, to May 4, 2019. A comparative analysis was performed with and without digital exposure, and an association analysis was performed among the variables from the survey and cognitive function.

Results: Significant differences were observed in demographic and socioeconomic characteristics such as school location, family type, and family income according to digital device exposure. The results of the Cambridge Neuropsychological Test Automated Battery tests, except for 1 test related to executive function, indicated no significant differences ($P>.05$) between group A and group B or among the 4 subgroups. Pretest digital device experience and amount of time spent using digital devices during the test had no significant impacts on the cognitive development of the children. Conversely, the multivariate analyses showed that cognitive function was associated with educational expenses per child, school (location), family type, and family income.

Conclusions: These results provide evidence to policy makers and practitioners on the importance of improving socioeconomic conditions, leading to investment in education by implementing programs for children's cognitive development through digital devices in LMICs.

(J Med Internet Res 2022;24(8):e31206) doi:10.2196/31206

KEYWORDS

low- and middle-income countries; digital device exposure; children; cognitive function; socioeconomic status

Introduction

Background

Resource-constrained health systems in low- and middle-income countries (LMICs) are considered a key obstacle to achieving sustainable development goals (SDGs), specifically SDG 3, to "ensure healthy lives and promote well-being for all at all ages," and SDG 4, to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" [1-3]. As LMICs are experiencing an unprecedented increase in the number and use of digital devices such as mobile phones, digital health and education initiatives that capitalize on the widespread use of these devices are emerging [4,5]. A digital device is a physical piece of equipment that uses digital data in some way, such as sending, receiving, storing, or processing data [6]. Despite controversies over various digital initiatives and programs using digital devices, such as "One Laptop Per Child," policy makers and practitioners are still interested in the potential of digital technology to address issues such as the digital divide, the impact of resource constraints on cognitive development, and developmental disabilities in children [7-11]. However, will digital device exposure and intensity of the exposure affect cognitive development in children in LMICs? This question is difficult to answer because most of the research targets high-income countries, adolescents, and adults [12-14].

In addition to foundational skills such as literacy and numeracy, digital literacy and skills are also key to implementing the SDGs in this digital age [2]. Therefore, many have suggested using digital devices for children's cognitive development to improve information processing, communication skills, and educational attainment [10,15]. Studies have demonstrated that digital devices provide only short-term improvements in children's cognitive abilities during interventions [11,16-19]. A recent study on patients with an intellectual disability diagnosis showed significant improvement in cognitive function using digital devices [20]. Moreover, most studies have focused on the experience of high-income countries or different age groups and patient-only analyses [14,20-23]. The evidence provided by these studies may not be relevant to LMICs because of their different socioeconomic environments and cultures. Although socioeconomic characteristics of households, both within and without, are known to significantly affect cognitive development, few studies on digital device–using interventions, including on the impact of various levels of digital device exposure on children's cognitive development, have been conducted in LMICs [24].

This Study

This study addressed the following questions through a survey and cognitive function tests of children in LMICs:

1. Do demographic and socioeconomic characteristics differ according to digital device exposure in LMICs?
2. Do digital device exposure and intensity of the exposure affect cognitive function in children in LMICs?
3. Is there an association among digital device exposure, demographic and socioeconomic characteristics, and cognitive function of children in LMICs?

Methods

Study Design

This cross-sectional study was designed to identify and analyze the correlation between cognitive function and exposure to digital devices in children aged 3 to 9 years at 2 schools in Cambodia. The study focused on the following urban and rural regions in Cambodia to consider the effects of various conditions such as average literacy rates, wealth and income distribution, and educational environments: Sisophon in Banteay Meanchey (rural population: 24%-73%) and Sangkat Chaom Chaov in Phnom Penh (rural population: 6%-40%). Sisophon has a relatively small share of the nationally estimated high-wealth quintile (22.9%) and a lower average of schooling years than Sangkat Chaom Chaov (high-wealth quintile: 84.4%). The schools selected were Xavier Jesuit School (rural) and Mirero School (urban). Xavier Jesuit School has both kindergarten and elementary classes. In total, 4 classes—2 kindergarten classes, 1 first grade class, and 1 second grade class—were selected. As Mirero School has only elementary classes, we selected 2 classes each from the first and second grades.

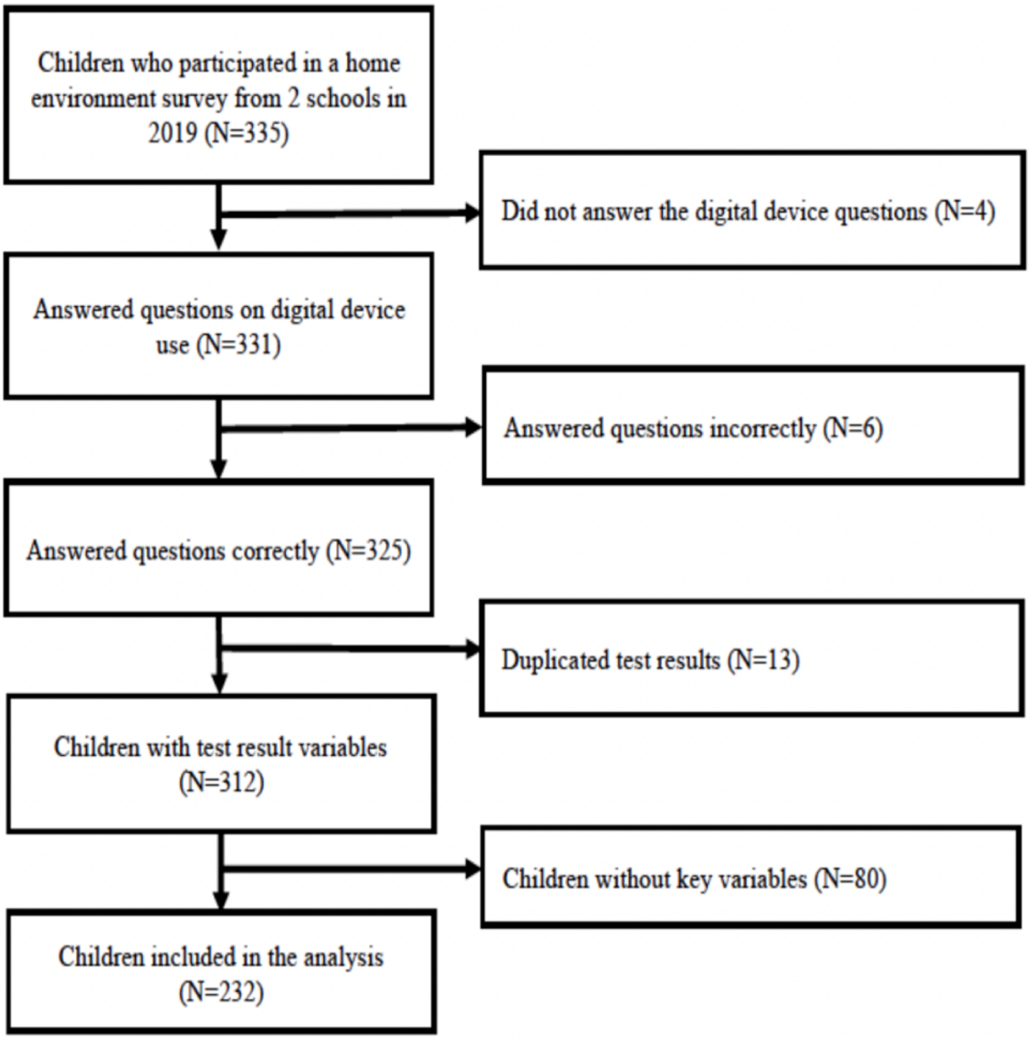
Participant Enrollment

To target children eligible for the cross-sectional study, we selected regions and schools with the advice of the Korea International Cooperation Agency Cambodia Office, which employs experts who are aware of the overall living environment, including education, in Cambodia. After the selection of areas and schools, classrooms and grades of age that met the inclusion criteria were selected.

Study participants were enrolled from 2 elementary schools in Cambodia's 2 regions: rural and urban. The inclusion criteria for children were as follows: (1) male and female students aged 3 to 9 years, (2) no environmental change in their home or school during the study, and (3) children who provided consent from their legal guardian or themselves. Students were excluded from the test if they could not use the study's digital device because of physical conditions or when legal guardians did not provide consent for the test. Figure 1 shows details of the participant selection flow.

The participants were divided into 2 groups based on their responses to the questionnaire about digital device experience: group A comprised children with digital device experience, and group B comprised those without digital device experience. The participants in group A were subsequently divided into 4 subgroups according to the duration (in minutes) of digital device use: group A-1: <30 minutes per day, group A-2: 30 to 60 minutes per day, group A-3: 60 to 90 minutes per day, and group A-4: >90 minutes per day. Of the 232 participants, 162 (69.8%) were in group A, with 95 (58.6%) in group A-1, 47 (29%) in group A-2, 10 (6.2%) in group A-3, and 10 (6.2%) in group A-4, whereas group B had 70 (30.2%) participants.

Figure 1. Participant selection flow.



Ethics Approval

All participants provided written informed consent before enrollment in the study. Consent was granted by their guardians after the study was explained in writing. Ethical clearance was obtained from the Ministry of Health's National Ethics Committee for Health Research in Cambodia (212).

Data Collection

We gathered data on demographic and socioeconomic characteristics through surveys and data on cognitive function using the Cambridge Neuropsychological Test Automated Battery (CANTAB) test. The survey targeted the basic socioeconomic characteristics and demographics of the participants and their guardians from April 22, 2019, to May 4, 2019. Most of the guardians were either the parents or grandparents. At both schools, the survey was delivered to the guardians living with the participant. The response rate was 82.9%. Most (335/404, 82.9%) of the participants returned the completed form to the school. In cases where the guardians could not read, schoolteachers assisted them in completing the survey at school. The questionnaire had 2 parts. The first elicited demographic information about the children, such as sex, age, and siblings, as well as digital device exposure and use. The second collected information about the family's socioeconomic status, including residential conditions, guardians' occupation and education level, household income, and educational expenditure (Multimedia Appendix 1). The participants' cognitive function, depending on their experience with digital devices, was evaluated through the CANTAB test using a tablet device. The CANTAB test includes highly sensitive, precise, and objective measures of cognitive function correlated with neural networks [25]. It includes tests that evaluate 4 cognitive areas: attention and psychomotor speed, executive function, memory, and social and emotional cognition. The participants were asked to perform 5 tests: motor screening task, reaction time, spatial working memory (SWM), pattern recognition memory, and spatial span. We collected 53 test result variables, including 11 key variables, for measuring the outcome of each test (Table 1 and Multimedia Appendix 2).

Table 1. Summary of the Cambridge Neuropsychological Test Automated Battery tests.

Cognitive function and text	Description	Key variable
Attention and psychomotor speed		
Motor screening	To evaluate response speed and pointing accuracy (selecting the cross), participants are asked to select the cross that appears on the screen as quickly and accurately as possible [26]	Mean latency from stimulus (MOTML[a])
Reaction time	Assesses simple reaction time and movement during simple and 5-choice reaction time trials	Median 5-choice reaction time (RTIFMDRT[b]); median 5-choice movement time (RTIFMDMT[c])
Executive function		
Spatial working memory	Test to find individual hidden tokens without returning to a box where one has previously been found [27]	Total between errors (SWMBE468[d]); between errors (4, 6, and 8 boxes; SWMBE 4,6, and 8, respectively); strategy score (SWMS[e])
Memory		
Pattern recognition memory	A 2-choice test of abstract visual pattern recognition memory [28]	Percent correct immediate (PRMPCI[f]); percent correct delayed (PRMPCD[g])
Spatial span	Test to recall the order in which a series of boxes was highlighted [28]	Longest successful sequence (SSPFSL[h])

[a] MOTML: motor screening task mean latency.

[b] RTIFMDRT: reaction time median 5-choice reaction time.

[c] RTIFMDMT: reaction time median 5-choice movement time.

[d] SWMBE468: spatial working memory between errors (4, 6, and 8 boxes).

[e] SWMS: spatial working memory strategy.

[f] PRMPCI: pattern recognition memory percent correct immediate.

[g] PRMPCD: pattern recognition memory percent correct delayed.

[h] SSPFSL: spatial span forward span length.

Statistical Analysis

To compare the demographic differences between groups A and B, a chi-square test was performed on categorical variables such as sex, and a Mann-Whitney *U* test was performed on continuous variables such as age. To compare the differences in cognitive function between the 2 groups, normality was tested for the variables using the Shapiro-Wilk test. Variables satisfying normality were examined using a 2-tailed *t* test, and those that did not satisfy normality were compared between the groups using the Mann-Whitney *U* test. For each comparison, the effect size was calculated for the 2 groups according to digital device exposure and for the 4 subgroups according to digital device use time, and the results of the normality test were compared for possible type 1 statistical errors. In comparing the 2 main groups, Cohen *d* was calculated for the *t* tests, and *r* was calculated using the Mann-Whitney *U* test. To compare the effect size among the 4 subgroups, eta squared (η^2) was calculated for ANOVA, and epsilon squared (^{2}) was used for the Kruskal-Wallis test. The threshold of statistical significance was set at $P<.05$, and an effect size greater than the small size, depending on its type (Cohen $d \approx \pm 0.20$: small, $r \approx \pm 0.10$: small, $\eta^2 \approx 0.01$: small, and $^2 \approx 0.01$: small), was considered significant for 2-tailed *t* tests. To confirm the association among demographic and socioeconomic characteristics, digital device familiarity, and cognitive function in children, univariate regression analysis was performed for 3 cognitive domains and 11 variables flowing from the CANTAB test. In the multivariate linear regression analysis, only variables that were statistically significant ($P<.05$) through univariate regression analysis were selected, and their effect on the cognitive function variable was evaluated with and without adjusting for age and sex. All statistical analyses were conducted using R (version 3.6.3; R Foundation for Statistical Computing) and Python (version 3.7; Python Software Foundation).

Results

Participant Flow

A total of 335 children participated in this study. Of these 335 children, 4 (1.2%) who did not answer the questions about digital device exposure and 6 (1.8%) who answered inconsistently (eg, talking about the purpose of the device, not the experience) were excluded from the data analysis; in addition, 13 (3.9%) children with duplicated test results and 80 (23.9%) without values for the key variables of each test were excluded. Ultimately, of the 335 children, 232 (69.3%) were included in the data analysis (Figure 1).

Overall Population

Of the 232 children, 162 (69.8%) were in group A, and 70 (30.2%) were in group B. The 162 students in group A comprised 110 (67.9%) Xavier Jesuit School students and 52 (32.1%) Mirero School students, a significant difference (*P*<.05). No significant difference between the groups was observed in terms of sex: in group A, 59.3% (96/162) of the participants were male students, whereas in group B, 56% (39/70) were male students (*P*=.72). The mean ages of the participants in group A and group B were 7.3 (SD 1.5) years and 7.6 (SD 1.3) years, respectively, without significant differences (*P*=.15). The proportion of participants with monthly family income of <US $150 was higher in group B (38/70, 54% vs 53/162, 32.7% in group A), with significant differences (*P*=.004). The proportion of participants with monthly family income of >US $350 was higher in group A (25/162, 15.4% vs 4/70, 6% in group B). The proportion of students whose mothers had secondary education and above was far higher in group A (73/162, 45.1% vs 17/70, 24% in group B), with significant differences (*P*=.01). The proportion of families who spent >US $30 per month on education per child was also higher in group A (24/162, 14.8% vs 4/70, 6% in group B), without overall differences between the 2 groups (*P*=.07; Table 2).

Table 2. Comparison of participants' demographic and socioeconomic characteristics according to digital device exposure[a].

Variable	Group A[b], n=162	Group B[c], n=70	Total, N=232	*P* value
School, n (%)				<.001
Xavier Jesuit (rural)	110 (67.9)	31 (44.3)	141 (60.8)	
Mirero (urban)	52 (32.1)	39 (55.7)	91 (39.2)	
Sex, n (%)				.72
Female	66 (40.7)	31 (44.3)	97 (41.8)	
Male	96 (59.3)	39 (55.7)	135 (58.2)	
Age (years), mean (SD)	7.3 (1.5)	7.6 (1.3)	7.4 (1.4)	.15
Family type, n (%)				.005
Other	11 (6.8)	4 (5.7)	15 (6.5)	
Only father	3 (1.9)	4 (5.7)	7 (3)	
Only mother	13 (8)	12 (17.1)	25 (10.8)	
Parents and grandparents living together	14 (8.6)	11 (15.7)	25 (10.8)	
Parents living together	121 (74.7)	37 (52.9)	158 (68.1)	
No response	0 (0)	2 (2.9)	2 (0.9)	
Family monthly income (US $), n (%)				.004
<150	53 (32.7)	8 (54.3)	91 (39.2)	
150 to 250	47 (29)	21 (30)	68 (29.3)	
250 to 350	37 (22.8)	6 (8.6)	43 (18.5)	
350 to 450	10 (6.2)	1 (1.4)	11 (4.7)	
>450	15 (9.3)	3 (4.3)	18 (7.8)	
Education expense per child per month (US $), n (%)				.07
15	53 (32.7)	38 (54.3)	91 (39.2)	
15 to 30	87 (53.7)	36 (51.4)	123 (53)	
>30	24 (14.8)	4 (5.7)	28 (12.1)	
No response	0 (0)	1 (1.4)	1 (0.4)	

[a] Full table has been presented in Multimedia Appendix 2.

[b] Digital device exposure group.

[c] Digital device nonexposure group.

Digital Device Exposure

When comparing the results of the CANTAB test between the 2 groups, the SWM between errors (4 boxes; SWMBE4) variable of group A had a median of 2.0 (IQR 1.0-3.0), whereas that of group B had a median of 2.0 (2.0-3.0); there were significant differences (*P*=.01). However, the effect size was small (*P*=.02). The SWM strategy (SWMS) variable of group A had a median of 12.0 (IQR 10.0-58.3), whereas that of group B had a median of 10.0 (IQR 9.0-12.0); there were significant differences (*P*.002). The effect size was small (*P*=.04). There were no significant differences between groups A and B for the other variables. The smaller values for the motor screening task mean latency (MOTML) and reaction time median 5-choice reaction time (RTIFMDRT) variables were positive, but the median values of group A were 15.5 and 21.2 points higher, respectively, than those in group B. Conversely, the reaction time median 5-choice movement time variable for the exposure group was 13.5 points lower, without a significant difference in median values among the other variables (Table 3).

Table 3. Comparison of cognitive function according to digital device exposure.

Cognitive function and variable	Group A[a], n=162, median (IQR)	Group B[b], n=70, median (IQR)	Total, N=232, median (IQR)	*P* value	Effect size (*r*)
Attention and psychomotor speed					
MOTML[c] (ms)[d]	781.1 (690.8-928.9)	765.6 (680.5-968.1)	776.1 (685.6-953.7)	.93	–0.004
RTIFMDRT[e] (ms)[d]	517.0 (462.0-578.0)	495.8 (460.0-551.0)	512.0 (461.0-566.0)	.29	0.001
RTIFMDMT[f] (ms)[d]	271.5 (232.0-325.5)	285.0 (238.5-342.5)	275.0 (235.0-330.5)	.11	0.007
Memory					
PRMPCI[g] (%)	75.0 (58.3-91.7)	70.8 (50.0-83.3)	75.0 (50.0-91.7)	.52	–0.003
PRMPCD[h] (%)	66.7 (50.0-75.0)	58.3 (50.0-75.0)	66.7 (50.0-75.0)	.08	0.009
SSPFSL[i] (n)	4.0 (3.0-5.0)	4.0 (3.0-5.0)	4.0 (3.0-5.0)	.99	–0.004
Executive function					
SWMBE468[j] (n)[d]	24.0 (20.0-28.0)	25.0 (22.0-28.0)	24.0 (20.0-28.0)	.18	0.003
SWMBE4[k] (n)[d]	2.0 (1.0-3.0)	2.0 (2.0-3.0)	2.0 (1.0-3.0)	.01	0.023
SWMBE6[l] (n)[d]	7.0 (5.0-9.0)	7.0 (6.0-9.0)	7.0 (5.0-9.0)	.60	–0.003
SWMBE8[m] (n)[d]	14.5 (12.0-17.0)	15.0 (13.0-17.0)	15.0 (13.0-17.0)	.61	–0.003
SWMS[n] (n)[d]	12.0 (10.0-58.3)	10.0 (9.0-12.0)	11.0 (10.0-50.0)	.002	0.036

[a] Digital device exposure group.

[b] Digital device nonexposure group.

[c] MOTML: motor screening task mean latency.

[d] Smaller values indicate more positive changes.

[e] RTIFMDRT: reaction time median 5-choice reaction time.

[f] RTIFMDMT: reaction time median 5-choice movement time.

[g] PRMPCI: pattern recognition memory percent correct immediate.

[h] PRMPCD: pattern recognition memory percent correct delayed.

[i] SSPFSL: spatial span forward span length.

[j] SWMBE468: spatial working memory between errors (4, 6, and 8 boxes).

[k] SWMBE4: spatial working memory between errors (4 boxes).

[l] SWMBE6: spatial working memory between errors (6 boxes).

[m] SWMBE8: spatial working memory between errors (8 boxes).

[n] SWMS: spatial working memory strategy.

Digital Device Use Time

When comparing the differences in cognitive levels based on the duration of digital device use, the results showed no significant differences among the 4 subgroups, with all effect sizes being small. Although there were no significant differences, the median values of the 2 variables measuring attention and psychomotor speed in group A-4 were higher than those in group A-1. Regarding the visual memory variables, there were no significant differences among the 4 subgroups, and the distribution did not show a linear relationship. Regarding the executive function test, as use time increased, the number of errors decreased, leading to positive results. Regarding the SWM test, the distributions of the 4 subgroups were not different (Table 4).

Table 4. Comparison of Cambridge Neuropsychological Test Automated Battery test results according to digital device use time.

Cognitive function and variable	Digital device use time					*P* value	Effect size (ε^2)
	Group A-1[a], n=95, median (IQR)	Group A-2[b], n=47, median (IQR)	Group A-3[c], n=10, median (IQR)	Group A-4[d], n=10, median (IQR)	Total, N=162, median (IQR)		
Attention and psychomotor speed							
MOTML[e] (ms)	785.2 (699.3-988.3)	760.5 (660.4-866.8)	835.2 (641.9-909.1)	864.5 (739.8-1065.9)	781.1 (690.8-928.9)	.24	0.143
RTIFMDRT[f] (ms)	510.0 (453.0-569.0)	519.5 (468.8-560.2)	596.5 (478.0-661.0)	530.5 (468.5-607.0)	517.0 (462.0-578.0)	.48	0.070
RTIFMDMT[g] (ms)	278.0 (240.0-333.2)	260.0 (225.2-295.5)	270.5 (243.5-324.0)	257.8 (191.0-418.0)	271.5 (232.0-325.5)	.26	0.037
Memory							
PRMPCI[h] (%)	66.7 (50.0-83.3)	83.3 (62.5-91.7)	83.3 (58.3-91.7)	75.0 (66.7-91.7)	75.0 (58.3-91.7)	.21	0.023
PRMPCD[i] (%)	66.7 (50.0-75.0)	66.7 (54.2-75.0)	70.8 (58.3-75.0)	70.8 (58.3-83.3)	66.7 (50.0-75.0)	.35	0.033
SSPFSL[j] (n)	4.0 (3.0- 5.0)	4.0 (3.0- 5.0)	3.5 (3.0- 4.0)	4.0 (4.0- 5.0)	4.0 (3.0- 5.0)	.64	–0.002
Executive function							
SWMBE468[k] (n)	24.0 (20.0-28.0)	24.0 (20.0-28.0)	24.5 (21.0-33.0)	21.0 (19.0-28.0)	24.0 (20.0-28.0)	.70	0.000
SWMBE4[l] (n)	2.0 (0.5- 3.0)	2.0 (1.0- 2.0)	2.0 (1.0- 4.0)	1.5 (1.0- 4.0)	2.0 (1.0- 3.0)	.87	0.020
SWMBE6[m] (n)	7.0 (5.0-10.0)	7.0 (5.0- 9.0)	8.0 (6.0-10.0)	6.0 (4.0-10.0)	7.0 (5.0- 9.0)	.66	–0.006
SWMBE8[n] (n)	15.0 (13.0-17.0)	14.0 (12.5-17.5)	15.5 (12.0-19.0)	13.5 (12.0-15.0)	14.5 (12.0-17.0)	.64	–0.004
SWMS[o] (n)	9.0 (8.5-10.0)	10.0 (9.0-11.0)	9.0 (8.0-11.0)	10.0 (9.0-11.0)	10.0 (9.0-11.0)	.07	0.176

[a] Digital device use time <30 minutes per day.

[b] Digital device use time 30 to 60 minutes per day.

[c] Digital device use time 60 to 90 minutes per day.

[d] Digital device use time >90 minutes per day.

[e] MOTML: motor screening task mean latency.

[f] RTIFMDRT: reaction time median 5-choice reaction time.

[g] RTIFMDMT: reaction time median 5-choice movement time.

[h] PRMPCI: pattern recognition memory percent correct immediate.

[i] PRMPCD: pattern recognition memory percent correct delayed.

[j] SSPFSL: spatial span forward span length.

[k] SWMBE468: spatial working memory between errors (4, 6, and 8 boxes).

[l] SWMBE4: spatial working memory between errors (4 boxes).

[m] SWMBE6: spatial working memory between errors (6 boxes).

[n] SWMBE8: spatial working memory between errors (8 boxes).

[o] SWMS: spatial working memory strategy.

Cognitive Function and Socioeconomic Status

All results of univariate linear regression analysis to determine the relationship between demographic characteristics and socioeconomic status and CANTAB test score have been presented in Multimedia Appendix 2.

The measures showed that the 3 variables MOTML, RTIFMDMT, and RTIFMDRT from the *Attention and psychomotor speed* cognitive function were significantly associated with *Age* (all *P*<.001), *School* (all *P*<.001), and *Education expense per child* (all *P*<.001) in the univariate model and multivariate model 1. However, in multivariate model 2, adjusted for age and sex, MOTML was significantly associated with *School*, and RTIFMDRT was significantly associated with *Family type* (Table 5).

Table 5. Univariate and multivariate regression models for relationship between cognitive function and survey variables.

Cognitive function and variable	Survey variable	Univariate model		Multivariate model 1		Multivariate model 2[a]	
		Coefficients	*P* value	Coefficients	*P* value	Coefficients	*P* value
Attention and psychomotor speed							
MOTML[b]							
	Age (years)	*–69.718*[c]	*<.001*	*–42.027*	*.001*	*–42.851*	*<.001*
	Education expense per child per month (US $; vs less than US $15)						
	15 to 30	*131.021*	*<.001*	–59.960	.08	–36.11	.29
	>30	–5.502	.91	43.200	.08	32.85	.18
	School (vs Mirero)	*185.077*	*<.001*	*157.79*	*<.001*	*88.74*	*.01*
RTIFMDMT[d]							
	Age (years)	*–16.631*	*<.001*	*–11.968*	*.003*	*–11.654*	*.004*
	Education expense per child per month (US $; vs less than US $15)						
	15 to 30	*41.118*	*<.001*	–20.140	.07	–14.577	.19
	>30	5.143	.72	*17.669*	*.03*	15.32	.05
	School (vs Mirero)	*33.564*	*<.001*	*24.352*	*.01*	7.923	.49
RTIFMDRT[e]							
	Age	*–40.904*	*<.001*	*–40.238*	*<.001*	*–40.142*	*<.001*
	Education expense per child per month (US $; vs less than US $15)						
	15 to 30	*61.556*	*<.001*	–20.140	.07	–14.577	.19
	>30	4.296	.83	*17.669*	*.03*	15.32	.05
	Family type (vs Other)						
	Only father	–46.919	.29	–50.240	.23	–27.001	.46
	Only mother	*–77.933*	*.01*	*–64.930*	*.03*	–41.667	.11
	Parents and grandparents living together	*–62.333*	*.049*	*–74.410*	*.01*	–45.816	.08
	Parents living together	*–64.801*	*.01*	*–61.050*	*.01*	*–43.347*	*.04*
	School (vs Mirero)	*68.184*	*<.001*	*69.650*	*<.001*	1.669	.90
Memory							
PRMPCI[f]							
	Age (years)	*4.267*	*<.001*	*4.263*	*<.001*	*4.326*	*<.001*
	Family type (vs other)						
	Only father	16.587	.08	16.911	.07	14.835	.11
	Only mother	11.778	.09	10.505	.12	8.324	.21
	Parents and grandparents living together	11.111	.10	12.293	.07	10.379	.12
	Parents living together	*15.246*	*.01*	*15.054*	*.01*	*13.540*	*.01*
	School (vs Mirero)	*–6.705*	*.02*	*–6.817*	*.02*	0.521	.88
PRMPCD[g]							
	Age (years)	*3.535*	*<.001*	*3.297*	*<.001*	*3.297*	*<.001*
	Education expense per child per month (US $; vs less than US $15)						
	15 to 30	*–5.389*	*.03*	*6.632*	*.01*	3.686	.17

Cognitive function and variable	Survey variable	Univariate model		Multivariate model 1		Multivariate model 2[a]	
		Coefficients	*P* value	Coefficients	*P* value	Coefficients	*P* value
	>30	3.122	.39	–1.203	.56	–0.408	.84
	Family income per month (US $; vs less than US $150)						
	150 to 250	2.497	.37	1.811	.52	1.221	.66
	250 to 350	5.715	.07	6.012	.06	5.846	.06
	350 to 450	–5.824	.29	–7.054	.19	–5.608	.29
	>450	*10.000*	*.03*	*10.095*	*.02*	*11.375*	*.008*
SSPFSL[h]							
	Age (years)	*0.226*	*<.001*	*0.226*	*<.001*	*0.225*	*<.001*
Executive function							
SWMBE468[i]							
	Education expense per child per month (US $; vs less than US $15)						
	15 to 30	–0.606	.46	–1.438	.11	–1.069	.25
	>30	*–2.628*	*.03*	–1.276	.06	*–1.397*	*.04*
	Sex (vs male)	*1.509*	*.048*	–1.582	.04	–1.570	.04
SWMBE4[j]							
	Digital device exposure (vs nonexposure group)	*–0.473*	*.02*	–0.389	.06	–0.390	.06
	Education expense per child per month (US $; vs less than US $15)						
	15 to 30	*–0.504*	*.01*	0.096	.67	0.119	.60
	>30	–0.286	.33	–0.262	.01	–0.272	.09
	School (vs Mirero)	*–0.574*	*.002*	*–0.412*	*.04*	*–0.480*	*.04*
SWMBE8[k]							
	Age (years)	*–0.372*	*.047*	*–0.379*	*.04*	*–0.379*	*.04*
	Sex (vs male)	*1.124*	*.04*	*–1.141*	*.03*	*–1.141*	*.03*
SWMS[l]							
	Age (years)	*–5.086*	*<.001*	*2.496*	*.046*	*2.540*	*.04*
	Digital device exposure (vs nonexposure group)	*13.423*	*<.001*	*7.336*	*.03*	*7.190*	*.04*
	Family income per month (US $; vs less than US $150)						
	150 to 250	*–9.464*	*.03*	–1.301	.72	–1.105	.76
	250 to 350	–1.479	.76	–4.089	.32	–4.101	.31
	350 to 450	–6.306	.46	–7.818	.26	–7.365	.29
	>450	–1.204	.86	*–14.057*	*.01*	*–14.628*	*.009*
	Family type (vs other)						
	Only father	0.331	.98	0.551	.96	–0.869	.93
	Only mother	*–17.289*	*.047*	–10.895	.12	–12.283	.08
	Parents and grandparents living together	–4.302	.62	–9.254	.20	–10.276	.16
	Parents living together	–7.994	.27	–6.319	.28	–7.223	.22

Cognitive function and variable	Survey variable	Univariate model		Multivariate model 1		Multivariate model 2[a]	
		Coefficients	*P* value	Coefficients	*P* value	Coefficients	*P* value
	School (vs Mirero)	*32.062*	*<.001*	*32.221*	*<.001*	*37.065*	*<.001*

[a] Adjusted for age and sex.

[b] MOTML: motor screening task mean latency.

[c] Statistically significant data are shown in italics.

[d] RTIFMDMT: reaction time median 5-choice movement time.

[e] RTIFMDRT: reaction time median 5-choice reaction time.

[f] PRMPCI: pattern recognition memory percent correct immediate.

[g] PRMPCD: pattern recognition memory percent correct delayed.

[h] SSPFSL: spatial span forward span length.

[i] SWMBE468: spatial working memory between errors (4, 6, and 8 boxes).

[j] SWMBE4: spatial working memory between errors (4 boxes).

[k] SWMBE8: spatial working memory between errors (8 boxes).

[l] SWMS: spatial working memory strategy.

The 3 *Memory* cognitive function variables pattern recognition memory percent correct immediate, pattern recognition memory percent correct delayed, and spatial span forward span length were associated with *Age*, *Education expense per child*, *Family income*, *Family type*, and *School*. In the univariate model and multivariate model 1, pattern recognition memory percent correct immediate was significantly associated with *Age* (all *P*<.001), *Family type* (all *P*=.07), and *School* (all *P*<.05). However, in multivariate model 2, it was significantly related only to *Age* (*P*<.001) and *Family type* (*P*=.01). In the univariate model and multivariate model 1, pattern recognition memory percent correct delayed was significantly related to *Age* (all *P*<.001), *Education expense per child* (all *P*<.05), and *Family income* (all *P*<.05). However, in multivariate model 2, it was significantly associated only with *Age* (*P*<.001) and *Family income* (all *P*=.08). In all 3 regression models, spatial span forward span length was strongly related to *Age* (*P*<.001; Table 5).

The variables SWM between errors (4, 6, and 8 boxes), SWMBE4, SWM between errors (8 boxes), and SWMS corresponding to *Executive function* were related to *Age*, *Digital device exposure*, *Education expense per child*, *Family income per month*, *Family type*, *Sex*, and *School*. In the univariate model, SWM between errors (4, 6, and 8 boxes) was significantly related to *Education expense per child* (*P*=.03) and *Sex* (*P*=.05). However, in multivariate model 1, it was not associated with any variable. SWMBE4 was significantly associated with *School* (all *P*<.05) in the 3 regression models. Both *Age* (all *P*<.05) and *Sex* (all *P*<.05) had a significant association in the 3 regression analyses of SWM between errors (8 boxes). The SWMS variable showed a significant relationship with the 3 variables *Age* (all *P*<.05), *Digital device exposure* (all *P*<.05), and *School* (all *P*<.001) in all regression analyses (Table 5).

Discussion

Comparison With Prior Work

Since 2015, when the 2030 Agenda for Sustainable Development was adopted at the United Nations General Assembly, many initiatives for the application of digital technologies to policies and programs for development, particularly those associated with SDGs 3 and 4, emerged in various countries, particularly LMICs [5,29]. However, there is insufficient evidence on how digital device use affects children's cognitive improvement in LMICs [8,30,31]. This study focused on urban and rural Cambodian schools to provide a comprehensive perspective on the adoption and use of digital devices among children. We used a cross-sectional study to examine the cognitive level of 3 domains using the CANTAB, focusing on how exposure to digital devices affects cognitive development in elementary school–age children.

Principal Findings

Our study found significant differences in demographic and socioeconomic characteristics such as school location, family type, and family income according to digital device exposure. We found that children demonstrated neither superiority nor inferiority in cognitive scores on the CANTAB in 3 cognitive domains depending on digital device use. Given the lack of empirical research on the impact of digital device exposure on cognitive development, we offer two key findings that make a significant contribution to policies and programs for the application of digital devices to health care and education: (1) there is no significant association between exposure to digital devices and cognitive development; (2) however, socioeconomic conditions such as school location, family income, family type, and education expenditure are significantly related to cognitive function [31].

Strengths and Limitations

Several policy implications can be drawn from this study's findings. The first is the importance of an enabling environment that maximizes the impact of digital devices on cognitive function and development. The evidence demonstrates that socioeconomic factors significantly affect cognitive function and development in infancy, including memory and enforcement functions [24,27]. For example, US studies found that children from higher socioeconomic backgrounds had higher achievements in visuospatial, memory, and executive functions than those from middle or low socioeconomic backgrounds. Similarly, students from higher socioeconomic backgrounds in Brazil were found to have better memory and executive function results [24]. To what extent do socioeconomic conditions affect children's cognitive function and development in LMICs? The findings of our study demonstrate that these conditions, specifically education expenditure, enhanced cognitive function and development to a greater extent than did children's exposure to digital devices. Comments such as "we have higher priorities than laptops" expressed by many delegates to the 2005 World Summit on the Information Society, during which the "One Laptop Per Child" devices were introduced, were supported by our findings. Proper investment in education is a much more significant task than purchases of, and exposure to, digital devices, particularly in LMICs.

The second implication concerns the design and setting of digital devices. Many initiatives on the use of digital devices for children's cognitive function and development are based on a theoretical framework called the "brain's rearrangement capacity," which posits that children learn to associate what they see, hear, and know with symbolic characters [31]. Thus, we collected data indicating the symbolic characters that children see and hear on digital devices (Multimedia Appendix 1). Most children with digital devices experience a desire to play games and watch videos, such as those available on YouTube. By playing games and watching videos, children may not learn to associate what they see, hear, or know with symbolic characters. Our finding on the insignificant impacts of exposure to, or use of, digital devices on cognitive development suggests that the design and setting of the devices that children are exposed to should be effective enough to stimulate and accelerate the "brain's rearrangement capacity." Giving children access to digital devices with suitable designs and settings is more important than ever amid a pandemic such as the COVID-19 pandemic, during which almost all schools had closed for long periods worldwide [28].

Our study had several limitations. First, the levels of children's cognitive function and development were measured only during the survey; the long-term impact of digital device use on cognitive function and development was not assessed. This is a typical limitation of cross-sectional studies with a defined time frame. A longitudinal panel study based on groups of children with different socioeconomic backgrounds and levels of exposure to digital devices is needed to measure the long-term effects of digital devices on cognitive function and development. Second, our study was conducted in 2 areas of Cambodia. We focused on both rural and urban areas to examine the impact of differences in socioeconomic levels. However, this did not cover a sufficient area. Nevertheless, considering the lack of studies on the impact of digital device exposure on children's cognitive function and development in LMICs, particularly those focusing on both digital device exposure and socioeconomic conditions, our study expands the understanding of the enabling environment required for digital device–use initiatives aimed at children's cognitive function and development.

Acknowledgments

This work was supported by a grant from the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A3A2075117), Republic of Korea; a grant from the Institute for Information & Communications Technology Planning & Evaluation (IITP 2022-0-00234) funded by the Ministry of Science and ICT, Republic of Korea; and a grant from the Korea International Cooperation Agency Creative Technology Solution program, Republic of Korea. The funders had no role in the study design, data collection, analysis, manuscript preparation, or decision to publish. The corresponding author has full access to all data and the final responsibility for the decision to submit the paper for publication.

Data Availability

The data set of anonymized individual participants collected during the study period, along with the data specification, is available upon request from the author with access to the data.

Authors' Contributions

H Heon Kim, SH, and YRP envisioned, planned, and supervised this study. H Heon Kim and SH drafted the paper, including the statistical analysis of the data, the effectiveness of digital devices, and educational expenditures. IY, SK, HSS, JK, and SY contributed to cognitive abilities, effectiveness of digital devices, and expertise in sustainable development goals. DRK and YC contributed to data collection through the implementation of the Cambodia Cambridge Neuropsychological Test Automated Battery. H Hyeon Kim and JHL analyzed the data and wrote and modified the paper at the revision stage. All authors reviewed, commented on, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Basic information survey for the project.
[DOCX File , 26 KB - jmir_v24i8e31206_app1.docx]

Multimedia Appendix 2
Demographic characteristics and socioeconomic status of eligible participants.
[DOCX File , 53 KB - jmir_v24i8e31206_app2.docx]

References

1. Puchalski Ritchie LM, Khan S, Moore JE, Timmings C, van Lettow M, Vogel JP, et al. Low- and middle-income countries face many common barriers to implementation of maternal health evidence products. J Clin Epidemiol 2016 Aug;76:229-237 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.02.017] [Medline: 26931284]
2. Digital inclusion for low-skilled and low-literate people. UNESCO. URL: https://en.unesco.org/icted/content/digital-inclusion-low-skilled-and-low-literate-people [accessed 2022-08-01]
3. O'Donnell O. Access to health care in developing countries: breaking down demand side barriers. Cad Saude Publica 2007 Dec;23(12):2820-2834 [FREE Full text] [doi: 10.1590/s0102-311x2007001200003] [Medline: 18157324]
4. Dentzer S. E-health's promise for the developing world. Health Aff (Millwood) 2010 Feb;29(2):229. [doi: 10.1377/hlthaff.2010.0006] [Medline: 20348063]
5. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
6. The use of digital devices in marketing library products in an inclusively engaged academic library. In: Handbook of Research on Digital Devices for Inclusivity and Engagement in Libraries. Pennsylvania, United States: IGI Global; 2020.
7. Kraemer KL, Dedrick J, Sharma P. One laptop per child. Commun ACM 2009 Jun;52(6):66-73. [doi: 10.1145/1516046.1516063]
8. Warschauer M, Ames M. Can one laptop per child save the world’s poor? J Int Affairs 2010;64(1):33-51.
9. World Development Report 2016 : Digital Dividends. Washington, DC: World Bank; 2016.
10. United Nation's Children's Fund (UNICEF). The State of the World's Children 2017 Children in a Digital World. New York, United States: United Nations; 2018.
11. Cristia J, Ibarrarán P, Cueto S, Santiago A, Severín E. Technology and child development: evidence from the one laptop per child program. Am Econ J Appl Econ 2017 Jul 01;9(3):295-320. [doi: 10.1257/app.20150385]
12. Danovitch JH. Growing up with Google: how children's understanding and use of internet‐based devices relates to cognitive development. Human Behav and Emerg Tech 2019 Apr 26;1(2):81-90. [doi: 10.1002/hbe2.142]
13. Meherali S, Rahim KA, Campbell S, Lassi ZS. Does digital literacy empower adolescent girls in low- and middle-income countries: a systematic review. Front Public Health 2021 Dec 16;9:761394 [FREE Full text] [doi: 10.3389/fpubh.2021.761394] [Medline: 34976923]
14. Oh SS, Kim K, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. J Med Internet Res 2021 Feb 3;23(2):e26145. [doi: 10.2196/26145]
15. Information literacy competency standards for higher education. American Library Association Institution Repository. URL: https://alair.ala.org/handle/11213/7668 [accessed 2022-08-01]
16. Neisser U, Boodoo G, Bouchard TJ, Boykin AW, Brody N, Ceci SJ, et al. Intelligence: knowns and unknowns. Am Psychol 1996 Feb;51(2):77-101. [doi: 10.1037/0003-066x.51.2.77]
17. Malamud O, Pop-Eleches C. Home computer use and the development of human capital. Q J Econ 2011 May;126(2):987-1027 [FREE Full text] [doi: 10.1093/qje/qjr008] [Medline: 22719135]
18. Beuermann DW, Cristia J, Cueto S, Malamud O, Cruz-Aguayo Y. One laptop per child at home: short-term impacts from a randomized experiment in Peru. Am Econ J Appl Econ 2015 Apr 01;7(2):53-80. [doi: 10.1257/app.20130267]
19. Haskins R. Beyond metaphor: the efficacy of early childhood education. Am Psychol 1989;44(2):274-282. [doi: 10.1037/0003-066x.44.2.274]
20. Torra Moreno M, Canals Sans J, Colomina Fosch MT. Behavioral and cognitive interventions with digital devices in subjects with intellectual disability: a systematic review. Front Psychiatry 2021 Apr 13;12:647399 [FREE Full text] [doi: 10.3389/fpsyt.2021.647399] [Medline: 33927655]
21. Di Giacomo D, Ranieri J, Lacasa P. Digital learning as enhanced learning processing? Cognitive evidence for new insight of smart learning. Front Psychol 2017 Aug 03;8:1329 [FREE Full text] [doi: 10.3389/fpsyg.2017.01329] [Medline: 28824508]
22. Falloon G. Young students using iPads: app design and content influences on their learning pathways. Comput Educ 2013 Oct;68:505-521. [doi: 10.1016/j.compedu.2013.06.006]
23. Masataka N. Development of reading ability is facilitated by intensive exposure to a digital children's picture book. Front Psychol 2014 May 02;5:396 [FREE Full text] [doi: 10.3389/fpsyg.2014.00396] [Medline: 24822051]

24. Hackman DA, Farah MJ, Meaney MJ. Socioeconomic status and the brain: mechanistic insights from human and animal research. Nat Rev Neurosci 2010 Sep;11(9):651-659 [FREE Full text] [doi: 10.1038/nrn2897] [Medline: 20725096]
25. CANTAB The most sensitive and validated cognitive research software available. Cambridge Cognition. URL: https://www.cambridgecognition.com/cantab [accessed 2022-08-01]
26. Wilmer HH, Sherman LE, Chein JM. Smartphones and cognition: a review of research exploring the links between mobile technology habits and cognitive functioning. Front Psychol 2017;8:605 [FREE Full text] [doi: 10.3389/fpsyg.2017.00605] [Medline: 28487665]
27. Noble KG, McCandliss BD, Farah MJ. Socioeconomic gradients predict individual differences in neurocognitive abilities. Dev Sci 2007 Jul;10(4):464-480. [doi: 10.1111/j.1467-7687.2007.00600.x] [Medline: 17552936]
28. Policy Brief: education during COVID-19 and beyond. United Nations. URL: https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf [accessed 2022-05-20]
29. The economic impacts of learning losses. OECD. URL: https://www.oecd.org/education/The-economic-impacts-of-coronavirus-covid-19-learning-losses.pdf [accessed 2022-07-21]
30. Chang A, Tilahun L, Breazeal B. Visualisations of data from the literacy tablet reading project in rural Ethiopia. In: Proceedings of the Electronic Visualisation and the Arts (EVA 2014) (EVA). 2014 Presented at: Electronic Visualisation and the Arts (EVA 2014) (EVA); Jul 8 - 10, 2014; London UK. [doi: 10.14236/ewic/eva2014.35]
31. The reading brain, global literacy, and the eradication of poverty. In: Bread and Brain, Education and Poverty. Vatican City: Libreria Editrice Vaticana; 2015.

Abbreviations

CANTAB: Cambridge Neuropsychological Test Automated Battery
LMIC: low- and middle-income country
MOTML: motor screening task mean latency
RTIFMDRT: reaction time median 5-choice reaction time
SDG: sustainable development goal
SWM: spatial working memory
SWMBE4: spatial working memory between errors (4 boxes)
SWMS: spatial working memory strategy

Edited by T Leung; submitted 13.06.21; peer-reviewed by K Patel, S Bhattacharjya, M Salimi; comments to author 14.12.21; revised version received 18.06.22; accepted 13.07.22; published 31.08.22.

Please cite as:
Kim HH, Lee J, Kim HH, Hwang S, Yi I, Kao S, Kim D, Sohn HS, Kim J, Choi Y, Yoon S, Park YR
Digital Device Exposure and Cognition Levels of Children in Low- and Middle-Income Countries: Cross-sectional Study in Cambodia
J Med Internet Res 2022;24(8):e31206
URL: https://www.jmir.org/2022/8/e31206
doi:10.2196/31206
PMID:36044246

Original Paper

The Disparities in Patient Portal Use Among Patients With Rheumatic and Musculoskeletal Diseases: Retrospective Cross-sectional Study

Enid Y Sun[1], MPH, MD; Carolina Alvarez[2], MSc; Leigh F Callahan[2,3], PhD; Saira Z Sheikh[2,3], MD

[1]Section of Rheumatology, Department of Medicine, Temple Lewis Katz School of Medicine, Philadelphia, PA, United States

[2]Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, NC, United States

[3]Division of Rheumatology, Allergy, and Immunology, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, United States

Corresponding Author:
Saira Z Sheikh, MD
Thurston Arthritis Research Center
University of North Carolina
3300 Thurston Building
CB #7280
Chapel Hill, NC, 27599
United States
Phone: 1 919 843 6619
Email: szsheikh@email.unc.edu

Abstract

Background: During the COVID-19 pandemic, the shift to virtual care became essential for the continued care of patients. Individuals with rheumatic and musculoskeletal diseases (RMDs) especially require frequent provider visits and close monitoring. To date, there have been limited studies examining inequities in health technology use among patients with RMDs.

Objective: Our goal was to identify characteristics associated with patient portal use before and after the COVID-19 pandemic in a convenience sample of patients with RMDs from a large academic medical center.

Methods: In this cross-sectional study, Epic electronic medical record data were queried to identify established patients of the University of North Carolina Hospitals adult rheumatology clinic between November 1, 2017, through November 30, 2019. Demographic and clinical data were collected to compare MyChart (Epic's patient portal) users with nonusers before and after the COVID-19 pandemic. MyChart activation and use were modeled using logistic regression and adjusted odds ratios, and confidence intervals were estimated.

Results: We identified 5075 established patients with RMDs who met the inclusion criteria. Prior to the pandemic, we found that younger age (P<.001), suburban residence (P=.05), commercial/state insurance (P<.001), military insurance (P=.05), and median income >US $50,000 ($P$<.001) were associated with significantly higher odds of MyChart activation. Male sex (P<.001), being of Black or African American (P<.001) or "other" race (P<.001), Spanish as a primary language (P<.001), rural residence (P=.007), Medicaid insurance (P<.001), and median income of <US $25,000 ($P$=.01) were associated with lower odds of MyChart activation. Following COVID-19, younger age (P<.001), commercial insurance (P=.03), state insurance (P=.02), and median income of US $50,000-75,000 ($P$=.01) were associated with significantly higher odds of MyChart use. However, being of Black or African American (P<.001) or "other" race (P=.01), Spanish as a primary language (P=.002), male sex (P=.004), rural residence (P=.005), and having no insurance (P<.001) or Medicaid (P=.008) were associated with lower odds of MyChart use.

Conclusions: Residence in a rural area, being of minority race/ethnicity, older age, male sex, lower median income, Medicaid, being uninsured, and non-English primary language are associated with lower odds of patient portal activation and use. Future health policy and clinical practice measures should focus on reducing barriers to health technology adoption among these groups.

(J Med Internet Res 2022;24(8):e38802) doi:10.2196/38802

KEYWORDS

COVID-19; telemedicine; telehealth; health technology; health care disparities; patient portal; rheumatology; musculoskeletal diseases; chronic disease; digital health; MyChart; rural area; minority population; virtual care

Introduction

Rheumatic and musculoskeletal diseases (RMDs) are complex chronic conditions that require lifelong care. Patients may experience flares or acute complications related to their disease, requiring close communication with their rheumatologist. Patients with these conditions often take medications that require frequent monitoring and irregular dosing schedules. These aspects of RMD management require a high level of patient agency and open avenues for patient-provider contact and communication. Digital technology such as patient portals, health apps, and wearable technologies allow patients to manage and participate in their own care [1].

Many studies have shown positive effects on patient outcomes and satisfaction when patients are engaged in their own care through digital technologies [1-4]. In a systematic review by de Jong et al [2], patients who were able to communicate with their physicians had increased knowledge and self-management regarding their chronic condition, decreased health care visits, and improved psychosocial and clinical outcomes. In another study of patients with rheumatoid arthritis, patients who received weekly SMS text messages had better medication adherence than patients who did not receive the SMS text messaging intervention [3]. Participation in a web-based arthritis self-management program was associated with improved health status measures and self-efficacy in a study of patients with RMDs (ie, rheumatoid arthritis, osteoarthritis, or fibromyalgia) [4].

Health technology use became a necessity in early 2020 following the SARS-CoV-2 (COVID-19) outbreak, which forced health care systems around the world to adapt in the face of uncertainty. During this period, there has been a large shift to virtual care. Although this change has the possibility to close the gap in health care delivery in the United States, studies have shown that there are disparities in health technology use and the use of technology in general [5-11]. These studies have shown that low health literacy, lower educational attainment, residence in a rural area, being of minority race/ethnicity, and older age are associated with lower rates of health app and general technology use (ie, computer and cellphone ownership) [5-11].

To date, there have been limited studies examining inequities in health technology use among patients with rheumatologic conditions, and to our knowledge, none have looked at how COVID-19 has affected the patterns of health technology use among this patient population. Our goal was to identify the characteristics of patient portal users versus nonusers from a group of patients at a large hospital-based rheumatology clinic. We aimed to identify disparities and potential barriers to telehealth adoption among patients with rheumatologic conditions to help close the gap in health technology use.

Methods

Study Subjects

In this cross-sectional study, Epic electronic medical record data were queried to identify established patients of the University of North Carolina Hospitals (UNCH) adult rheumatology clinic between November 1, 2017, through November 30, 2019. "Established" patients were defined as patients who had at least one return visit during the 2-year study period. We specifically excluded "new" patients since these individuals may be seen for 1 consultative visit without further follow-up in the UNCH system. We felt that including these subjects could underestimate patient portal activation or use among our population.

Variables

Demographic and clinical data were collected from patient- and provider-entered information on Epic and used to compare the patients who activated Epic's patient portal (MyChart) to patients who did not activate MyChart at the time of the initial data acquisition. Additional data on MyChart usage were collected for the following year to compare MyChart use 8 months prior to the start of telemedicine visits at our clinic (from July 1, 2019, to March 30, 2020; "prepandemic") to the 8 months following the clinic's adoption of virtual care (from April 1, 2020, to December 2, 2020; "postpandemic"). MyChart "activation" indicates that the patient, or a patient proxy, has enrolled for patient portal access. MyChart "usage" was defined as the patient or patient-assigned proxy using MyChart to read or send patient-provider messages or manage appointments.

Demographic information collected included age, sex, race or ethnicity as documented in the electronic medical record (American Indian or Alaska Native or Native Hawaiian or Pacific Islander, Asian, Black or African American, Hispanic or Latino, White, or "other" race), primary language (English, Spanish, or "other"), zip code and county of primary residence, and primary insurance payor. Patients were grouped into generational categories based on age at the time of initial data collection (November 2019): born from 1997 to the present ("Gen Z," ages 17-22 years), born from 1981-96 ("Millennials," ages 23-38 years), born from 1965-80 ("Gen X," ages 39-54 years), born from 1946-64 ("Baby Boomers," ages 55-73 years), and born from 1928-45 ("Silent Generation," ages 74-91 years) [12]. North Carolina (NC) "rural," "suburban," and "urban" county designations were defined as average population densities of ≤250 people/square mile, 250-750 people/square mile, or ≥750 people/square mile, respectively, based on densities as reported in 2014 US Census population estimates [13]. Using individual income 2017 zip code data for NC from the Internal Revenue Service (IRS), we used the median gross income of each zip code to estimate individuals' annual median gross income. Estimated median adjusted gross income was grouped in quartiles as reported in the IRS data [14]. Income data were not available for all zip codes; income data was not reported for zip codes with a low number of returns or in cases of nonresidential zip codes [14]. Thus, patients with NC zip codes without income information and patients with out-of-state zip codes were excluded from analysis. The clinical data collected included the most recent outpatient visit date and number of clinic visits (≤2 vs ≥3) within the study period.

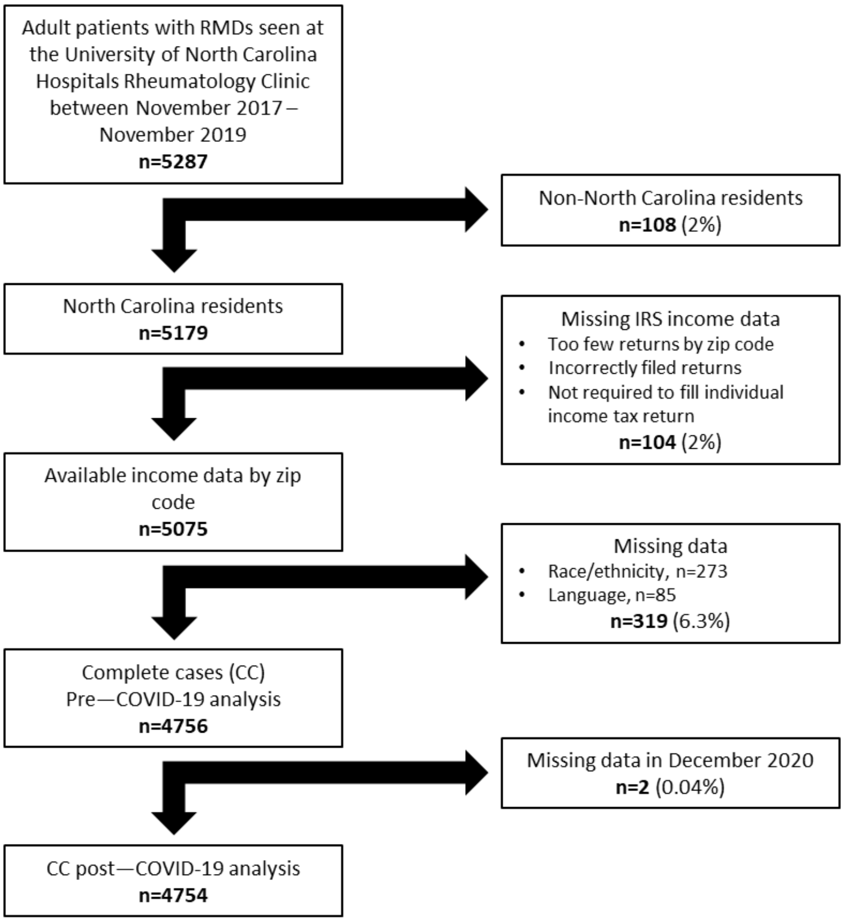
Statistical Analysis

Descriptive statistics were used to summarize study subjects and relevant variables. Counts and percentages were produced for categorical variables, whereas mean (SD) or median (IQR) were computed for continuous variables. Multivariable logistic regressions were separately modeled for the log odds of MyChart activation pre–COVID-19 (Model 1), MyChart use pre–COVID-19 and post–COVID-19 (Model 2), and MyChart use post–COVID-19 among nonusers pre–COVID-19 (Model 3). Model 2 used generalized estimating equations to account for the correlation between a patient's pre–COVID-19 and post–COVID-19 MyChart use and the interaction of pre–COVID-19 or post–COVID-19 time, with all covariables tested and retained if *P*<.05; otherwise, overall effects were shown for this model.

Models included all previously defined variables: visit date, the number of visits, age group, sex, race or ethnicity, primary language, county of residence, insurance, and median zip code–based income to produce adjusted odds ratios (aORs) and 95% CI. Complete cases were included for multivariable analyses, excluding subjects with missing variable items given that missing data rates (319/5075, 6.3%) were well below <10% (Figure 1) among NC residents with available IRS income by zip code.

Sensitivity analyses using multiple imputations of variables with missing information (race or ethnicity and primary language) were performed to assess the consistency of results. These variables were imputed using logistic regression by fully conditional specification methods for binary variables, which performs best for missing-at-random patterns and a missing proportion of less than 50%, to generate 10 imputed data sets for analyses [15]. Due to the exploratory nature of our study, corrections of statistical significance level were not performed [16]. All analyses were performed with SAS statistical software (version 9.4; SAS Institute Inc). Statistical significance was determined at *P*=.05.

Figure 1. Subject inclusion and exclusion. IRS: Internal Revenue Service; RMD: rheumatic and musculoskeletal disease.



Ethics Approval

Our study was reviewed and approved by the University of North Carolina's Institutional Review Board (protocol 19-3155) and adheres to the ethical principles of the Declaration of Helsinki. A waiver of informed consent was obtained due to the retrospective nature of our study.

Results

General Characteristics

We identified 5287 established patients who were seen at the UNCH rheumatology clinic during our study period, of whom 5075 patients were NC residents with available income data based on zip code. There were 4756 complete cases included in the pre–COVID-19 MyChart activation analyses and 4754 in post–COVID-19 MyChart use analyses (Figure 1).

Descriptive statistics for key characteristics are shown for the cohort of NC residents with income data (N=5075; Table 1). The mean age of the NC cohort was 54.7 (SD 15.4) years, and 73.9% (n=3749) were female. Of the 5075 established patients, 51% (n=2586) identified as White and 26.4% (n=1342) identified as Black or African American. In all, 88% (n=4478) identified English as their primary language, whereas 9.2% (n=469) reported Spanish as their primary language. Patients were almost evenly split between urban (n=1679, 33.1%), suburban (n=1506, 29.7%), and rural (n=1890, 37.2%) residences. Over two-thirds (n=3563, 70.2%) had a median adjusted gross income between US $25,000 to <US $50,000. Regarding insurance, 37% (n=1851) had Medicare, 25.1% (n=1276) had commercial insurance, 10.1% (n=512) had Medicaid, and 16.9% (n=859) were uninsured.

We examined MyChart activation among 3759 MyChart “activators” by patient characteristics (Table 2). We found that 74.1% (3759/5075) of our cohort had activated MyChart (Table 2). For age groups, 91% (111/122) of Gen Z patients had activated MyChart, whereas 69.7% (352/505) of those aged ≥75 years had activated MyChart. The rates of MyChart activation were 77.1% (2890/3749) among women and 65.5% (869/1326) among men. For race and ethnicity, 82% (2130/2586) of White patients and 87% (87/100) of Asian patients activated MyChart, whereas 63.6% (854/1342) of Black or African American patients and 65.4% (409/625) of Hispanic or Latino patients activated MyChart. MyChart activation was 76.3% (3416/4478) among English speakers and 57.4% (269/469) among Spanish speakers. The rates of MyChart activation among patients residing in a suburban county was 81.3% (1224/1506) and 65.4% (1237/1890) among patients residing in a rural county. Only 53.4% (86/161) of patients with an estimated median adjusted gross income of <US $25,000 activated MyChart, compared to 92.5% (124/134) of patients with a gross income of US $75,000 to <US $100,000. The rates of MyChart activation were high among individuals with commercial (1099/1276, 86.1%), state (319/350, 91.1%), and military (113/132, 85.6%) insurance, whereas 71.6% (1325/1851) of Medicare beneficiaries, 65.6% (336/512) of Medicaid recipients, and 65.3% (561/859) of uninsured individuals activated MyChart.

Table 1. Characteristics of patients—North Carolina residents with available income (N=5075).

Characteristic	Patients
Age group (years), n (%)	
17-24 (Gen Z)	122 (2.4)
25-39 (Millennials)	828 (16.3)
40-54 (Gen X)	1538 (30.3)
55-74 (Baby Boomers)	2082 (41)
≥75 (Silent Gen)	505 (10)
Sex, n (%)	
Female	3749 (73.9)
Male	1326 (26.1)
Race/ethnicity[a], n (%)	
American Indian or Alaska Native or Native Hawaiian or Pacific Islander	31 (0.6)
Asian	100 (2)
Black or African American	1342 (26.4)
Hispanic or Latino	625 (12.3)
White	2586 (51)
Other race	118 (2.3)
Primary language[b], n (%)	
English	4478 (88.2)
Spanish	469 (9.2)
Other	43 (0.8)
County of residence, n (%)	
North Carolina urban	1679 (33.1)
North Carolina suburban	1506 (29.7)
North Carolina rural	1890 (37.2)
Number of visits, median (IQR)	3 (1-4)
Median adjusted gross income (US $), n (%)	
<25,000	161 (3.2)
25,000 to <50,000	3563 (70.2)
50,000 to <75,000	1217 (24)
75,000 to <100,000	134 (2.6)
Insurance, n (%)	
Medicare	1851 (36.5)
Commercial	1276 (25.1)
Uninsured	859 (16.9)
Medicaid	512 (10.1)
State	350 (6.9)
Military	132 (2.6)
Department of Correction	95 (1.9)
Primary visit diagnosis, n (%)	
Rheumatoid arthritis	1320 (26)
Seronegative spondyloarthropathies	452 (8.9)

Characteristic	Patients
Crystal-induced arthropathies	175 (3.4)
Osteoarthritis and other arthropathies	524 (10.3)
Metabolic bone diseases and other musculoskeletal conditions	517 (10.2)
Miscellaneous inflammatory and autoimmune conditions	412 (8.1)
Lupus and other systemic connective tissue disorders	1252 (24.7)
Vasculitis	265 (5.2)
Other[c]	158 (3.1)

[a] Missing race/ethnicity (n=273).

[b] Missing primary language (n=85).

[c] Nonrheumatologic conditions, nonspecific symptoms, or laboratory abnormalities.

Table 2. MyChart activation by patient demographics (N=3759).

Characteristic	Patient, n/N (%)
Age group (years)	
17-24 (Gen Z)	111/122 (91)
25-39 (Millennials)	670/828 (80.9)
40-54 (Gen X)	1153/1538 (75)
55-74 (Baby Boomers)	1473/2082 (70.7)
≥75 (Silent Gen)	352/505 (69.7)
Sex	
Female	2890/3749 (77.1)
Male	869/1326 (65.5)
Race/ethnicity	
American Indian or Alaska Native or Native Hawaiian or Pacific Islander	24/31 (77.4)
Asian	87/100 (87)
Black or African American	854/1342 (63.6)
Hispanic or Latino	409/625 (65.4)
White	2130/2586 (82.4)
Other race	81/118 (68.6)
Primary language	
English	3416/4478 (76.3)
Spanish	269/469 (57.4)
Other	29/43 (67.4)
County of residence	
North Caroline urban	1298/1679 (77.3)
North Carolina suburban	1224/1506 (81.3)
North Carolina rural	1237/1890 (65.4)
Median adjusted gross income (US $)	
<25,000	86/161 (53.4)
25,000 to <50,000	2503/3563 (70.2)
50,000 to <75,000	1046/1217 (85.9)
75,000 to <100,000	124/134 (92.5)
Insurance	
Medicare	1325/1851 (71.6)
Commercial	1099/1276 (86.1)
Uninsured	561/859 (65.3)
Medicaid	336/512 (65.6)
State	319/350 (91.1)
Military	113/132 (85.6)
Department of Correction	6/95 (6.3)

MyChart Activation Pre–COVID-19

Using data from complete cases among NC residents (n=4756), we calculated the aORs of MyChart activation by patient characteristics prior to the COVID-19 pandemic (Table 3). Compared to Baby Boomers, Gen Z patients were 5 times more likely to activate MyChart (aOR 5.39, 95% CI 2.67-10.9), followed by Millennials (aOR 2.86, 95% CI 2.22-3.69) and Gen X (aOR 1.72, 95% CI 1.42-2.08) patients. Male patients were significantly less likely to activate MyChart than female patients (aoR 0.61, 95% CI 0.51-0.71; *P*<.001). Compared to White patients, Black or African American patients (aOR 0.39, 95%

CI 0.33-0.47; *P*<.001) and patients of “other” race (aOR 0.44, 95% 0.27-0.70; *P*<.001) had significantly lower odds of MyChart activation. Spanish as a primary language was associated with significantly lower odds of MyChart activation (aOR 0.31, 95% CI 0.20-0.48; *P*<.001) than English. Suburban residence was associated with significantly higher odds of MyChart activation (aOR 1.22, 95% CI 1.00-1.49; *P*=.05), whereas rural residence was associated with significantly lower odds of activation (aOR 0.78, 95% CI 0.65-0.93; *P*=.007) than urban residence. Compared to patients insured through Medicare, there were significantly higher odds of MyChart activation among subjects with commercial insurance (aOR 1.77, 95% CI 1.41-2.23; *P*<.001), state insurance (aOR 2.67, 95% CI 1.76-4.05; *P*<.001), and military insurance (Tricare; aOR 2.20, 95% CI 1.19-4.10; *P*=.05), whereas Medicaid insurance was associated with significantly lower odds of MyChart activation (aOR 0.64, 95% CI 0.49-0.83; *P*<.001). Compared to thte median gross income level of US $25,000 to <US $50,000, median income of <US $25,000 was significantly associated with lower odds of MyChart activation (aOR 0.62, 95% CI 0.42-0.90; *P*=.01), whereas higher income levels were significantly associated with MyChart activation: US $50,000 to <US $75,000 (aOR 1.89, 95% CI 1.53-2.33; *P*<.001) and US $75,000 to <US $100,000 (aOR 3.61, 95% CI 1.74-7.47; *P*<.001). Results from the analysis using multiple imputed data were consistent with these results; thus, the results from complete case analyses are reported.

Table 3. Adjusted odds ratios (aORs) and 95% CI of MyChart activation pre–COVID-19 (n=4756; Model 1)[a].

Characteristic	aOR (95% CI)	*P* value
Age group (years)		
17-24 (Gen Z)	5.39 (2.67-10.9)	<.001
25-39 (Millennials)	2.86 (2.22-3.69)	<.001
40-54 (Gen X)	1.72 (1.42-2.08)	<.001
55-74 (Baby Boomers; ref)	1.00	
≥75 (Silent Gen)	0.8 (0.62-1.03)	.08
Sex		
Female (ref)	1.00	
Male	0.61 (0.51-0.71)	<.001
Race/ethnicity		
American Indian or Alaska Native or Native Hawaiian or Pacific Islander	1.09 (0.44-2.70)	.85
Asian	1.04 (0.52-2.09)	.92
Black or African American	0.39 (0.33-0.47)	<.001
Hispanic or Latino	0.86 (0.56-1.31)	.48
White (ref)	1.00	
Other race	0.44 (0.27-0.70)	<.001
Primary language		
English (ref)	1.00	
Spanish	0.31 (0.20-0.48)	<.001
Other	0.46 (0.21-1.01)	.05
County of residence		
North Carolina urban (ref)	1.00	
North Carolina suburban	1.22 (1.00-1.49)	.05
North Carolina rural	0.78 (0.65-0.93)	.007
Insurance		
Medicare (ref)	1.00	
Commercial	1.77 (1.41-2.23)	<.001
Uninsured	0.85 (0.67-1.08)	.19
Medicaid	0.64 (0.49-0.83)	<.001
State	2.67 (1.76-4.05)	<.001
Military	2.2 (1.19-4.10)	.05
Median income (US $		
<25,000	0.62 (0.42-0.90)	.01
25,000 to <50,000 (ref)	1.00	
50,000 to <75,000	1.89 (1.53-2.33)	<.001
75,000 to <100,000	3.61 (1.74-7.47)	<.001

[a] Model 1 covariables include the most recent visit date, the number of visits, age group, sex, race/ethnicity, primary language, county of residence, insurance, and median income.

MyChart Use Pre–COVID-19 and Post–COVID-19

To determine changes in patient portal use during the COVID-19 pandemic, we calculated the aORs of MyChart usage in the 8 months prior to and the first 8 months following telemedicine adoption (Table 4). We also calculated the odds of becoming a MyChart user during the COVID-19 pandemic among those who were previously nonusers (Table 5).

Some disparities remained despite the rapid and nearly complete transition from in-person to remote care starting in April 2020 (Table 4; Model 2). The associations between MyChart use and sex, race or ethnicity, language, residency rurality, and insurance were similar to those observed with MyChart activation (Table 3; Model 1) and were not significantly different by pre–COVID-19 or post–COVID-19 timing. However, Gen Z patients had higher odds of MyChart use post–COVID-19 (aOR 2.52, 95% CI 1.63-3.89) than pre–COVID-19 (aOR 1.54, 95% CI 0.99-2.39). Interestingly, there was no difference in MyChart use after the pandemic between the highest earners (US $75,000 to <US $100,000) and the reference group (US $25,000 to <US $50,000), perhaps reflecting an increase in MyChart use among the reference group.

Among prior nonusers (n=3086; Table 5; Model 3), we observed that Gen Z was associated with significantly higher odds of becoming a MyChart user during the pandemic (aOR 2.80, 95% CI 1.32-5.94; *P*=.007). Prior male nonusers were less likely to become a MyChart user (aOR 0.58, 95% CI 0.41-0.83), as well as nonusers of rural residence compared to nonusers of urban residence (aOR 0.62, 95% CI 0.44-0.87).

Table 4. Adjusted odds ratios (aORs) and 95% CI of MyChart use pre–COVID-19 and post–COVID-19 (Model 2)[a,b].

Characteristic	July 2019 to March 2020		April 2020 to November 2020	
	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Age group (years)				
17-24 (Gen Z)	1.54 (0.99-2.39)	.05	2.52 (1.63-3.89)	<.001
25-39 (Millennials)	1.60 (1.30-1.97)	<.001	1.51 (1.22-1.86)	<.001
40-54 (Gen X)	1.37 (1.15-1.63)	<.001	1.31 (1.11-1.56)	.002
55-74 (Baby Boomers; ref)	1.00			
≥75 (Silent Gen)	0.80 (0.63-1.01)	.06	0.96 (0.76-1.22)	.76
Sex				
Female (ref)	1.00			
Male	0.81 (0.70-0.93)			.004
Race/ethnicity				
American Indian or Alaska Native or Native Hawaiian or Pacific	0.95 (0.47-1.90)			.89
Asian	0.75 (0.49-1.14)			.17
Black or African American	0.61 (0.52-0.70)			<.001
Hispanic or Latino	1.23 (0.90-1.69)			.19
White (ref)	1.00			
Other race	0.60 (0.40-0.89)			.01
Primary language				
English (ref)	1.00			
Spanish	0.43 (0.29-0.62)	<.001	0.55 (0.38-0.79)	.002
Other	1.26 (0.63-2.54)	.52	1.55 (0.78-3.09)	.21
County of residence				
North Carolina urban (ref)	1.00			
North Carolina suburban	1.11 (0.95-1.29)			.18
North Carolina rural	0.80 (0.69-0.94)			.005
Insurance				
Medicare (ref)	1.00			
Commercial	1.21 (1.02-1.45)			.03
Uninsured	0.67 (0.54-0.83)			<.001
Medicaid	0.73 (0.57-0.92)			.008
State	1.36 (1.05-1.75)			.02
Military	1.10 (0.72-1.68)			.67
Median income (US $)				
<25,000	0.70 (0.46-1.08)	.11	0.96 (0.64-1.44)	.86
25,000 to <50,000 (ref)	1.00			
50,000 to <75,000	1.50 (1.27-1.77)	<.001	1.23 (1.05-1.45)	.01
75,000 to <100,000	1.88 (1.23-2.85)	.003	1.23 (0.82-1.84)	.32

[a] Model additionally controls for last visit date from 2017-19 and the number of visits.

[b] Interaction terms were tested between pre–COVID-19 and post–COVID-19 time frames and all demographic covariables, and interaction terms are used to show effects by pre–COVID-19 or post–COVID-19 time frame if the interaction term was *P*<.05. Otherwise, the overall main effect is shown and is not significantly different by the pre–COVID-19 or post–COVID-19 time frame.

Table 5. Adjusted odds ratios (aORs) and 95% CI of MyChart use among previous nonusers (Model 3; n=3086)[a].

Characteristic	aOR (95% CI)	*P* value
Age group (years)		
17-24 (Gen Z)	2.80 (1.32-5.94)	.007
25-39 (Millennials)	1.17 (0.75-1.81)	.49
40-54 (Gen X)	1.20 (0.85-1.70)	.29
55-74 (Baby Boomers; ref)	1.00	
≥75 (Silent Gen)	0.76 (0.45-1.28)	.30
Sex		
Female (ref)	1.00	
Male	0.58 (0.41-0.83)	.003
Race/ethnicity		
American Indian or Alaska Native or Native Hawaiian or Pacific	1.55 (0.34-6.99)	.57
Asian	0.49 (0.14-1.78)	.28
Black or African American	1.08 (0.79-1.48)	.63
Hispanic or Latino	0.70 (0.31-1.59)	.39
White (ref)	1.00	
Other race	0.50 (0.17-1.48)	.21
Primary language		
English (ref)	1.00	
Spanish	1.41 (0.58-3.43)	.45
Other	1.98 (0.38-10.2)	.41
County of residence		
North Carolina urban (ref)	1.00	
North Carolina suburban	0.97 (0.70-1.36)	.88
North Carolina rural	0.62 (0.44-0.87)	.006
Insurance		
Medicare (ref)	1.00	
Commercial	1.00 (0.67-1.49)	.99
Uninsured	0.76 (0.48-1.20)	.24
Medicaid	0.73 (0.43-1.23)	.23
State	1.29 (0.75-2.24)	.36
Military	0.71 (0.27-1.88)	.49
Median income (US $)		
<25,000	0.84 (0.37-1.88)	.67
25,000 to <50,000 (ref)	1.00	
50,000 to <75,000	0.95 (0.67-1.35)	.78
75,000 to <100,000	0.34 (0.08-1.45)	.14

[a] Model additionally controls for last visit date from 2017-19 and the number of visits.

Discussion

Principal Findings

We found that after the start of telemedicine visits at our institution, there was a significant increase in patient portal use among our youngest patients (Gen Z generation). However, MyChart usage following the implementation of telemedicine remained significantly lower among those of Black or African American race and “other” race, having Spanish as a primary language, being uninsured, and having Medicaid. Male sex and rural residence were also associated with lower odds of MyChart use post–COVID-19, and individuals in these groups were significantly less likely to become MyChart users during the pandemic. To our knowledge, this is the first study describing disparities in patient portal use among patients with RMDs and the first to evaluate changes in health technology use during the current COVID-19 pandemic.

Comparison With Prior Work

The results of our study build upon previous work, underscoring inequities in telehealth use and further highlighting that these disparities existed prior to the COVID-19 pandemic and continue to persist. A large cross-sectional study conducted prior to the COVID-19 pandemic showed that the most frequently cited barriers to patient portal adoption were preference for direct communication with providers and inexperience with computers, which were associated with lower income and older age. In this study, other commonly reported barriers included having no patient portal available or difficulty accessing the portal, lack of internet access, and privacy concerns [17].

Various studies examining disparities in telehealth use during the pandemic found similar results to ours. A study by Pierce and Stevermer [18] showed that those of non-White race and those who resided in rural postal codes had lower rates of telehealth use. Another study also showed that older age, living alone, and rural residence were associated with lower telehealth use [19]. However, these patterns are not restricted to rural populations. A study of New York City residents showed that Black and Hispanic patients, non-English speakers, and older patients were less likely to use telehealth for COVID-19–related care [20].

Some factors that may explain these inequities in telehealth use include limited access to affordable and reliable internet, low computer ownership, and low digital technology literacy among certain groups. It is estimated that 24 million Americans do not have access to affordable high-speed internet, with rural residents being disproportionately affected [21,22]. In a survey of NC residents, the cost and lack of access to broadband were the 2 most cited reasons for not having internet access [23]. Cellular phones offer an alternative method to access the internet and is sometimes the only option for certain individuals. A study of computer or laptop ownership among Americans showed that younger age, non-White race, lower educational attainment, and lower income were associated with "smartphone" dependency and a lack of computer or laptop ownership [24]. However, the ability to access the internet via cellular devices is subject to the availability of reliable cellular data networks.

In our study, we noted that men had lower patient portal activation and use before and during the pandemic. This finding may be explained by men's overall lower engagement with health care and hesitancy toward help-seeking behaviors compared to women [25,26]. Similar to our findings, Yang et al [27] found that Medicaid enrollees were less likely to adopt eHealth tools compared to the non-Medicaid population in part due to lower odds of internet access.

We found that Black or African American individuals were less likely to use the patient portal even after controlling for factors such as age, insurance, residence type, and income. Thus, although there are certainly social and economic factors tied to race that contribute to lower health technology use, there seem to be other elements that influence health technology use among Black or African American individuals. Some of these results may stem from poorer access to reliable internet, privacy concerns, or preference for speaking directly to their health care providers [28,29]. A qualitative study by Lyles et al [30] evaluating barriers to patient portal use among Latinx and Black patients showed that difficulty navigating the patient portal and concern that patient portal use would diminish the patient-provider relationship were 2 major themes observed across age, income, and geographical groups. Spanish and other non-English speakers are less likely to access health care or use telehealth modalities due to difficulty communicating with providers and using health technology platforms not available in their language [31,32]. UNCH MyChart is only available in English, which prevents non–English-speaking patients from using this resource.

Age is also a large factor in telehealth use. A study investigating health technology "readiness" among older adults showed that 41.4% of Medicare beneficiaries lacked access to a computer with high-speed internet access, 40.9% lacked a smartphone with a wireless data plan, and 26.3% lacked either form of digital access [33]. Other difficulties that older adults may face include age-related impairments (eg, hearing loss, vision loss, and dementia) or low overall use and unfamiliarity with using technology, and these barriers seem to be amplified in patients who are male; single; Black or African American or Hispanic or Latino; reside in a nonmetropolitan area; and have less education, lower income, and poorer self-reported health [34].

Strengths and Limitations

The strengths of this study include our large real-world cohort of patients with RMDs and the racial and socioeconomic diversity of our study population. We were also able to compare changes in health technology use pre–COVID-19 and post–COVID-19 in the same patients. Some of the limitations of our current study include our retrospective study design and the use of zip code as a proxy for socioeconomic variables, which risks homogenizing certain populations. Of the 5179 NC residents in our cohort, 104 were not included in the analysis due to lack of IRS income data. Of these individuals, 63% (n=66) were from rural counties, 23% (n=24) from urban counties, and 13% (n=14) from suburban counties. As expected, the majority of these individuals resided in rural counties. However, given the small percentage of the NC cohort that these patients represented (2%), excluding them from the analysis is unlikely to affect our results. We did not specifically assess remote telecommunication visits among our patients in this study and therefore cannot draw conclusions on whether there are disparities in telemedicine use among our patients. Additional factors that we did not include, but are possible confounders, include health literacy, smartphone ownership, computer or laptop ownership, broadband access, and cellular data access.

Conclusions

As technology is increasingly used for health care delivery, addressing disparities in health technology use has never been more important. A recently published perspective piece in the New England Journal of Medicine highlighted the scope of the issue and addressed the newly enacted Infrastructure Investment and Jobs Act. This law includes funding toward broadband infrastructure development and affordability, improving connectivity in rural and tribal communities, the creation of digital literacy programs, and preventing digital discrimination [35]. Although the Infrastructure Investment and Jobs Act does not specifically address digital inequities in health care, the potential changes that may occur as a result of this law will undoubtedly affect accessibility to health technology. Although many of the issues that contribute to inequitable technology access are multifaceted and cannot be changed immediately, we hope that increased research and resources are invested in making health technology accessible for all. Future studies should focus on ongoing barriers and potential solutions in avenues such as internet accessibility, health and digital literacy, and attitudes and perceptions toward health technology.

Acknowledgments

This research was funded by The Fund for Excellence in Lupus & Sjogren’s at the University of North Carolina Thurston Arthritis Research Center.

Editorial assistance was provided by Tessa Englund, PhD, MPH, Research Associate at the University of North Carolina Thurston Arthritis Research Center.

Authors' Contributions

EYS, CA, LFC, and SZS conceptualized, designed, and created the methodology of the study. EYS and CA performed data curation. CA completed the formal analysis and validation. EYS, CA, LFC, and SZS contributed to the original draft preparation, review, and editing. EYS and SZS provided supervision and project administration. SZS acquired the funding. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

SZS has served as a consultant for GlaxoSmithKline, served on advisory boards for Aurinia Pharmaceuticals Inc and AstraZeneca, and received grant funding from Pfizer outside of this project. All other authors declare no other conflicts of interest.

References

1. Kataria S, Ravindran V. Digital health: a new dimension in rheumatology patient care. Rheumatol Int 2018 Nov;38(11):1949-1957. [doi: 10.1007/s00296-018-4037-x] [Medline: 29713795]
2. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
3. Mary A, Boursier A, Desailly Henry I, Grados F, Séjourné A, Salomon S, et al. Mobile phone text messages and effect on treatment adherence in patients taking methotrexate for rheumatoid arthritis: a randomized pilot study. Arthritis Care Res (Hoboken) 2019 Oct;71(10):1344-1352. [doi: 10.1002/acr.23750] [Medline: 30192070]
4. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
5. Arcury TA, Quandt SA, Sandberg JC, Miller DP, Latulipe C, Leng X, et al. Patient portal utilization among ethnically diverse low income older adults: observational study. JMIR Med Inform 2017 Nov 14;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]
6. Bailey SC, O'Conor R, Bojarski EA, Mullen R, Patzer RE, Vicencio D, et al. Literacy disparities in patient access and health-related use of internet and mobile technologies. Health Expect 2015 Dec;18(6):3079-3087 [FREE Full text] [doi: 10.1111/hex.12294] [Medline: 25363660]
7. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among caucasian, filipino, korean, and latino americans: the digital link to health survey. JMIR mHealth uHealth 2014 Oct 22;2(4):e43 [FREE Full text] [doi: 10.2196/mhealth.3710] [Medline: 25339246]
8. Claudel SE, Bertoni AG. Exploring the use of personal technology in type 2 diabetes management among ethnic minority patients: cross-sectional analysis of survey data from the Lifestyle Intervention for the Treatment of Diabetes study (LIFT Diabetes). JMIR Diabetes 2018 Feb 22;3(1):e5 [FREE Full text] [doi: 10.2196/diabetes.8934] [Medline: 30291086]
9. Greenberg AJ, Haney D, Blake KD, Moser RP, Hesse BW. Differences in access to and use of electronic personal health information between rural and urban residents in the United States. J Rural Health 2018 Feb;34 Suppl 1:s30-s38 [FREE Full text] [doi: 10.1111/jrh.12228] [Medline: 28075508]

10. van der Vaart R, Drossaert CHC, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MAFJ. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014 Mar 26;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
11. Vangeepuram N, Mayer V, Fei K, Hanlen-Rosado E, Andrade C, Wright S, et al. Smartphone ownership and perspectives on health apps among a vulnerable population in East Harlem, New York. Mhealth 2018;4:31 [FREE Full text] [doi: 10.21037/mhealth.2018.07.02] [Medline: 30221166]
12. Dimock M. Defining generations: where Millennials end and Generation Z begins. Pew Research Center. 2019 Jan 17. URL: https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/ [accessed 2019-10-03]
13. The North Carolina Rural Center. About Us. NC Rural Center. URL: https://www.ncruralcenter.org/about-us/ [accessed 2019-10-21]
14. SOI tax stats - individual income tax statistics - 2017 ZIP code data (SOI). Internal Revenue Service. URL: https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2017-zip-code-data-soi [accessed 2022-08-26]
15. Allison P. Imputation of categorical variables with PROC MI. 2005 Presented at: Proceedings of the Thirtieth Annual SAS Users Group International Conference; April 10-13, 2005; Philadelphia, PA p. 113-130 URL: https://support.sas.com/resources/papers/proceedings/proceedings/sugi30/113-30.pdf
16. Turkiewicz A, Luta G, Hughes HV, Ranstam J. Statistical mistakes and how to avoid them - lessons learned from the reproducibility crisis. Osteoarthritis Cartilage 2018 Nov;26(11):1409-1411 [FREE Full text] [doi: 10.1016/j.joca.2018.07.017] [Medline: 30096356]
17. Turner K, Clary A, Hong Y, Alishahi Tabriz A, Shea CM. Patient portal barriers and group differences: cross-sectional national survey study. J Med Internet Res 2020 Sep 17;22(9):e18870 [FREE Full text] [doi: 10.2196/18870] [Medline: 32940620]
18. Pierce RP, Stevermer JJ. Disparities in use of telehealth at the onset of the COVID-19 public health emergency. J Telemed Telecare 2020 Oct 21:1357633X20963893 [FREE Full text] [doi: 10.1177/1357633X20963893] [Medline: 33081595]
19. Jaffe DH, Lee L, Huynh S, Haskell TP. Health inequalities in the use of telehealth in the United States in the lens of COVID-19. Popul Health Manag 2020 Oct;23(5):368-377. [doi: 10.1089/pop.2020.0186] [Medline: 32816644]
20. 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]
21. Bridging the digital divide for all Americans. Federal Communications Commission. URL: https://www.fcc.gov/about-fcc/fcc-initiatives/bridging-digital-divide-all-americans [accessed 2020-12-20]
22. Secretary Perdue applauds broadband investment included in Omnibus. United States Department of Agriculture. 2018 Mar 23. URL: https://www.usda.gov/media/press-releases/2018/03/23/secretary-perdue-applauds-broadband-investment-included-omnibus [accessed 2020-12-20]
23. The Homework Gap in North Carolina. North Carolina Department of Information Technology. URL: https://www.ncbroadband.gov/media/59/open [accessed 2020-12-20]
24. Mobile technology and home broadband 2019. Pew Research Center. 2019 Jun 13. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2020-12-26]
25. Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. Am Psychol 2003 Jan;58(1):5-14. [doi: 10.1037/0003-066x.58.1.5] [Medline: 12674814]
26. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract 2000 Feb;49(2):147-152. [Medline: 10718692]
27. Yang X, Yang N, Lewis D, Parton J, Hudnall M. Patterns and influencing factors of eHealth tools adoption among Medicaid and non-Medicaid populations from the Health Information National Trends Survey (HINTS) 2017-2019: questionnaire study. J Med Internet Res 2021 Feb 18;23(2):e25809 [FREE Full text] [doi: 10.2196/25809] [Medline: 33599619]
28. Chunara R, Zhao Y, Chen J, Lawrence K, Testa PA, Nov O, et al. Telemedicine and healthcare disparities: a cohort study in a large healthcare system in New York City during COVID-19. J Am Med Inform Assoc 2021 Jan 15;28(1):33-41 [FREE Full text] [doi: 10.1093/jamia/ocaa217] [Medline: 32866264]
29. Elkefi S, Yu Z, Asan O. Online medical record nonuse among patients: data analysis study of the 2019 Health Information National Trends Survey. J Med Internet Res 2021 Feb 22;23(2):e24767 [FREE Full text] [doi: 10.2196/24767] [Medline: 33616539]
30. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. "I want to keep the personal relationship with my doctor": understanding barriers to portal use among African Americans and Latinos. J Med Internet Res 2016 Oct 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
31. Cashen MS, Dykes P, Gerber B. eHealth technology and internet resources: barriers for vulnerable populations. J Cardiovasc Nurs 2004;19(3):209-14; quiz 215. [doi: 10.1097/00005082-200405000-00010] [Medline: 15191264]
32. Clayman ML, Manganello JA, Viswanath K, Hesse BW, Arora NK. Providing health messages to Hispanics/Latinos: understanding the importance of language, trust in health information sources, and media use. J Health Commun 2010;15 Suppl 3:252-263 [FREE Full text] [doi: 10.1080/10810730.2010.522697] [Medline: 21154097]

33. Roberts ET, Mehrotra A. Assessment of disparities in digital access among Medicare beneficiaries and implications for telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
34. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
35. Rodriguez JA, Shachar C, Bates DW. Digital inclusion as health care - supporting health care equity with digital-infrastructure initiatives. N Engl J Med 2022 Mar 24;386(12):1101-1103. [doi: 10.1056/NEJMp2115646] [Medline: 35302722]

Abbreviations

aOR: adjusted odds ratio
IRS: Internal Revenue Service
NC: North Carolina
RMD: rheumatic and musculoskeletal disease
UNCH: University of North Carolina Hospitals

Edited by C Basch; submitted 16.04.22; peer-reviewed by S Hajesmaeel Gohari, M Hudnall; comments to author 03.08.22; revised version received 17.08.22; accepted 22.08.22; published 31.08.22.

Please cite as:
Sun EY, Alvarez C, Callahan LF, Sheikh SZ
The Disparities in Patient Portal Use Among Patients With Rheumatic and Musculoskeletal Diseases: Retrospective Cross-sectional Study
J Med Internet Res 2022;24(8):e38802
URL: https://www.jmir.org/2022/8/e38802
doi:10.2196/38802
PMID:36001872

Original Paper

The Human Need for Equilibrium: Qualitative Study on the Ingenuity, Technical Competency, and Changing Strategies of People With Dementia Seeking Health Information

Emma Dixon[1], BS, PhD; Jesse Anderson[1], BS; Diana C Blackwelder[1], BS, MS; Mary L Radnofsky[1], BA, MA, PhD; Amanda Lazar[1], BS, PhD

[1] College of Information Studies, University of Maryland, College Park, MD, United States

Corresponding Author:
Emma Dixon, BS, PhD
College of Information Studies
University of Maryland
4130 Campus Dr 4th floor
College Park, MD, 20742
United States
Phone: 1 843 609 7249
Email: eedixon@umd.edu

Abstract

Background: Prior research on health information behaviors of people with dementia has primarily focused on examining the types of information exchanged by people with dementia using various web-based platforms. A previous study investigated the information behaviors of people with dementia within a month of their diagnosis. There is an empirical gap in the literature regarding the evolution of health information needs and behaviors of people with dementia as their condition progresses.

Objective: Our work primarily investigated the information behaviors of people with dementia who have been living with the condition for several (4 to 26) years. We also aimed to identify their motivations for changing their information behaviors over time. Our primary research questions were as follows: how do people with dementia get informed about their condition, and why do people with dementia seek information about their condition?

Methods: We adopted an action research approach by including 2 people with dementia as members of our research team. Collaboratively, we conducted 16 remote 1-hour contextual inquiry sessions with people living with mild to moderate dementia. During the study sessions, the first 40 minutes included semistructured interviews with participants concerning their information behaviors, followed by a 20-minute demonstration of their information-seeking strategies. Data from these interviews were analyzed using a constructivist grounded theory approach.

Results: Participants described their information needs in terms of managing the disrupted physiological, emotional, and social aspects of their lives following a diagnosis of dementia. They used various information behaviors, including active search, ongoing search, monitoring, proxy search, information avoidance, and selective exposure. These information behaviors were not stagnant; however, they were adapted to accommodate the changing circumstances of their dementia and their lives as they worked to re-establish equilibrium to continue to engage in life while living with a degenerative neurological condition.

Conclusions: Our research revealed the motivations, changing abilities, and chosen strategies of people with dementia in their search for information as their condition evolves. This knowledge can be used to develop and improve person-centered information and support services for people with dementia so that they can more easily re-establish equilibrium and continue to engage in life.

(J Med Internet Res 2022;24(8):e35072) doi:10.2196/35072

KEYWORDS

dementia; health information behavior; action research; equilibrium; postdiagnostic experience; mobile phone

Introduction

Background

Information behavior refers to the ways in which people interact (and do not interact) with information [1]. Understanding health information behavior is important as decisions based on health information sources can influence a person's health trajectory, quality of life, and health outcomes [1,2]. Populations with unique health information needs may exhibit different health information behaviors. Therefore, researchers have studied the information behaviors of populations such as people with diabetes [3,4], breast cancer [5], and prostate cancer [6]. We studied the health information needs and behaviors of people with dementia as, similar to those with other health conditions, they want to be fully informed about their diagnosis [7-10]; however, the cognitive nature of their medical condition makes this inherently difficult. Therefore, we must identify the information that they seek, as well as find the most appropriate ways to provide it, to meet their needs.

To date, nearly all dementia-related information behavior research has analyzed data created by people with dementia using social media (eg, Twitter [11,12] and Facebook [13,14]), web-based dementia forums [15-17], and a web-based dementia advocacy platform to share their experiences [18]. These studies revealed the importance of web-based communities for information exchange among people with dementia, although they did not explore the broader information behaviors of people with dementia.

Most research on health information research on dementia has focused on informal caregivers' information behaviors. A recent scoping literature review of 20 studies found 4 studies on the information needs and information-seeking behaviors of people with dementia, with only 2 studies distinguishing the behaviors of people with dementia from those of caregivers [19]. Studies that delineate the unique information needs and information behaviors of people with dementia found that caregivers were more interested in searching for topics specific to dementia, whereas people with dementia predominantly searched for support groups and were generally disinterested in seeking information regarding dementia [20, 21].

Harland et al [10] found that perceptions of dementia were major factors affecting information-seeking behaviors immediately after diagnosis. For example, people with dementia avoided health information if they felt unable to influence their situation, whereas others valued and sought this information to "confirm their suspicions and provide explanations" [10]. Harland et al [10] called for further work to understand "how information needs and behaviors change over time and with the progression of the disease"—a call to which this paper responds.

Objective

The primary aim of this study was to discover the nature, content, and evolution of information behaviors of people living several years after a dementia diagnosis. Our secondary aim was to identify the motivations for changing information behaviors over time. Therefore, we sought to address the following research questions:

1. How do people with dementia stay informed about their condition?
2. Why do people with dementia seek information about their condition?

To answer these research questions, we adopted an action research approach [22], which included contextual inquiry sessions with 16 people with dementia to understand the factors that motivated them to initially seek information. We then examined their evolving information behaviors and investigated their motivations for adapting their information behaviors. In addition, our findings led to revelations beyond the scope of our original interview questions, which unveiled information behaviors in the realms of personal well-being and equilibrium. Equilibrium, defined as a state of balance between a person's mental framework (cognitive and emotional) and the environment [23], is a concept that has been applied in domains ranging from education to health sciences. Specifically, researchers have noted ways in which people strive for equilibrium when coping with a new medical diagnosis [24,25], which triggers disequilibrium, a state of disorder and imbalance in the mind [23]. Past research has noted how individuals with conditions such as cancer [24] and diabetes [25] actively search for information as a way of restoring equilibrium. In our study, several kinds of information behavior, including actively searching for information, played a key role in restoring equilibrium in participants with dementia. We compare our findings to prior research and discuss the implications of our findings for medical professionals who work with people with dementia, as well as health information resource providers such as web-based content developers interested in creating more accessible content.

Methods

Study Overview

Our research team comprised university researchers and 2 dementia advisers (people living with dementia who are knowledgeable and active in dementia advocacy). We selected an action research approach [22] that involved members of the target culture throughout the research process [22]. The 2 dementia advisers on our team served as empowered researchers and decision-makers. The participation of dementia advisers afforded the research team greater access to otherwise hidden truths in the population of people with dementia. The research team collaboratively determined the topic of the study, identified target questions, designed the study protocol, collected and analyzed the data, and wrote the final report.

A contextual inquiry approach [26] allowed us to collect data through participants' verbal explanations, as well as their demonstration of information-seeking behaviors. This strategy gives participants the chance to explain and attempt to perform their intentions, thereby not limiting data collection to recall or verbal descriptions of information behaviors [26,27]. This is particularly useful when conducting research with people with cognitive disabilities. We chose a qualitative approach rather than using a priori theories and methodologies to describe complex, dynamic participant experiences.

Recruitment

We recruited people with dementia through convenience sampling from our networks, which included members of peer support groups, large dementia advocacy organizations, and snowball sampling. To qualify for the study, participants had to self-report a clinical diagnosis of mild to moderate dementia and seek dementia information since being diagnosed. The research team first sent 30 potential participants an initial recruitment email with details outlining the study and the criteria for participation in the study. Of the 30 participants, 16 (53%) responded that they met the participation criteria and were interested and available to participate. Before scheduling each study session, the authors verified the eligibility of the participants by verbally inquiring about the participation criteria. Interview and observation sessions were scheduled and completed between July and September 2020.

Data Collection

The 16 contextual inquiries comprised 2 parts: an interview (about 40 minutes), followed by a 20-minute observation session. Sessions were conducted on Zoom videoconferencing (Zoom Video Communications Inc) because of pandemic-related research restrictions. The first 4 contextual inquiry sessions were conducted by a team of 2 (ED and JA) academic researchers. The remaining 12 contextual inquiry sessions were conducted by blended teams (1 academic researcher [ED teamed with DCB] and 1 dementia adviser [JA teamed with MLR]), with the dementia adviser leading each session.

Ethics Approval

We assumed consent capacity, which aligns with best practices when working with people having mild to moderate dementia [28] and is required by law in many countries [29]. As a precaution, we were attentive to a possible lack of capacity and were prepared to use the University of California Davis protocol [30] to determine whether the person was able to do the following: (1) make a choice to participate, (2) show an understanding of what the study entailed, (3) describe their rationale for participating in the study, and (4) show an appreciation of the potential risks and benefits of the study. This protocol was not used for any participants, as there was no indication that any participant might not have the capacity to consent.

In each contextual inquiry session, participants first provided verbal consent and then completed a short demographics form. In semistructured interviews, we asked questions such as “How do you typically obtain dementia-related information?” and “How would you improve your process of obtaining dementia information?” We subsequently followed up with probing questions to pursue topics raised by the informants (eg, conversations with their physicians). During each observation, the screen-sharing technology allowed participants to demonstrate their information-seeking strategies.

On the basis of our related work involving people with dementia, we created 8 scenarios as prompts for information seeking (see Multimedia Appendix 1 for scenarios), and then invited participants to select and conduct a search while describing their intentions. During this phase, we probed participants to reveal their thought processes using questions such as “Why did you choose this site?” Following each session, participants received US $20 Amazon gift cards as compensation.

Raw data comprised session transcriptions, screen videos of web-based searches, field notes, methodological memos, and Zoom video and audio recordings, which resulted in approximately 16 hours of recorded data. This was transcribed using Otter.ai (AISense, Inc) and reviewed and verified by an academic researcher. All procedures were approved by the University of Maryland institutional review board (approval number 1316631-40).

Data Analysis

We used a constructivist grounded theory approach to analyze the interview data [31] as it required taking into consideration and accounting for our own perspectives as researchers. This was critical, given that our research team included both dementia advisers and academic researchers.

Each of us open coded the transcriptions, creating early codes such as “offline information gathering strategies” and “vetting people/organizations.” We collaboratively discussed our codes and determined emergent themes, such as “information sources and systems,” “information gathering strategies,” and “motivations for seeking information.” Individual codes were merged into collaborative themes, resulting in a code book. The team then collectively edited, clarified, and refined the code book. The resulting code book was applied during the refining process of focused coding using the most significant initial codes to sift through large amounts of data to categorize the data incisively and completely [31].

Each transcript was focus coded twice—once by a dementia adviser and then by an academic researcher—to ensure that we correctly understood what each participant’s intended meaning. Through this process, our analysis led us to understand how individuals use different types of information behaviors, which has been previously defined in the literature (although, to the best of our knowledge, not in the context of dementia).

We followed an iterative process of engaging with the data, comparing codes, performing pattern recognition, and memoing for several months with weekly team meetings to build connections between codes and emergent theories that would explain the nature of participants’ health information–seeking behaviors. This process led to a dementia adviser connecting the concept of “finding equilibrium,” based on the theory of “optimizing re-equilibration” [23], to the motivation that participants described for their information-seeking strategies and their subsequent emergent changes in information behaviors, which we describe further in the results.

When studying the lived experiences of individuals or groups of people, we may (consciously and unconsciously) superimpose our own perspectives, cultures, and experiences onto the analyses [31]. Therefore, we explicitly reveal our beliefs and status as academics or people living with dementia, so that the reader may interpret our findings with a full understanding of our roles [31]. Although based at a research university, the academic researchers have also been informed by dementia activists who advocate for people with dementia and are known to practice self-determination well into the progression of the condition [32-34]. In addition, technology research has informed our practice of supporting people with dementia [35] and other disabling conditions to take an active role in their own health and well-being [36-39]. These perspectives have influenced the ways in which we conduct research on how people with dementia use information.

Participants

A total of 16 participants aged 57 to79 years self-reported at least one type of mild to moderate dementia with a range of 4 to 26 years since their diagnosis before the study (Table 1). Of the 16 participants, 15 (94%) identified ethnically as White and 1 (6%) as Asian. All participants reported employment statuses of retired, retired on disability, or volunteering. Nearly all claimed being “somewhat confident” using technology (Thomas was “not confident,” whereas Lila, Lucy, and Carter were “very confident”). All participants were dementia advocates and members of online peer support dementia groups.

Table 1. Participant demographics.

Pseudonym	Age (years)	Sex	Country of residence	Type of dementia	Years since diagnosis	Education	Technical confidence
Arnold	68	Male	Canada	Vascular dementia	≥10	Some college; no degree	Somewhat confident
Dawson	73	Male	United States	Early onset Alzheimer disease	2 to 5	Some college; no degree	Somewhat confident
Lucy	67	Female	United Kingdom	Early onset Alzheimer disease	5 to 10	Bachelor’s degree	Very confident
Carter	61	Male	United Kingdom	Vascular dementia	5 to 10	Some college; no degree	Very confident
Michael	61	Male	United States	Functional neurological disorder	5 to 10	Some college; no degree	Somewhat confident
Sadie	79	Female	United States	Alzheimer disease	≥20	High school diploma	Somewhat confident
Lila	Range 60-70	Female	Canada	Early onset Alzheimer disease with a Lewy body component	5 to 10	Multiple bachelor’s degrees	Very confident
Carly	62	Female	United States	Frontotemporal dementia	≥10	Master’s degree	Somewhat confident
Gale	71	Female	Australia	Frontotemporal dementia	5 to 10	Master’s degree	Somewhat confident
Eva	57	Female	United States	Lewy body dementia with behavioral disturbances	5 to 10	Bachelor’s degree	Somewhat confident
Levy	61	Male	No answer	Lewy body or Parkinson disease	2 to 5	Some college; no degree	Only a little confident
Velma	61	Female	Canada	Vascular dementia	5 to 10	Bachelor’s degree	Somewhat confident
Thomas	68	Male	United States	Variant of slow-moving Alzheimer disease	5 to 10	High school diploma	Not at all confident
Kevin	79	Male	United States	Frontotemporal dementia	≥10	Master’s degree	Somewhat confident
Toby	61	Male	United States	Early onset Alzheimer disease and Lewy body	5 to 10	Master’s degree	Somewhat confident
Donna	62	Female	Australia	Semantic variant of primary progressive aphasia	≥10	Master’s degree	Somewhat confident

Results

Overview

We discovered the information needs of people with dementia concerning the physiological, emotional, and social aspects of their life. These included the factors that motivated them to seek information on their condition to re-establish equilibrium, both initially after diagnosis and in evolving ways over the years (or decades, in some cases) through active searching, monitoring and ongoing searching, proxy information search, information avoidance, and selective exposure.

The Goal of Health Information Seeking: Re-establishing Equilibrium After Diagnosis

Overview

Disequilibrium is defined as a state of disorder and imbalance in the mind such that individuals cannot assimilate new information into their schema because of its contradictions or inconsistencies with their prior knowledge or experience [23]. In the instances described by the participants, disequilibrium arose with a medical diagnosis of dementia. After hearing that they had an incurable brain condition, participants described feeling sometimes relieved at receiving an actual diagnosis but also still feeling uninformed, confused, worried, sad, upset, or alone in terms of the physiological, emotional, and social aspects of life. As they strived to resolve this disequilibrium, they often did so by seeking relevant health information; however, just as often, they did not find what they needed.

Information Needs Concerning Physiological Aspects of Life

Upon diagnosis, participants usually expected physicians to satisfy their information needs regarding their physiological health. However, as nearly all respondents explained, most of their physicians (and even local dementia organizations) failed to provide adequate information regarding the origin, progression, treatment, and management of dementia, both initially and during the evolution of their condition. Arnold explained, “I wasn’t given any referrals, any information, who to talk to, where to go.” Donna, too, recounted as follows:

> *I was told, “The type of dementia you’ve got, there’s nothing I can do, and no medication available.”...And then, “See you in six months.”*

This lack of information at diagnosis led to a state of disequilibrium for participants, as they were unsure of the physiological factors related to what to expect out of life with their new dementia diagnosis.

Even after realizing that their questions might yield unpleasant answers (or none at all), several respondents persevered in researching the nature of their condition. Gale explains this as follows:

> *When I was first told the diagnosis, I didn’t really know anything...I really wanted to study [frontotemporal dementia], find out what it was, and confirm [to] myself that I thought the symptoms matched.*

Participants asked the following:

> *...medical, mental kinds of stuff...a cure* [Kevin]
>
> *Is this being caused by the [reduced] blood flow?* [Velma]
>
> *Have I taken the proper steps?* [Levy]

By educating themselves about the physiological aspects of dementia, some participants hoped to learn how to mitigate the effects of dementia on their lives as a way of re-establishing equilibrium. Lila explained the following:

> *I’ve come to realize that there are different forms [of dementia], and each form comes with a different set of problems. And if I know the problem, maybe I can avert the long term effects.*

Information Needs Concerning the Emotional Aspects of Life

The emotional toll of the information provided at diagnosis often left the participants hopeless or in an extremely low emotional state. When Velma received a diagnosis of vascular dementia, her physicians told her, “There was nothing that they could do, nothing that they really had for me, and that I needed to get my affairs in order to try to enjoy my window of time.” This led her to an emotional “state of numbness and shock for probably a good six months or so...I couldn’t even figure out what questions I needed to actually ask, because all that kept going through my head was three to eight years,” which is the reported average life expectancy for individuals with vascular dementia. Similarly, Thomas described his diagnosis as “excruciatingly painful to hear,” leading him to “go home, and then you put a blanket over your head, start having some very dark thoughts about...what’s the point of carrying on, if this is all I’ve got?” He added the following:

> *A lot of health practitioners...don’t practice good emotional care. And emotional care is every bit, if not more important, particularly when you're being told you have a fatal disease.*

Participants had previously expected a different trajectory for their lives. Therefore, the diagnosis of dementia was emotionally destabilizing.

Participants searched for encouraging health information to balance the negative news they had received with any positive news—sometimes just to find hope to re-establish equilibrium. However, Thomas realized he was not the best person to do the research:

> *We need...the wisdom of what to do with this new diagnosis as [physicians] give the diagnosis, not send people home and let them go looking for it.*

Disequilibrium upon diagnosis was compounded by failed expectations that physicians would not only fully address his physical and cognitive condition but also his emotional state of being, as it related to the changes he was undergoing.

Other participants experienced not only receiving pessimistic information from their physicians but also a dehumanizing delivery of their diagnosis. Lila’s neurologist “never looked at [her] once.” Instead, the physician announced to her husband, “‘Your wife has early onset Alzheimer's. You can bring her back when she can’t dress herself.’” Lila was furious:

> *I had to keep my mouth shut. I was biting my tongue so hard, I thought the blood was gonna start pouring out of my mouth...*

She later called her primary physician and was referred to another specialist. These kinds of interactions with physicians could last for years, which is explained by Donna as follows:

> *I had a very good rapport with [the doctor], but after the dementia diagnosis, he wouldn’t see me on my own. After about three years, I took a piece of paper with big type, big black font on it: “Talk to Me.”...He said, “What do you mean?” I said, “Well, before dementia, you used to talk to me about my health.*

> *Now you'll only talk to my husband; it's like I don't even exist in the room."*

Such dehumanizing interactions with physicians at diagnosis left at least one participant yearning for moral support and information "that touch the person's heart" (Thomas).

Information Needs Concerning Social Aspects of Life

Participants said that their social lives changed from the moment they shared their diagnoses, reporting an apparent lack of empathy from friends or family members who had no prior experience with dementia. Thomas explained the following:

> *When you finally work up the nerve to tell people what you're going through, the first thing out of their mouth, 95% of the time, isn't empathetic. It's like, "You don't look like you have dementia! At least you don't have cancer"...[A spouse] loves you dearly, but doesn't get the fear that courses through your veins*

Lacking a sense of belonging to a given community also contributed to the participants' state of disequilibrium, driving them to seek information that would help with the social aspects of their lives.

Participants described the need for social connections with other people with dementia and sought further information on their similar lived experiences:

> *I knew that I couldn't be the only one...So I was searching—where are these people? I have to find them. I know they're out there* [Velma]

Connecting to other people with dementia was key to understanding how to live well; knowledge from their lived experiences could not be gleaned from other relationships. Lila described the following:

> *We need to learn from each other. We need to hear how other people do things and what they've gone through.*

Given that many participants could not locate others with dementia nearby, they "had to go online" (Sadie) to accommodate information needs concerning social aspects of life to re-establish equilibrium.

Evolving Information Behaviors of People With Dementia

Overview

Although participants expressed the need for more health information immediately following the diagnosis of dementia, they also described changes in information needs as their medical condition progressed, as new medical discoveries occurred, and as access to different types of information became known. Their initial needs and formerly successful strategies to seek information no longer provided them with equilibrium; therefore, they adapted. At the time of the study, 31% (5/16) of participants had been living with dementia for >10 years, and one participant for even 20 years, and we learned about changes in their information needs and behaviors, along with those of the more recently diagnosed people.

Active Searching

Overview

The first information behavior most participants described following diagnosis reflected the Wilson active search technique, meaning they were "intentionally choosing to browse or search for information" [40]. Participants indicated that the goal of their active searching immediately after diagnosis aided them in recovering from initial postdiagnostic reactions such as depression or denial. Active searching also took place months and years later from digital resources and other people with dementia in online peer support groups.

Digital Resources

As many participants felt they were provided insufficient information at diagnosis, they described actively "searching out every little thing" (Levy) to satisfy their information needs. Carly described herself as "a questioning person...a thinking person," where she tries to clarify the things she's read by "immediately look[ing] [it] up."

Given the multiple types of dementia and their myriad of symptoms, participants had varied physiological information needs, which led them to actively search for web-based resources. Velma, who described herself as "always learning—everyday I'm learning," read "a lot of new reports that come out in JAMA...I'm learning how important things like nutrition, exercise, socialization and connections with people are, in staying well." When experiencing potential symptoms of dementia, Lila searched for "'dementia and left side brain' So, you know, because of the regions in the brain, depending on which side you have dementia attacks and the areas it attacks, it will also tell you what symptoms could be happening to you." Other participants wanted to understand whether symptoms, such as loss of smell and taste (Velma), were related to dementia or to another condition that they might need to treat.

Participants routinely conducted active searches on the internet as they knew that research is "evolving all the time" (Dawson). Even after 12 years of living with dementia, Donna looks up new information "all the time, because everything's changing so much...even stuff about diagnosing categories." Similarly, Dawson acknowledged that "what was in Google out there 10 years ago, 8 years ago, 5 years ago, 10 minutes ago, is always changing..."

For many participants, the sources of information they used were dependent on the type of information they were trying to find. Carly described: "I have many systems—dependent upon the question..." For basic information, Carly would use the Google search engine. When this was "not enough," she would turn to peers with dementia. These peer interactions during the process of actively seeking web-based information were typically synchronous and presented in a mutually beneficial manner, as discussed in the following section.

Counterparts in Peer Support Groups

Without exception, all participants said that they actively sought information about dementia from others with similar diagnoses in online support groups, citing this as one of the most important strategies for gathering useful information.

Several participants described actively seeking physiological information from other people with dementia who had been living with the disease longer, and seemingly better, through web-based synchronous communication using videoconferencing platforms such as Zoom. For example, Eva and Michael both referred to a web-based social group called "Dementia Mentors" [41], which provides a safe, nonjudgmental, stress-free environment several times a week. Through this program, people recently diagnosed with dementia can communicate with people who have lived with the condition for many years via web-based interactions.

The participants also mentioned organizations that offered online peer support groups, such as the Alzheimer's Association, Association for Frontotemporal Dementia, Lewy Body Association, Alzheimer's Society, Dementia Alliance International, and Dementia Australia. After joining a peer support group, Velma "started to have a better understanding of how much there was that we could do to help ourselves." For Toby, the group provided an opportunity to observe other group members, and "almost [see] a case study in what's going to happen to me vicariously," which satisfied his need for information to understand the physiological changes he might experience in the future. Similarly, Dawson explained the following:

> *You can see they've experienced it and you have faith and trust in that individual...if they've had a good experience with a particular medication or some activity that they've done, then it may be something I might want to consider...we share that information.*

Such online peer support groups helped participants identify, and often actually provided, the information necessary to re-establish equilibrium in their lives.

Online peer-to-peer support groups provide a place where participants "share your problems with people who get it...all the different nuances and fears," similar to "a foxhole connection you get when you're in war" (Thomas). Information sharing created a bond that "wasn't the social thing; it was much deeper than that...this connection is heartfelt, emotional, mental and spiritual" (Thomas). This connection helped Thomas to "overcome my own severe negative feelings about people with dementia, and recognize that there's still life worth living, because I was just like everyone else." Arnold expressed how "what we say stays in the group, and we're not pre-judged." Knowing that everyone is similarly vulnerable and that confidentiality is preserved seemed to make peer groups feel safe, even allowing for personal growth. These web-based communities seem to be highly valued resources for re-establishing equilibrium by meeting the information needs concerning the social and emotional aspects of life.

Monitoring and Ongoing Searching

Overview

Participants explained how an active search helped them build "a base of knowledge" (Toby) on dementia and re-establish equilibrium by meeting their physiological, social, and emotional information needs after diagnosis. With these needs met, and equilibrium restored, some shifted to searching less often, such as Gale:

> *In all honesty I tend not to look for anything specific nowadays, because I kind of feel that I've got all the basic information.*

These participants adopted the information behavior of ongoing search, "where active searching has already established the basic framework of knowledge, but where occasional continuing search is carried out to update or expand one's framework" [40]. To facilitate an ongoing search, participants set up what Ellis defines as monitoring strategies, which include any actions that enable someone to stay updated with new developments in their field [42]. For example, Gale explained that when "somebody has prompted me with a subscription or said something, I'd say, 'Oh yeah; I'll look that up and see.'" When new information was inconsistent with the existing knowledge base of the participants, they would slip into disequilibrium, prompting the need to conduct an ongoing search for further information to re-establish equilibrium.

Monitoring strategies included attending peer support groups, curating social media accounts, and receiving push notifications such as subscribing to newsletters, as discussed in the following sections.

Peer Support Groups

Monitoring and ongoing searches in peer support groups of people with dementia were important sources of health information for the participants when "someone else in the group may bring one [research article] forward and go, 'Hey, did you see this one?'" (Velma). Thomas believed that current clinical research "would come up in group, 'cause we have people in our group that are proactive about those kinds of things," alluding to peers who routinely conducted active searches. Arnold, who is in several peer support groups, explained, "I've been caught off-guard with some information...These groups are a tremendous source of information and help."

Lucy occasionally conducted ongoing searches:

> *I don't actually go online unless...a discussion has come up and I want to know more about it...I will go on Google Search and find out...a little bit more about what they're talking about.*

She monitors information in the "nine different groups that I'm involved in" with the goal of "shar[ing] all the information I've collected from others...[because] it's important, isn't it? Because they don't know, and that's something I'm still capable of...getting all this information and letting them know."

Social Media

Participants also used social media platforms such as Twitter to monitor rather than actively search for relevant health information. Arnold explained that "About 95% of the people that I follow are dementia-related," such as researchers and advocacy organizations. He elaborated that he is "not using Twitter to get information; it's more seeing what people in the dementia world are up to," reflecting a monitoring strategy. He "look[s] at everything that's posted. And then I'll click the heart, which means I like it. And sometimes I will make a retweet with a comment. But that's basically it."

Donna checks, "if I see something on Twitter, and there's a lot of chatter about it, I think, 'Oh, I better read that one,'" reflecting an ongoing search strategy. Carter curated a network on Twitter, although for the "vast majority" of people he followed, he had also "met personally...[and] can vouch for their credibility."

Push Notifications and Subscriptions

Many participants set push notifications from subscription providers to monitor the latest dementia information from organizations such as the Dementia Engagement and Empowering Program (Carter), Alzheimer's Disease International (Donna), Lewy Body Dementia Association (Levy), international health organizations (World Health Organization and United Nations [Donna]), medical journals (The Journal of the American Medical Association [Velma] and Neuroscience [Gale]), and medical organizations (the Mayo Clinic [Donna and Velma]). Velma described this strategy as, "just wait for them [newsletters] to show up," as she did not "know how to find a list of what I'm subscribing to or not," but did regularly read her email.

The participants remarked that these monitoring strategies resulted in a large volume of new information. Gale "subscribed to any new research...I probably get three or four emails a day that have information about what's going on with dementia." Donna subscribed to "loads" of blogs that "come into my inbox on a Monday morning...I very rarely go to them in my inbox" as "I just get too many emails, about 600 emails a day across my different emails." However, this offer of information did not usually create disequilibrium. Instead, as Donna explained, with certain research organizations, such as the Mayo Clinic, "I get their newsletter, but I rarely read it unless it really jumps out as something that might be super relevant." Therefore, these monitoring strategies only led to ongoing searches when participants were confronted with new information that was not already part of their existing knowledge base or conflicted with their own experiences.

Proxy Information Search

Proxy information search is when "one individual tries to find information on another person's behalf" [43]. Owing to the progression of dementia and changes in abilities, some participants described having to transition from active or ongoing search to proxy information search to continue having access to the kind of dementia-related information they needed. For example, Toby described decreases in "the speed at which I can assimilate information...and the extent to which I can recall information." Thomas finds actively searching for information difficult as "I don't retain stuff really good anymore." Given that so much information seeking related to dementia took place on the web, active and ongoing search became inaccessible to Carter because of his difficulty with "working on the PC" where "things that I used to be able to do, not that long ago, that I can't do anymore." As a result, the physiological, social, and emotional aspects of life were inadequately addressed, often leaving participants in a state of disequilibrium. To re-establish equilibrium, some participants found a proxy to help them search for dementia information, thus accommodating their progressive cognitive disabilities.

To reiterate, the participants' shift to using the strategy of proxy information search did not appear to be based on a waning interest in dementia information or a lack of questions. Toby explained the following:

> *I generally have the ability to ask the questions and to know the relevant questions and pose them. But I have lost the ability to do some of the deep research that I normally would have done.*

In such instances, several participants referred to their spouses as assuming the role of proxy information seeker by "look[ing] stuff up for me" (Thomas) and being "a conduit for information flow" (Toby). Others relied on mentors (Arnold) and family members (Levy). Thomas also relied on people from his support groups.

The participants expressed much gratitude for the proxy searchers. However, some participants were deterred by the need to transition to greater dependency on information. Levy explained his reluctance to use a proxy but also recognized his own limitations:

> *I don't really go out looking. I used to...when I first got diagnosed...*

However, giving up the power to search for information was a "struggle...because you feel you're going to lose everything. And it's like, now I gotta give this up [searching on his own]. It's killing me" (Levy).

People with dementia often have activities slowly taken away because of their own changing cognitive abilities [32] or from well-intentioned carers to relieve them from certain responsibilities. In either case, the transition to a proxy search may be seen as yet another activity being taken from them. Although participants noted their disappointment and dissatisfaction with having to rely on others to assist with information search, they retained an active role in choosing when to engage in a proxy search.

Some participants described serving as proxy researchers for their peers. When Dawson gave dementia-related talks, people asked him questions about specific types of dementia or how to get tested for dementia. Hence, he would do "a little more digging," then "usually I'll share with them at a very high level what I have been able to learn, but I will also ask them to check with their own doctor or check with an association to get specific[s]." By contrast, Gale did not simply conduct searches for people but rather "show[ed] them how to do it [look up information] and where to look," demonstrating for those capable of looking up dementia information but in need of training on where and how to find it.

Information Avoidance and Selective Exposure

Some participants expressed how, after a certain point, continuing to search for dementia information led to a state of disequilibrium rather than providing a means to reach equilibrium. Michael said, "I'm burned out. I'm burned out on looking at that stuff—I get it. I get it—it is what it is." In response, participants chose information avoidance, or "the human tendency to avoid, ignore, and deny information, particularly in the context of health care" [44], to maintain their equilibrium. When Eva experienced degenerative changes because of dementia, she admitted, "I just let it slide by; I don't care. I don't want to know." In these instances, a participant may experience changes inconsistent with prior knowledge or experiences but still choose not to seek information that would help "go beyond his current state and strike out in new directions" [23]. Instead, some participants described avoiding dementia information or only selectively exposing themselves to dementia information, specifically to maintain their equilibrium.

However, information avoidance did not necessarily signify the end of participants' information seeking but was used more as a mitigation strategy to avoid shifting into disequilibrium because of information overload at a particular moment. Thomas occasionally used proxy search or information avoidance to focus on life and relationships because of the following:

> *What's the point? I mean I know there are some people, they're going to spend the rest of whatever is left of their life worrying about it, stressing over looking stuff up, and I'd rather go play with my dog and go camping and do what I can, while I can, with what time I've got left.*

Information avoidance allowed participants to deflect negative emotions and, instead, as Thomas explained, engage in "trying to be a person who's positive and happy and full of joy. And that's the way I want to live my life; that's the way I want to be known."

The perceived high volume of web-based dementia information was also a contributing factor to some participants' choices to engage in information avoidance. Eva described being overwhelmed and frustrated:

> *You get too much information sometimes, and you just don't want to know any more facts and information.*

She also expressed dissatisfaction with not being able to find answers to questions, sometimes simply as there were no answers. Thomas was more fatalistic:

> *What's the Internet gonna tell me about how you improve your memory?...As far as I know, there's no real way to improve your memory; you either have it or you don't.*

For some, the response to overwhelmingly negative dementia information was not avoiding it all but rather avoiding only emotionally triggering stories, which is a practice called selective exposure, wherein conscious decisions are made to consume only certain information [45]. Lila minimized information related to the stages of dementia to manage her anxiety:

> *Once you're diagnosed...Well, you know, I have dementia, I'm going to decline. I know that...I'm not going to even think about that. You know I have enough to worry about...Why add something else on my plate that could potentially make me more anxious?*

However, she continued to actively search for physiological information on her symptoms. In an observation session, Eva declined to explore the scenario to find information on coping with anxiety caused by dementia, saying, "I don't need to read it. I live it...I live a lot of my life in paranoia and worry and fear." She then chose a less sensitive scenario and demonstrated her technical skills without having to lose her equilibrium because of the potential emotional toll of the information content.

Discussion

Principal Findings

Our findings, based on the semistructured interviews with 16 participants with dementia recruited through convenience sampling, unveiled how the postdiagnostic information needs of people with dementia concerning their physiological, emotional, and social aspects of life motivated participants to actively search for information on their condition. Participants also used the information behaviors of the ongoing search, monitoring, proxy search, information avoidance, and selective exposure as they worked to re-establish equilibrium. These information behaviors were not stagnant but adapted to accommodate the changing circumstances and needs of their lives with dementia to continue to engage in life while living with a degenerative neurological condition.

Comparison With Prior Work

Past research examining how information behaviors change over time has included people with cancer [46] and diabetes [47]. Eheman et al [46] also specifically investigated how information-seeking strategies change from active to passive searching after people receive cancer treatment. Although cancer causes 10 million deaths a year, and diabetes 1.5 million, dementia now affects >55 million people worldwide, with >10 million new annual cases [48]. This fast-growing population merits investigation into its information-seeking behaviors, especially as there are several forms of dementia. with even those in the medical field being often unfamiliar with the dozens of diseases that can cause this condition. The result is that people diagnosed with dementia have had to learn about their complex symptoms for themselves while trying to live fulfilling lives.

Prior research on the information behaviors of people with dementia found that caregivers were more interested in searching for topics specific to dementia. On the other hand, people with dementia were more interested in finding support groups but appeared altogether disinterested in seeking dementia information [20,21]. Although our research validates the importance of support groups (and provides evidence for their role in filling physiological, social, and emotional information needs), our findings refute this claim of disinterest. Our findings are more aligned with the recent Harland et al [10] study, which highlights the interests of some people with dementia to seek information 4 weeks after diagnosis, whereas others avoided it as they felt they could not affect their situation [10]. Similarly, we found that some individuals, such as Velma, practiced information avoidance for up to 6 months after diagnosis. Some participants' justification for information avoidance was because they felt emotionally "numb" (Velma) rather than powerless to affect their situation, as reported in the study by Harland et al [10], and even reported the decisions to engage in various information behaviors, including actively seeking dementia information later in their lives. Because of our semistructured interviews with 16 participants with dementia recruited through convenience sampling, we were able to discover the transitions that many participants underwent in their information behavior.

Even participants who described intentionally avoiding dementia information or only exposing themselves to certain topics many years after diagnosis (because of the emotional toll they anticipated) illustrated conscious choices to maintain equilibrium. Such acts of self-preservation and self-care are perhaps a vital stage of effective dementia information seeking because of the pessimistic, overwhelming, and unsatisfactory extant dementia information. This finding provides a contrasting narrative to the typical assumption that people with dementia are incapable of self-regulation [49]. In fact, participants were able to recognize that searching for dementia information at times led them to an unhealthy state of mind; therefore, they adjusted their actions to avoid disequilibrium.

Thus, information avoidance may be worth exploring in future research, as it may inform physicians and support and advocacy organizations in their efforts to deliver more person-centered, positive dementia-related information. When participants chose to seek dementia information, they developed and used their own systems to monitor advances in research, including regular scrolling through curated social media accounts, which is an avenue for the dissemination of self-narratives by people with dementia [11,12]. Our findings also contribute to the field by demonstrating that some people with dementia use social media in the hopes of hearing about new breakthroughs in medical research regarding dementia. Future studies should consider using social media as an additional platform for disseminating information about support services for individuals with dementia.

Our findings depict that the information behaviors of a group of people with dementia differ from those studied in previous research, which found that people with mild cognitive impairment and dementia underuse social media and push notifications to access health information [50]. Although prior research reported apathy toward technology in the search for web-based dementia resources [51,52], individuals in this study demonstrated a wide range of technology savviness and interest, not only in finding web-based dementia resources but also in setting up monitoring strategies through social media and subscriptions.

We also discovered that participants regularly used web-based platforms to seek out peers with dementia on Twitter [11,12], Facebook [13], dementia advocacy websites [18], and web-based forums [15-17], echoing the importance of web-based communities. Notably, we found that people with dementia seek peers who have lived with the diagnosis longer than they had in synchronous support groups such as "Dementia Mentors" [41]. However, participants also leveraged technology to support others as friends and mentors, with one participant (Gale) teaching newly diagnosed people how to search for dementia information instead of merely serving as a proxy. Such mentors fit the concept of "Dementia Trailblazers" outlined by Johnson et al [17] in 2020—people with dementia who were "extremely active" and knowledgeable in web-based dementia forums in providing information support. Thus, our findings provide evidence of the value of creating support services that incorporate peer mentors to teach people newly diagnosed with dementia how to search for information to meet their physiological, social, and emotional information needs.

Our research on the evolving information behaviors of people with dementia has shown that their search strategies change over time, not only because of the degenerative nature of their condition but more importantly because they remain motivated to re-establish a sense of equilibrium in their lives. Past work has described how people living with cancer [24] and diabetes [25] strive for equilibrium while coping with their diagnosis. It is noteworthy that our study has shown that people with cognitive disabilities are motivated by the same human need for equilibrium in the physiological, social, and emotional aspects of life. To the best of our knowledge, our work is the first to theorize how information behaviors, and the transitioning between them, are motivated by the need to re-establish an equilibrium for people with dementia. Although past work has largely focused on the various motivational factors for information seeking or information avoidance [4,46,47], our findings enrich the health information behavior literature by illustrating a single cross-cutting motivational factor for the range of information behaviors—equilibrium.

Implications

Revelations from this research will serve to inform medical professionals and web-based content developers about the information needs of the 55 million people with dementia worldwide [48], not only to help them regain equilibrium after the initial diagnosis of a serious brain disease but also to continue supporting them throughout the progression of their condition as they undergo changes in physiological, emotional, and social aspects of their lives.

Medical Professionals

Participants described their initial diagnosis of dementia as the trigger to disequilibrium, a finding consistent with prior research that revealed the overwhelmingly negative manner used by physicians to inform patients of their condition, which led to multiple calls for physicians to adjust the language they use to communicate a diagnosis [10,53,54]. Therefore, we add our voices to the call for physicians to convey information in a way that is emotionally reassuring [10] and to "instill hope in the context of a dementia diagnosis" [54].

Left in an information desert after diagnosis, participants were forced to go in search of answers about their type of dementia on the web using only their random technological skills, general education, and access to computers to sift through approximately 1.2 million TB of internet content (as of 2021). Although many participants eventually found health information on the web, it was often not exactly what they needed, took an inordinate amount of time, and was easily lost because of the complexity involved in saving digital information. To combat these challenges, we join in the call of previous work for interprofessional education [55] and multidisciplinary care teams [56] to provide more effective postdiagnosis support to people newly diagnosed with dementia [55]. Our findings demonstrate the necessity of providing such information upon diagnosis in multiple formats and languages (eg, printed, web-based, verbal, pictorial, audio, and video) and when the newly diagnosed individual is emotionally more ready for it. Further work is needed to identify the full range of barriers that people with dementia may encounter when searching for health information on the web.

Web-Based Content Developers

In pursuit of dementia information on the web, participants described using social media platforms, search engines, dementia advocacy websites, medical organizations, journal websites, and online peer support groups, although they encountered challenges with each. To address some of these challenges, we ask content developers to better support the information needs of people with dementia in the following areas.

As many participants wanted to actively, although selectively, search for dementia information (to limit their exposure to certain dementia-related topics), we would suggest improving signposts to the content of pages, so that end users can more easily avoid their personal informational triggers to disequilibrium.

Furthermore, the nuisance of subscriptions was frequently discussed, despite the participants' stated desire to stay informed about health information. A potential solution could be to filter more finely any available web-based content, such as excluding paid advertisements, providing only open-access articles, or showing only reports whose accuracy has been academically verified.

Finally, because of executive function challenges (eg, speed of retrieval, recall, and retention), people with dementia sometimes resorted to proxy searches, which often led to negative emotions around the sense of losing agency. Therefore, we refer to web-based content developers to learn and address the cognitive accessibility needs of neurodiverse end users [57]. These include but are not limited to providing clear and understandable text, using simplified layouts to facilitate the locating of needed information, reducing extraneous material such as advertising and flashing notifications to maintain focus, and offering support for different ways of understanding content [57].

Limitations

In designing a study involving people with dementia, we were cognizant of the need to respect the time commitment and amount of work we should require. Therefore, we limited the scope of data collection to a single, approximately 1-hour session. We realize that a longitudinal study of participants' information-seeking behaviors would yield much more data; thus, we hope that future ethnographic or autoethnographic studies will follow the progression of people with dementia, as well as the evolution of their information behaviors.

Participants in this study were limited in racial, geographic, linguistic, and age diversity because of the nature of our convenience sampling recruitment strategy and access to volunteers. All but 1 participant identified as White, and all resided in the United States, the United Kingdom, Canada, or Australia and spoke English, which does not represent the global population currently living with dementia. Prior research shows a higher prevalence of dementia in the United States in African American and Latinx communities [58]. Therefore, future work is needed to ensure that our understanding of information behavior is more representative.

Given that 50% (8/16) of our participants were aged ≤65 years, we also acknowledge an overrepresentation of people with younger onset dementia [59], which only accounts for 9% of global diagnoses [48], although this number may be an underrepresentation of actual cases. Recruiting this relatively younger group of participants may have resulted from the hesitation of the older generation to reveal their diagnoses because of stigma [32,60,61] and misinformation about dementia, which leads to an unwillingness to discuss personal experiences with researchers [62]. The overrepresentation may also have been because participants were recruited using convenience sampling from the most visible peer support and advocacy groups that regularly expressed an interest in helping dementia researchers. Finally, most participants in this study used technology confidently, which may not be representative of the general population with mild to moderate dementia, although this is trending upward: 54.14% of people with mild cognitive impairment or dementia report using their smartphones and tablets almost daily [63].

Conclusions

Using an action research methodology with 2 dementia advisers and academic researchers, we collaboratively identified the motivations and evolution of the information behaviors exhibited by people with dementia after diagnosis. Participants demonstrated their ingenuity and changing abilities to search for information themselves, shifting their information behaviors (eg, from active to monitoring or proxy searches and from ongoing searches to information avoidance and selective exposure), and discussing such changes to maintain and re-establish their sense of equilibrium in life.

Participants also demonstrated and reflected on their abilities, challenges, and frustrations when seeking dementia information on the web. We found that people with dementia (similar to people with cancer, diabetes, and other conditions that trigger disequilibrium) were motivated to adapt their information behaviors not only to meet their evolving cognitive needs but also to address the physiological, emotional, and social aspects of their lives to re-establish equilibrium when faced with inconsistent or conflicting information, even as their condition progressed.

By informing physicians, support organizations, web-based health resource developers, and other factors that influence information behaviors of people with dementia, we aim to see innovation in communication technologies and media platforms to facilitate their use by a neurodiverse population. In face-to-face interactions, we hope that groups will incorporate peer mentors to teach people newly diagnosed with dementia how and where to search for trustworthy health information on the internet, as well as from their medical providers. We also anticipate that our findings will guide readers newly diagnosed with dementia and who are experiencing disequilibrium to try some of the participants' strategies to search for the web-based health information that they need to re-establish equilibrium in their own lives.

Acknowledgments

We thank the participants, Beth St Jean, and the anonymous reviewers who provided feedback on versions of this paper. This work was supported, in part, by grant 90REGE0008, US Admin for Community Living, NIDILRR, Department of Health and Human Services, NSF grant IIS-2045679, and the National Science Foundation Graduate Research Fellowship Program under grant DGE 1840340. Opinions expressed do not necessarily represent official policy of the federal government.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview and observation guide used to understand information search by people living with dementia.
[PDF File (Adobe PDF File), 195 KB - jmir_v24i8e35072_app1.pdf]

References

1. Understanding Human Information Behavior When, How, and Why People Interact with Information. Lanham, Maryland, United States: Rowman & Littlefield Publishers; 2021.
2. Johnson F, Rowley J, Sbaffi L. Modelling trust formation in health information contexts. J Inf Sci 2015 Mar 25;41(4):415-429. [doi: 10.1177/0165551515577914]
3. Mamykina L, Nakikj D, Elhadad N. Collective sensemaking in online health forums. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15: CHI Conference on Human Factors in Computing Systems; Apr 18 - 23, 2015; Seoul Republic of Korea. [doi: 10.1145/2702123.2702566]
4. St. Jean B. Factors motivating, demotivating, or impeding information seeking and use by people with type 2 diabetes: a call to work toward preventing, identifying, and addressing incognizance. J Assoc Inf Sci Technol 2016 Mar 02;68(2):309-320. [doi: 10.1002/asi.23652]
5. Sillence E. Giving and receiving peer advice in an online breast cancer support group. Cyberpsychol Behav Social Network 2013 Jun;16(6):480-485. [doi: 10.1089/cyber.2013.1512] [Medline: 23659724]
6. Wong F, Stewart DE, Dancey J, Meana M, McAndrews MP, Bunston T, et al. Men with prostate cancer. J Psychosomatic Res 2000 Jul;49(1):13-19. [doi: 10.1016/s0022-3999(99)00109-9]
7. Pinner G, Bouman WP. Attitudes of patients with mild dementia and their carers towards disclosure of the diagnosis. Int Psychogeriatrics 2003 Sep 10;15(3):279-288. [doi: 10.1017/s1041610203009530] [Medline: 14756163]
8. Beattie A, Daker-White G, Gilliard J, Means R. 'How can they tell?' A qualitative study of the views of younger people about their dementia and dementia care services. Health Social Care Community 2004 Jul;12(4):359-368. [doi: 10.1111/j.1365-2524.2004.00505.x] [Medline: 15272891]
9. Harman G, Clare L. Illness representations and lived experience in early-stage dementia. Qual Health Res 2006 Apr 01;16(4):484-502. [doi: 10.1177/1049732306286851] [Medline: 16513992]
10. Harland J, Bath PA, Wainwright A, Seymour J. Making sense of dementia. ASLIB J Inf Manag 2017 May 15;69(3):261-277. [doi: 10.1108/ajim-08-2016-0141]
11. Thomas B. Whose story is it anyway?: following everyday accounts of living with dementia on social media. Style 2017;51(3):357-373. [doi: 10.1353/sty.2017.0030]
12. Talbot CV, O'Dwyer ST, Clare L, Heaton J, Anderson J. How people with dementia use twitter: a qualitative analysis. Comput Human Behav 2020 Jan;102:112-119. [doi: 10.1016/j.chb.2019.08.005]
13. Craig D, Strivens E. Facing the times: a young onset dementia support group: Facebook™ style. Australasian J Ageing 2016 Mar 24;35(1):48-53. [doi: 10.1111/ajag.12264] [Medline: 27010874]
14. Lawless M, Augoustinos M, LeCouteur A. Dementia on Facebook: requesting information and advice about dementia risk-prevention on social media. Discourse Context Media 2018 Oct;25:44-51. [doi: 10.1016/j.dcm.2018.01.011]
15. Rodriquez J. Narrating dementia: self and community in an online forum. Qual Health Res 2013 Sep 01;23(9):1215-1227. [doi: 10.1177/1049732313501725] [Medline: 23907588]
16. Johnson J, Black R, Chen Y, Hayes G. Older adults with dementia in an online forum: a preliminary analysis. In: Proceedings of the Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing.

2019 Presented at: CSCW '19: Computer Supported Cooperative Work and Social Computing; Nov 9 - 13, 2019; Austin TX USA. [doi: 10.1145/3311957.3359477]

17. Johnson J, Black RW, Hayes GR. Roles in the discussion: an analysis of social support in an online forum for people with dementia. Proc ACM Human Comput Interact 2020 Oct 14;4(CSCW2):1-30. [doi: 10.1145/3415198]
18. Lazar A, Dixon EE. Safe enough to share: setting the dementia agenda online. Proc ACM Human Comput Interact 2019 Nov 07;3(CSCW):1-23. [doi: 10.1145/3359187]
19. Soong A, Au ST, Kyaw BM, Theng YL, Tudor Car L. Information needs and information seeking behaviour of people with dementia and their non-professional caregivers: a scoping review. BMC Geriatrics 2020 Feb 14;20(1):61 [FREE Full text] [doi: 10.1186/s12877-020-1454-y] [Medline: 32059648]
20. Edelman P, Kuhn D, Fulton BR, Kyrouac GA. Information and service needs of persons with Alzheimer's disease and their family caregivers living in rural communities. Am J Alzheimer's Disease Other Dementias 2006 Sep 04;21(4):226-233 [FREE Full text] [doi: 10.1177/1533317506290664] [Medline: 16948286]
21. Forbes D, Finkelstein S, Blake C, Gibson M, Morgan D, Markle-Reid M, et al. Knowledge exchange throughout the dementia care journey by Canadian rural community-based health care practitioners, persons with dementia, and their care partners: an interpretive descriptive study. Rural Remote Health 2012 Nov 26:2201. [doi: 10.22605/rrh2201]
22. Hayes GR. The relationship of action research to human-computer interaction. ACM Trans Comput Human Interact 2011 Jul 01;18(3):1-20. [doi: 10.1145/1993060.1993065]
23. Piaget J. The Equilibration of Cognitive Structures The Central Problem of Intellectual Development. Chicago: University of Chicago Press; 1985.
24. Carlick A, Biley FC. Thoughts on the therapeutic use of narrative in the promotion of coping in cancer care. Eur J Cancer Care 2004 Sep;13(4):308-317. [doi: 10.1111/j.1365-2354.2004.00466.x] [Medline: 15305898]
25. Kelly G, Lawrence J, Dodds A. Women’s developmental experiences of living with type 1 diabetes. Forum Qual Social Res 2005;6(1). [doi: 10.17169/fqs-6.1.521]
26. Holtzblatt K, Beyer H. Contextual Design Defining Customer-Centered Systems. Amsterdam, Netherlands: Elsevier Science; 1997.
27. Lazar J, Feng J, Hochheiser H. Research Methods in Human-Computer Interaction. Amsterdam: Elsevier Science; 2017.
28. Hegde S, Ellajosyula R. Capacity issues and decision-making in dementia. Annals Indian Academy Neurol 2016 Oct;19(Suppl 1):S34-S39 [FREE Full text] [doi: 10.4103/0972-2327.192890] [Medline: 27891023]
29. Mental Capacity Act. The Health Research Authority. 2005. URL: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/mental-capacity-act/ [accessed 2021-06-29]
30. Policy and procedures for assessing capacity to consent for research. UC Davis Alzheimer's Disease Center. URL: https://health.ucdavis.edu/clinicaltrials/StudyTools/Documents/ResearchCapacityPolicy7=02.pdf [accessed 2002-07-10]
31. Charmaz K. Constructing Grounded Theory A Practical Guide Through Qualitative Analysis. Thousand Oaks, California: SAGE Publications; 2014.
32. Swaffer K. Dementia and prescribed dis-engagement. Dementia (London) 2015 Jan 11;14(1):3-6. [doi: 10.1177/1471301214548136] [Medline: 25503628]
33. Bryden C. Dancing with Dementia My Story of Living Positively with Dementia. London, United Kingdom: Jessica Kingsley Publishers; 2005.
34. Taylor R. Alzheimer's from the Inside Out. Towson, MD: Health Professions Press; 2007.
35. Dixon E, Piper A, Lazar A. “Taking care of myself as long as I can”: how people with dementia configure self-management systems. 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: 10.1145/3411764.3445225]
36. Nunes F. Designing self-care technologies for everyday life: a practice approach. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing System. 2015 Presented at: CHI '15: CHI Conference on Human Factors in Computing Systems; Apr 18 - 23, 2015; Seoul Republic of Korea. [doi: 10.1145/2702613.2702624]
37. Nunes F. From medicalized to mundane self-care technologies. Interactions 2019 Apr 23;26(3):67-69. [doi: 10.1145/3319374]
38. Buls D, Rooksby J. Technology for self-management of rosacea: a survey and field trial. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA. [doi: 10.1145/3027063.3053080]
39. de Barros A, Cevada J, Bayés, Alcaine S, Mestre B. User-centred design of a mobile self-management solution for Parkinson's disease. In: Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia. 2013 Presented at: MUM '13: 12th International Conference on Mobile and Ubiquitous Multimedia; Dec 2 - 5, 2013; Luleå Sweden. [doi: 10.1145/2541831.2541839]
40. Wilson T. Human information behavior. Inf Sci 2000;3:49-56. [doi: 10.28945/576]
41. Dementia Mentors - Online. DEMENTIA MENTORS. URL: http://www.dementiamentors.org/home.html [accessed 2021-08-03]
42. Ellis D. A behavioural approach to information retrieval system design. J Document 1989;45(3):171-212. [doi: 10.1108/eb026843]

43. McKenzie P. Positioning theory and the negotiation of information needs in a clinical midwifery setting. J Am Society Inf Sci Technol 2004 Jun;55(8):685-694 [FREE Full text] [doi: 10.1002/asi.20002]
44. Case DO, Andrews JE, Johnson JD, Allard SL. Avoiding versus seeking: the relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. J Med Library Assoc 2005 Jul;93(3):353-362 [FREE Full text] [Medline: 16059425]
45. Adamic L, Glance N. The political blogosphere and the 2004 U.S. election: divided they blog. In: Proceedings of the 3rd international workshop on Link discovery. 2005 Presented at: KDD05: The Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 21 - 25, 2005; Chicago Illinois. [doi: 10.1145/1134271.1134277]
46. Eheman CR, Berkowitz Z, Lee J, Mohile S, Purnell J, Rodriguez EM, et al. Information-seeking styles among cancer patients before and after treatment by demographics and use of information sources. J Health Commun 2009 Jul 31;14(5):487-502 [FREE Full text] [doi: 10.1080/10810730903032945] [Medline: 19657928]
47. St Jean BL. Information behavior of people diagnosed with a chronic serious health condition: a longitudinal study. University of Michigan. 2012. URL: https://deepblue.lib.umich.edu/handle/2027.42/91570
48. Dementia. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2022-03-15]
49. Ismail Z, Gatchel J, Bateman DR, Barcelos-Ferreira R, Cantillon M, Jaeger J, et al. Affective and emotional dysregulation as pre-dementia risk markers: exploring the mild behavioral impairment symptoms of depression, anxiety, irritability, and euphoria. Int Psychogeriatrics 2017 Sep 13;30(2):185-196. [doi: 10.1017/s1041610217001880]
50. LaMonica HM, English A, Hickie IB, Ip J, Ireland C, West S, et al. Examining internet and eHealth practices and preferences: survey study of Australian older adults with subjective memory complaints, mild cognitive impairment, or dementia. J Medical Internet Res 2017 Oct 25;19(10):e358 [FREE Full text] [doi: 10.2196/jmir.7981] [Medline: 29070481]
51. Sherman C, Liu CS, Herrmann N, Lanctôt KL. Prevalence, neurobiology, and treatments for apathy in prodromal dementia. Int Psychogeriatrics 2017 Apr 18;30(2):177-184. [doi: 10.1017/s1041610217000527]
52. van der Linde RM, Matthews FE, Dening T, Brayne C. Patterns and persistence of behavioural and psychological symptoms in those with cognitive impairment: the importance of apathy. Int J Geriatric Psychiatry 2017 Mar 27;32(3):306-315. [doi: 10.1002/gps.4464] [Medline: 27017917]
53. Tyrrell J, Genin N, Myslinski M. Freedom of choice and decision-making in health and social care. Dementia 2016 Jul 26;5(4):479-502. [doi: 10.1177/1471301206069915]
54. Aminzadeh F, Byszewski A, Molnar FJ, Eisner M. Emotional impact of dementia diagnosis: exploring persons with dementia and caregivers' perspectives. Aging Mental Health 2007 May;11(3):281-290. [doi: 10.1080/13607860600963695] [Medline: 17558579]
55. Dreier-Wolfgramm A, Michalowsky B, Austrom MG, van der Marck MA, Iliffe S, Alder C, et al. Dementia care management in primary care : current collaborative care models and the case for interprofessional education. Zeitschrift für Gerontologie und Geriatrie 2017 May 31;50(Suppl 2):68-77. [doi: 10.1007/s00391-017-1220-8] [Medline: 28364258]
56. Frost R, Rait G, Aw S, Brunskill G, Wilcock J, Robinson L, PriDem team. Implementing post diagnostic dementia care in primary care: a mixed-methods systematic review. Aging Mental Health 2021 Aug 10;25(8):1381-1394. [doi: 10.1080/13607863.2020.1818182] [Medline: 32911966]
57. Making content usable for people with cognitive and learning disabilities. W3C. URL: https://www.w3.org/TR/coga-usable/#dfn-cognitive-and-learning-disabilities [accessed 2022-03-04]
58. Alzheimer's disease facts and figures. Alzheimer's Association. URL: https://www.alz.org/alzheimers-dementia/facts-figures [accessed 2022-07-19]
59. Younger/early-onset Alzheimer's. Alzheimer's Association. URL: https://alz.org/alzheimers-dementia/what-is-alzheimers/younger-early-onset [accessed 2020-03-16]
60. Behuniak SM. The living dead? The construction of people with Alzheimer's disease as zombies. Ageing Soc 2010 Sep 17;31(1):70-92. [doi: 10.1017/s0144686x10000693]
61. Swaffer K. Dementia: stigma, language, and dementia-friendly. Dementia (London) 2014 Nov 17;13(6):709-716. [doi: 10.1177/1471301214548143] [Medline: 25326220]
62. Slegers K, Wilkinson A, Hendriks N. Active collaboration in healthcare design: participatory design to develop a dementia care app. In: Proceedings of the CHI '13 Extended Abstracts on Human Factors in Computing Systems. 2013 Presented at: CHI '13 Extended Abstracts on Human Factors in Computing Systems; Apr 27-May 2, 2013; Paris France. [doi: 10.1145/2468356.2468440]
63. Guzman-Parra J, Barnestein-Fonseca P, Guerrero-Pertiñez G, Anderberg P, Jimenez-Fernandez L, Valero-Moreno E, et al. Attitudes and use of information and communication technologies in older adults with mild cognitive impairment or early stages of dementia and their caregivers: cross-sectional study. J Medical Internet Res 2020 Jun 01;22(6):e17253 [FREE Full text] [doi: 10.2196/17253] [Medline: 32442136]

Edited by T Leung; submitted 03.12.21; peer-reviewed by B St. Jean, E Rubenstein, S Song; comments to author 23.02.22; revised version received 05.05.22; accepted 27.05.22; published 11.08.22.

Please cite as:
Dixon E, Anderson J, Blackwelder DC, Radnofsky ML, Lazar A
The Human Need for Equilibrium: Qualitative Study on the Ingenuity, Technical Competency, and Changing Strategies of People With Dementia Seeking Health Information
J Med Internet Res 2022;24(8):e35072
URL: https://www.jmir.org/2022/8/e35072
doi:10.2196/35072
PMID:35969426

Original Paper

Web-Based Educational Intervention to Improve Knowledge of Systematic Reviews Among Health Science Professionals: Randomized Controlled Trial

Marina Krnic Martinic[1], MD, PhD; Marta Čivljak[2], MD, PhD; Ana Marušić[3], MD, PhD; Damir Sapunar[4], MD, PhD; Tina Poklepović Peričić[3], DMD, PhD; Ivan Buljan[3], PhD; Ružica Tokalić[3], MD, PhD; Snježana Mališa[2], PhD; Marijana Neuberg[5], PhD; Kata Ivanišević[6], MSc; Diana Aranza[7], MSc; Nataša Skitarelić[8], MD, PhD; Sanja Zoranić[9], PhD; Štefica Mikšić[10], DMD, PhD; Dalibor Čavić[2], MSc; Livia Puljak[2], MD, PhD

[1]University Hospital Split, Split, Croatia
[2]Catholic University of Croatia, Zagreb, Croatia
[3]Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia
[4]University of Split School of Medicine, Split, Croatia
[5]University North, Varazdin, Croatia
[6]Faculty of Health Studies, University of Rijeka, Rijeka, Croatia
[7]University Department of Health Studies, University of Split, Split, Croatia
[8]Department of Health Studies, University of Zadar, Zadar, Croatia
[9]Department of Nursing, University of Dubrovnik, Dubrovnik, Croatia
[10]Faculty of Dental Medicine and Health Osijek, The Josip Juraj Strossmayer University of Osijek, Osijek, Croatia

Corresponding Author:
Livia Puljak, MD, PhD
Catholic University of Croatia
Ilica 242
Zagreb, 10000
Croatia
Phone: 385 13706 600
Email: livia.puljak@gmail.com

Abstract

Background: Lack of knowledge of systematic reviews (SRs) could prevent individual health care professionals from using SRs as a source of information in their clinical practice or discourage them from participating in such research.

Objective: In this randomized controlled trial, we evaluated the effect of a short web-based educational intervention on short-term knowledge of SRs.

Methods: Eligible participants were 871 Master's students of university health sciences studies in Croatia; 589 (67.6%) students who agreed to participate in the trial were randomized using a computer program into 2 groups. Intervention group A (294/589, 49.9%) received a short web-based educational intervention about SR methodology, and intervention group B (295/589, 50.1%) was presented with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist. The participants' knowledge of SRs was assessed before and after the intervention. The participants could not be blinded because of the nature of the intervention. The primary outcome was the difference in the percentage of correct answers about SR methodology per participant between the groups after the intervention, expressed as relative risk and 95% CI.

Results: Results from 162 and 165 participants in the educational intervention and PRISMA checklist groups, respectively, were available for analysis. Most of them (educational intervention group: 130/162, 80.2%; PRISMA checklist group: 131/165, 79.4%) were employed as health care professionals in addition to being health sciences students. After the intervention, the educational intervention group had 23% (relative risk percentage) more correct answers in the postintervention questionnaire than the PRISMA checklist group (relative risk=1.23, 95% CI 1.17-1.29).

Conclusions: A short web-based educational intervention about SRs is an effective tool for short-term improvement of knowledge of SRs among health care studies students, most of whom were also employed as health care professionals. Further studies are needed to explore the long-term effects of the tested education.

Trial Registration: OSF Registries 10.17605/OSF.IO/RYMVC; https://osf.io/rymvc

(J Med Internet Res 2022;24(8):e37000) doi:10.2196/37000

KEYWORDS

educational intervention; systematic review; health science professionals; knowledge; randomized controlled trial

Introduction

Background

Evidence-based medicine (EBM), which is interchangeably also called evidence-based practice (EBP) or evidence-based health care (EBHC) [1], is credited with a major impact on health care [2]. Systematic reviews (SRs) are considered the gold standard evidence that helps in making decisions about health within the concept of EBM [3].

However, multiple studies have shown a low level of knowledge of EBM among health care professionals. Low awareness of EBM was reported by Novak et al [4] among physicians in Croatia, and limited knowledge but a positive attitude toward EBM was reported by Ulvenes et al [5] among Norwegian physicians. A study conducted by Munroe et al [6] showed that only 3% of nurses evaluated their knowledge of EBP as very good.

Knowledge of SRs is considered important for health sciences and medical students as well because it is important that clinicians know how to find and appraise evidence [7]. Knowledge of SRs in trainees can help not only in developing useful skills in critical appraisal but also in addressing important clinical questions and serve as a strong basis to design new, original research studies that will fill the gaps and answer relevant and unsolved clinical questions [8].

The importance of medical students' exposure to EBM was shown by Vrdoljak et al [9], who reported that knowledge and attitudes of mentors toward EBM in general practice can be influenced by using medical students as academic detailers. It has been shown that better knowledge and more positive attitudes toward EBM among medical students are associated with the exposure to the vertical subject on research in biomedicine and activities of The Cochrane Collaboration [10]. Glass et al [11] reported that summarized research evidence delivered in a poster format can increase student nurses' access to the evidence base. This intervention has increased their knowledge to guide their clinical practice. Thus, knowledge of EBM is a variable that can be influenced. A lack of knowledge of SRs and EBM could prevent individual health care professionals from using SRs as a source of information in their clinical practice or discourage them from participating in such research. Several studies have shown the effectiveness of educational programs on changing the beliefs on and attitudes toward EBM of health care professionals and their readiness to use evidence from EBM sources such as the Cochrane Library or SRs to solve clinical problems [12-15].

Web-based educational interventions are low-cost, easy to implement, easily refined and stored for later use, and easily accessible by health care professionals. Educational interventions conducted via the internet related to various topics in medicine have been shown to be effective [16,17]. Several studies have also proved the effectiveness of web-based educational interventions among health care professionals on knowledge of EBP [12-14].

A 2017 Campbell SR on the effectiveness of e-learning in improving knowledge of EBHC showed that, compared with no learning, pure e-learning improved knowledge of and skills regarding EBHC but not attitudes and behaviors [18]. Varnell et al [12] showed that an accelerated 8-week training program influenced a statistically significant positive change in beliefs on and attitudes toward EBP. A controlled trial examining the effect of an educational intervention on knowledge of EBM among physicians in Israel [14] reported a significant improvement in the level of knowledge of and attitudes toward EBM but not a significant impact on clinical practice [14].

Objectives

We were not able to find studies evaluating the effectiveness of educational interventions dedicated to learning about SRs and SR methodology. In this randomized controlled trial (RCT), we evaluated the effect of a short web-based educational intervention about SRs on short-term knowledge of SR among students of health sciences studies in Croatia.

Methods

Ethics Approval

The study protocol was approved by the Ethics Committee of the Catholic University of Croatia on March 1, 2021 (Klasa: 641-03/21-01/03; Urbroj: 498-03-02-06-02/1-21-02). Subsequently, the ethics committees of all participating institutions also approved the study protocol. The participants provided written informed consent to take part in the study.

Guidelines for Reporting

The manuscript was reported in line with the CONSORT (Consolidated Standards of Reporting Trials) checklist [19]. The CONSORT checklist for this manuscript is available in Multimedia Appendix 1. The educational intervention was reported in line with the Guideline for Reporting Evidence-based Practice Educational Interventions and Teaching (GREET) checklist [20].

Trial Registration

The study protocol was prospectively registered (ie, before enrolling the first participants) on the Open Science Framework website [21]. There were no differences between the protocol and the conducted trial.

Study Design

We conducted an RCT with 2 parallel groups and 1:1 participant allocation.

Participants

Inclusion Criteria

The participants were students of Master's university health sciences studies in Croatia. The study programs available at the participating universities were Nursing, Radiological Technology, Clinical Nutrition, Physiotherapy, and other programs. Full-time and part-time students were eligible to take part in the study. Many of these students were already employed in health care; students were eligible for participation regardless of their employment status.

Institutions

There were 8 eligible institutions in Croatia for this study, and we invited all of them. The following 7 institutions accepted the invitation to participate: Catholic University of Croatia; University Department of Health Studies Split; University Department of Health Studies Zadar; University of Dubrovnik, Nursing Studies; University North, Faculty of Dental Medicine and Health; University of Osijek, Faculty of Health Studies; and University of Rijeka. One institution declined the invitation to participate in the study (University of Zagreb School of Medicine).

Contacting the Students

Students from eligible institutions were contacted via email by coauthors (MC, MN, KI, DA, NS, SZ, and SM) employed in these institutions and invited to participate in the study on brief web-based education about SRs of the literature. Students who agreed to participate were randomized by simple randomization using the Randomizer website. After randomization, they were sent an email invitation to access the web-based platform on which materials for participants from the educational intervention and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist groups were available.

The text of the email provided information about the study and provisions related to the anonymity of the participants according to the General Data Protection Regulation, and students were invited to click on the link to further participate in the study. For this study, 2 separate interfaces for the participants were created on the SurveyMonkey platform (Momentive Inc). One interface was created for participants in the educational intervention group and the other for participants enrolled in the PRISMA checklist group. Each group accessed their interface using a separate link.

The link in the email took the participants to their respective SurveyMonkey web-based interface. The text of the email to the participants is presented in Multimedia Appendix 2. In the SurveyMonkey interface, the participants were initially asked to confirm that they voluntarily took part in the research and that they were providing informed consent to participate in the study by entering the next page.

Intervention Group A

In the web-based interface, intervention group A received a newly developed intervention created by the authors of this study with expertise in medical and health sciences education and research methodology. The educational intervention was written in the Croatian language. It consisted of 11 short educational texts on the methodology of producing SRs. A module describing the forest and funnel plots contained figures of those 2 graphs. The content of the educational intervention was an abbreviated version of the information contained in Cochrane's educational materials for web-based learning about SRs of the literature (Cochrane Interactive Learning). The complete content of the educational intervention, translated into English, is presented in Multimedia Appendix 3.

The learning objectives of the educational intervention anticipated that the participants would be able to define EBM, recognize different levels of evidence, define an SR, ask a clinical question, define the steps for preparing and registering an SR protocol, describe literature search and screening, explain the risk-of-bias assessment, and describe the process of data analysis and interpretation in SRs. In addition to theory, there was a practical learning objective: the participants were expected to be able to differentiate between an abstract of an SR and of a narrative literature review.

The first version of the educational intervention was iteratively revised within the team. Before conducting this trial, the web-based interface with the educational intervention was evaluated in a qualitative study among health care workers via semistructured interviews (Krnic Martinic et al, unpublished data, November 2021). The results of the users' feedback obtained in the qualitative study were used to revise the educational intervention.

The intervention was delivered as an asynchronous web-based education that did not include any components of live education or interaction.

The participants were able to go back and forth through the web-based interface with the educational modules and respective questions without a time limit.

Intervention Group B

Intervention group B was presented with the PRISMA checklist [22] for reporting on SRs (Multimedia Appendix 4) in their web-based interface, and the participants were asked to read it. It was presented to the participants in 11 separate sections to be as similar as possible to the number and form of the educational texts in intervention group A.

Pre- and Postintervention Questionnaires

Both groups completed a preintervention questionnaire containing questions about demographic characteristics and their knowledge of SRs before the presentation of the intervention (educational intervention or PRISMA checklist group; Multimedia Appendix 5). We were unable to find questionnaires on this topic and purpose in the literature. Thus, we designed the pre- and postintervention questionnaires specifically for this study. The pre- and postintervention questionnaires were not validated. Questions evaluating knowledge of SRs were based on the questions used in our previous studies on knowledge of SRs [23] and the definitions of SRs [24].

At the end of the educational intervention or PRISMA checklist presentation, the participants were asked to answer the postintervention questionnaire (Multimedia Appendix 5). The questionnaire contained the same questions on knowledge of SRs as in the preintervention questionnaire as well as questions about whether they agreed with the proposed characteristics of the definition of SRs. Finally, they were presented with 4 abstracts of published articles and asked to assess whether they were abstracts of SRs.

As part of the postintervention questionnaire, the participants were asked to express the level of their agreement on whether an SR should have 6 characteristics proposed earlier by Krnic Martinic et al [24]. They were asked to express their agreement with a number on a Likert scale ranging from 1 to 5 that best suited their opinion, where 1 meant *I do not agree at all* and 5 meant *I completely agree* (Multimedia Appendix 5).

After those questions, the participants were presented with 4 abstracts of published scientific articles, of which 2 (50%) were abstracts of SRs [25,26] and the other 2 (50%) were abstracts of narrative reviews of the literature [27,28]. They were chosen based on a nonstructured literature search of SRs where we tried to find SR abstracts that were simple to understand and appropriate for the target population. The abstracts did not contain any mention of the study design used. If the abstract reported that it was an abstract of an SR or if a systematic search was mentioned, that part of the abstract was removed. The participants were asked to assess whether the abstracts were abstracts of SRs. The 4 abstracts used for this assessment are presented in Multimedia Appendix 6 [25-28].

On the last page of the interface in both intervention groups A and B, the participants were invited to optionally leave their first and last name and email address if they wanted to receive a certificate of participation in the educational intervention. The certificate was prepared by Cochrane Croatia.

The entire questionnaire we administered to the participants was a survey and not a psychological instrument. Thus, we did not perform any psychometric calculations. For the 6 before-and-after questions about the opinion regarding SRs, we calculated that, at the first measurement (before the intervention), reliability was .89, expressed using Cronbach α.

Outcomes

The primary outcome was the difference in the percentage of correct answers per participant in the postintervention questionnaire between intervention groups A and B.

Secondary outcomes were the difference in the percentage of correct answers per participant in the pre- and postintervention questionnaires for the intervention group, the proportion of participants who correctly recognized an abstract describing an SR of the literature (percentage), and the proportion of participants who correctly recognized an abstract describing a simple narrative review of the literature (percentage).

Participant Timeline

After we obtained permission from the ethics committees, the participants were invited to take part. After collecting the names of students who agreed to participate and randomizing them, the invitation to participate in the study containing the link to the intervention A or intervention B interface was sent on June 7, 2021. The links were inactivated on June 20, 2021. The knowledge assessment was conducted immediately after the intervention.

Sample Size

The expected effect size was a difference of at least 20% for the primary outcome between intervention groups A and B. The calculation of the sample size to compare the proportions, predefining an α of .05 and β of .20, assuming a difference of at least 20% for the primary outcome between intervention groups A and B, determined that a sample size of 182 participants (91 participants per group) would be required. To compensate for the possible loss of participants after the beginning of the survey (incomplete answers) or the possibility that participants who initially agreed to take part might eventually choose not to take part, the plan was to include at least 20% more participants than calculated as necessary (n=218).

Encouraging the Inclusion of Participants (Recruitment)

After the initial email was sent to the participants with the link to their respective study arm, 3 more reminders were sent to the participants 4 days apart.

Randomization of Participants

The participants were randomized by simple randomization using the Randomizer website.

Allocation Concealment

After randomization, the participants were allocated to the study arms using a randomization sequence by a third person who was not included in other parts of the study.

Blinding

Blinding of Participants and Personnel

The intervention was of such a nature that the participants could not be blinded.

Blinding of Outcome Assessors

Only the first author (MKM) and the principal investigator (LP) had access to the complete raw data set generated by SurveyMonkey, which included the names and email addresses of the participants who wanted the certificate. MKM removed the participants' names and email addresses before the outcome assessor (IB) analyzed the data; thus, anonymized data were analyzed.

Data Management

One author downloaded Microsoft Excel worksheets from SurveyMonkey, which were anonymized in case any participant left a name and email address to obtain the certificate. The SurveyMonkey interface was configured not to collect any information about the participants, including IP addresses. The data were stored on a secure server until the time of analysis.

Statistical Analysis

To determine the normality of the variables' distribution, we used the Kolmogorov-Smirnov test. Categorical data were presented as frequencies and percentages, and numerical values were presented as medians with IQR for variables not following normal distribution and as arithmetic means with IQR for variables following normal distribution. Differences between intervention groups A and B for categorical variables were tested using the chi-square test. To express the difference between groups, numerical values were tested with 2-tailed *t* tests for independent samples (for variables following normal distribution) and Mann-Whitney tests (for variables not following normal distribution). Pre- and postintervention differences were evaluated using the chi-square test for categorical variables and the *t* test for independent samples for numerical variables. The effect size for the primary outcome (the difference between the percentage of correct answers between groups in the postintervention questionnaire) was expressed using relative risk (RR) and 95% CI, as was the difference between the number of correct answers in the pre- and postintervention questionnaires in both groups. The effect size for the secondary outcome was expressed using odds ratio with 95% CI.

We assessed the participants' opinions before and after using parametric procedures on Likert-type scales, which are usually analyzed using a nonparametric test. This was done because, after the initial analysis where we used nonparametric statistics, the results were not interpretable. When we presented results using median and 95% CIs, the results were similar in both groups, although there were significant differences after the intervention. Therefore, we proceeded with parametric testing, which gave the same results but was more precise as it enabled us to interpret the direction of the difference clearly.

All analyses were performed using the computer program JASP (version 0.14.1.0; JASP Team). Statistical significance was set at *P*<.05.

Results

Participant Flow

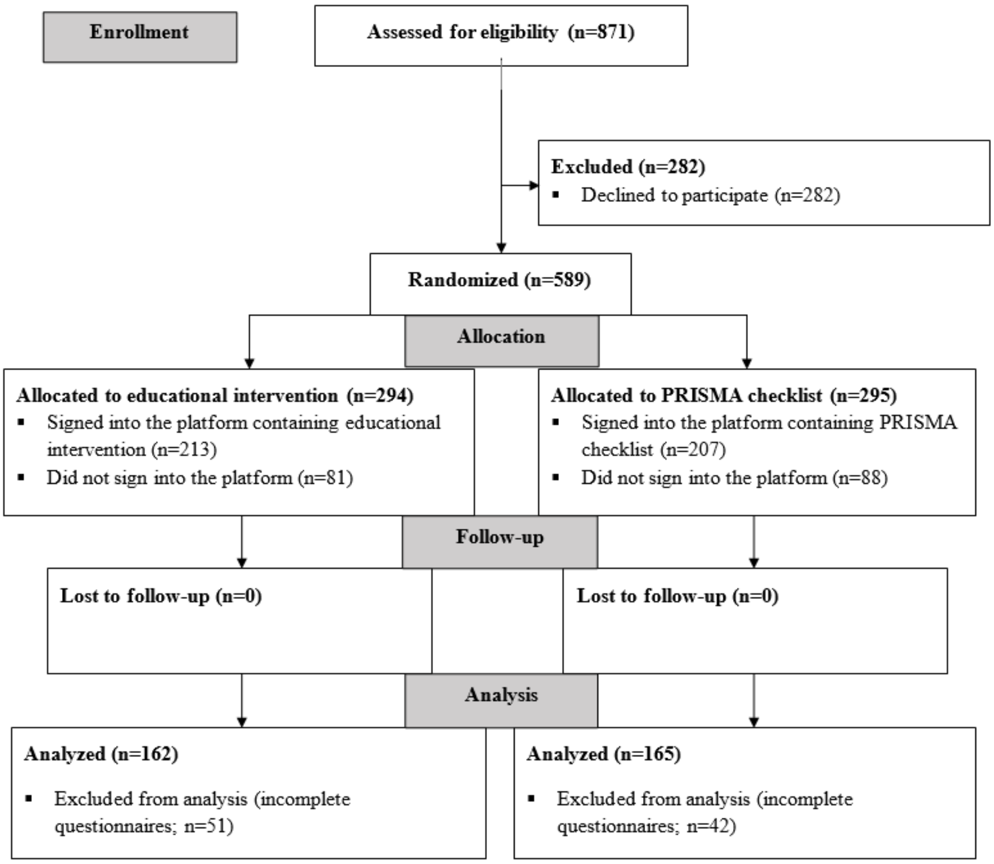
In this trial, 871 potential participants met the inclusion criteria, of whom 282 (32.4%) indicated that they did not want to participate in the study. Thus, 67.6% (589/871) of students were randomized: 31.1% (183/589) from University North; 23.1% (136/589) from the Catholic University of Croatia; 22.4% (132/589) from the Faculty of Health Studies, University of Rijeka; 14.3% (84/589) from the University Department of Health Studies Split; 6.6% (39/589) from the University Department of Health Studies Zadar; 1.7% (10/589) from the University of Dubrovnik, Nursing Studies; and 0.8% (5/589) from the Faculty of Dental Medicine and Health, University of Osijek.

Recruitment and Access to the Educational Platform

The link to participate in the study was sent via email to the addresses of the 589 students on June 7, 2021. The students were sent 3 reminders 4 days apart, and access to the web-based platforms was inactivated on June 20, 2021. A detailed participant flow diagram is shown in Figure 1.

The average time the participants took to complete the entire interface with questionnaires and educational materials was 21 (SD 9.00) minutes in the educational intervention group and 19 (SD 3.96) minutes in the PRISMA checklist group.

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



Baseline Participant Characteristics

The demographic participant data are presented in Table 1. More than 40% of the participants (educational intervention group: 66/162, 40.7%; PRISMA checklist group: 64/165, 38.8%) were from 1 institution (University North), >80% of the participants (educational intervention group: 134/162, 82.7%; PRISMA checklist group: 138/165, 83.6%) studied nursing, and >50% of the participants (educational intervention group: 97/162, 59.9%; PRISMA checklist group: 85/165, 51.5%) attended the second year of study. More than 80% of the participants (educational intervention group: 136/162, 84%; PRISMA checklist group: 138/165, 83.6%) were employed while studying for their Master's degree. Most participants were employed as health care workers (educational intervention group: 130/162, 80.2%; PRISMA checklist group: 131/165, 79.4%). The median length of working in health care was 9.9 years among participants who received the educational intervention and 9.8 years in the PRISMA checklist group. The median age of the participants in both groups was approximately 30 years, and >85% of the participants in both groups were women (educational intervention group: 140/162, 86.4%; PRISMA checklist group: 146/165, 88.5%; Table 1).

Participants in both groups rated their knowledge of SRs with a median grade of 3 (range 1-5). All participants (327/327, 100%) stated that they had heard of SRs, and approximately three-quarters of the participants in both groups (educational intervention group: 124/162, 76.5%; PRISMA checklist group: 123/165, 74.5%) stated that they had read an SR. In our sample, 17.3% (28/162) of the participants from the group that received the educational intervention and 18.2% (30/165) of the participants from the PRISMA checklist group stated that they had participated in producing an SR (Table 1).

Table 1. Demographic characteristics of the participants included in the analysis (N=327).

Variable and level	Educational intervention (n=162)	PRISMA[a] checklist (n=165)
Institution, n (%)		
Faculty of Dental Medicine and Health Osijek	3 (1.9)	2 (1.2)
Faculty of Health Studies, University of Rijeka	26 (16)	30 (18.2)
Croatian Catholic University	46 (28.4)	41 (24.8)
Health Department, University of Zadar	18 (11.1)	20 (12.1)
Health Studies, University of Dubrovnik	0 (0)	3 (1.8)
Health Studies, University of Split	3 (1.9)	5 (3)
University North	66 (40.7)	64 (38.8)
Study program, n (%)		
Physiotherapy	17 (10.5)	15 (9.1)
Clinical Nutrition	5 (3.1)	8 (4.8)
Radiological Technology	2 (1.2)	1 (0.6)
Nursing	134 (82.7)	138 (83.6)
Something else	4 (2.5)	3 (1.8)
Year of study, n (%)		
First	59 (36.4)	74 (44.8)
Second	97 (59.9)	85 (51.5)
Third	6 (3.7)	6 (3.6)
Currently employed (yes), n (%)	136 (84)	138 (83.6)
Currently employed as a health care worker (yes), n (%)	130 (80.2)	131 (79.4)
Length of work status (years), median (IQR)	7 (3-15)	6 (2-16)
Age (years), median (IQR)	28 (24-35)	26 (24-34)
Women, n (%)	140 (86.4)	146 (88.5)
Self-assessment of knowledge of EBM[b] (1-5), median (IQR)	3 (3-4)	3 (3-4)
Had heard about systematic reviews, n (%)	162 (100)	165 (100)
Had read a systematic review, n (%)	124 (76.5)	123 (74.5)
Had participated in writing a systematic review, n (%)	28 (17.3)	30 (18.2)

[a] PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

[b] EBM: evidence-based medicine.

Numbers Analyzed

Of the 420 participants who accessed the interface, 327 (77.9% response rate) completed the questionnaires, and their results were included in further analysis (Figure 1).

Owing to incomplete questionnaires, we excluded the results of 23.9% (51/213) of participants from the educational intervention group and 20.3% (42/207) of participants from the PRISMA checklist group. The results of 54.9% (162/295) of participants from the educational intervention group and 56.1% (165/294) of participants from the PRISMA checklist group were finally included in the analysis. There were no transfers of participants from one group to another.

Primary Outcome

In the postintervention questionnaire, of the 1458 potential correct answers, there were 1086 (74.49%) correct answers to knowledge questions in the educational intervention group (162/327, 49.5%). In the PRISMA checklist group (165/327, 50.5%), of the 1485 potential correct answers, there were 900 (60.61%) correct answers (Table 2). Thus, the effect size for the difference in the number of correct answers to knowledge questions between groups was an RR of 1.23 (95% CI 1.17-1.29); that is, the educational intervention group had 23% (relative risk percentage) more correct answers in the postintervention questionnaire than the PRISMA checklist group.

Table 2. Knowledge of systematic reviews (SRs) among participants who completed pre- and postintervention assessments (N=327).

Questionnaire and items	Educational intervention (n=162)	PRISMA[a] checklist (n=165)	*P* value[b]
Preintervention questionnaire (correct answer)			
It is sufficient to search one database to produce an SR (no), n (%)	128 (79)	139 (84.2)	.22
SRs must be produced by one author only (no), n (%)	103 (63.6)	98 (59.4)	.44
SRs must contain meta-analyses (no), n (%)	17 (10.5)	21 (12.7)	.52
SRs must have duplicate screening and data extraction (yes), n (%)	87 (53.7)	83 (50.3)	.54
A list of both included and excluded studies must be provided (yes), n (%)	117 (72.2)	116 (70.3)	.70
The quality of the included studies must be assessed (yes), n (%)	135 (83.3)	143 (86.7)	.40
In the case of meta-analyses, a heterogeneity test must be done to ensure the results of the studies can be combined (yes), n (%)	126 (77.8)	126 (76.4)	.76
Results of meta-analyses must be presented as a funnel plot (no), n (%)	31 (19.1)	13 (7.9)	.003
Results of publication bias analysis must be presented as a forest plot (no), n (%)	31 (19.1)	27 (16.4)	.51
Total correct answer scores, mean (95% CI)	4.8 (4.5-5.0)	4.6 (4.4-4.9)	.44
Postintervention questionnaire (correct answer)			
It is sufficient to search one database to produce an SR (no), n (%)	156 (96.3)	120 (72.7)[c]	<.001
SRs must be produced by one author only (no), n (%)	153 (94.4)	126 (76.4)	<.001
SRs must contain meta-analyses (no), n (%)	38 (23.5)[c]	33 (20)[c]	.45
SRs must have duplicate screening and data extraction (yes), n (%)	144 (88.9)	111 (67.3)[c]	<.001
A list of both included and excluded studies must be provided (yes), n (%)	144 (88.9)	141 (85.5)[c]	.35
The quality of the included studies must be assessed (yes), n (%)	155 (95.7)[c]	142 (86.1)	.003
In the case of meta-analyses, a heterogeneity test must be done to ensure the results of the studies can be combined (yes), n (%)	150 (92.6)[c]	146 (88.5)[c]	.20
Results of meta-analyses must be presented as a funnel plot (no), n (%)	80 (49.4)[c]	44 (26.7)[c]	<.001
Results of publication bias analysis must be presented as a forest plot (no), n (%)	66 (40.7)[c]	37 (22.4)[c]	<.001
Total correct answer scores, mean (95% CI)	6.7 (6.5-6.9)[c]	5.5 (5.3-5.7)[c]	<.001

[a] PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

[b] Comparison between educational intervention and PRISMA checklist groups. Chi-square test was used for categorical variables, and 2-tailed *t* test was used for independent samples for numeric variables.

[c] Comparison before and after the intervention. Chi-square test was used for categorical variables, and 2-tailed *t* test was used for dependent samples for numeric variables.

Secondary Outcomes

Difference in the Number of Correct Answers per Participant in the Pre- and Postintervention Questionnaires for the Educational Intervention Group

Both groups performed better on the postintervention questionnaire than on the preintervention questionnaire (Table 2). In the educational intervention group, the total number of correct answers was 53.16% (775/1458) in the preintervention questionnaire and 74.49% (1086/1458) in the postintervention questionnaire (RR=1.40, 95% CI 1.32-1.48; Table 2). In the PRISMA checklist group, the total number of correct answers was 51.58% (766/1485) in the preintervention questionnaire and 60.61% (900/1485) in the postintervention questionnaire (RR=1.17, 95% CI 1.10-1.25; Table 2).

Independent of the group, in the pre- and postintervention questionnaires, the smallest number of correct answers was to questions related to the concept of meta-analysis, whereas, in both groups, the highest number of correct answers was to the question about the necessity to assess the quality of research included in the SR (Table 2).

There was no difference in the overall results of the questionnaire assessing knowledge of SRs (the exact number of answers to all 9 knowledge questions) between the educational intervention and PRISMA checklist groups before the intervention (Table 2).

Proportion of Participants Who Correctly Recognized SR Abstracts

The first 2 presented summaries were identified accurately as summaries of SRs by 65.4% (106/162) and 74.1% (120/162) of participants from the educational intervention group and 71.5% (118/165) and 72.7% (120/165) of participants in the PRISMA checklist group, respectively (Table 3). There was no statistically significant difference between the groups in the ability to correctly detect an SR summary (Table 3).

The third and fourth summaries were recognized as a summary of a simple narrative review by 22.2% (36/162) and 46.3% (75/162) of participants from the educational intervention group and 34.5% (57/165) and 47.9% (79/165) of participants from the PRISMA checklist group, respectively (Table 3). There was no statistically significant difference between the groups in the recognition of summaries of narrative reviews (Table 3).

Table 3. Comparison of answers to questions on sources of information needed to answer a clinical question and recognition of a systematic review (SR) abstract (N=327).

Variable and level	Educational intervention (n=162), n (%)	PRISMA[a] checklist (n=165), n (%)	*P* value[b]
If you needed to search for information to solve a clinical problem, what would be the preferred information source for you?			
Colleagues	65 (40.1)	65 (39.4)	.98
Books	59 (36.4)	61 (37)	.98
Scientific literature	120 (74.1)	132 (80)	.25
SR of the literature	135 (83.3)	136 (82.4)	.93
Internet search engine (Google)	29 (17.9)	30 (18.2)	.98
Is this an SR abstract?			
Abstract 1[c]—correct answer “Yes”	106 (65.4)	118 (71.5)	.34
Abstract 2[d]—correct answer “Yes”	120 (74.1)	120 (72.7)	.56
Abstract 3[e]—correct answer “No”	36 (22.2)	57 (34.5)	.02
Abstract 4[f]—correct answer “No”	75 (46.3)	79 (47.9)	.78

[a] PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

[b] Chi-square test.

[c] A total of 7 answers missing.

[d] A total of 7 answers missing.

[e] A total of 8 answers missing.

[f] A total of 10 answers missing.

Additional Analyses

There was no statistical difference in the choice of information sources between the educational intervention and PRISMA checklist groups in the postintervention questionnaire, with multiple possible responses about where the participants would look for answers to a clinical question from their own clinical practice (Table 3). More than 80% of the participants in both groups (educational intervention group: 135/162, 83.3%; PRISMA checklist group: 136/165, 82.4%) stated that they would look for answers in an SR. Most participants (educational intervention group: 120/162, 74.1%; PRISMA checklist group: 132/165, 80%) responded that they would look for answers in scientific literature in general (Table 3). A third of the participants in both groups would look for an answer to a clinical question in a textbook (educational intervention group: 59/162, 36.4%; PRISMA checklist group: 61/165, 37%) or ask a coworker for an answer (educational intervention group: 65/162, 40.1%; PRISMA checklist group: 65/165, 39.4%). Less than one-fifth of the participants in both groups would search for an answer on an internet search engine such as Google (educational intervention group: 29/162, 17.9%; PRISMA checklist group: 30/165, 18.2%; Table 3).

In the preintervention assessment in both groups of participants, there was no significant difference in agreement with the proposed characteristics of an SR (Table 4). After the intervention, there was more agreement with these characteristics in the educational intervention group than in the PRISMA checklist group (Table 4).

Table 4. Responses regarding the characteristics of a systematic review in the pre- and postintervention questionnaires (N=327)[a].

Questionnaire and items	Educational intervention (n=162), mean (95% CI)	PRISMA[b] checklist (n=165), mean (95% CI)	*P* value[c]
Preintervention questionnaire			
Research question is defined	4.5 (4.4-4.7)	4.4 (4.2-4.5)	.09
Listed sources of literature searched, with repeatable search strategy (naming of databases, naming of search platforms, search date, and complete search strategy)	4.4 (4.2-4.5)	4.2 (4.0-4.3)	.08
Listed criteria for inclusion and exclusion of research	4.5 (4.3-4.6)	4.3 (4.1-4.4)	.04
Listed selection methods	4.4 (4.3-4.6)	4.3 (4.2-4.5)	.45
Critically evaluates and reports on the quality or risk of bias of the included studies	4.4 (4.2-4.5)	4.1 (4.0-4.3)	.04
Provides information on data analysis and synthesis that allows for the repeatability of the results	4.4 (4.2-4.5)	4.2 (4.1-4.4)	.11
Postintervention questionnaire			
Research question is defined	4.8 (4.7-4.9)	4.6 (4.5-4.7)	<.001
Listed sources of literature searched, with repeatable search strategy (naming of databases, naming of search platforms, search date, and complete search strategy)	4.7 (4.6-4.8)	4.6 (4.5-4.7)	.05
Listed criteria for inclusion and exclusion of research	4.8 (4.7-4.9)	4.5 (4.4-4.6)	<.001
Listed selection methods	4.8 (4.7-4.9)	4.6 (4.5-4.7)	.02
Critically evaluates and reports on the quality or risk of bias of the included studies	4.7 (4.6-4.8)	4.5 (4.4-4.6)	<.001
Provides information on data analysis and synthesis that allows for the repeatability of the results	4.7 (4.6-4.8)	4.5 (4.3-4.6)	<.001

[a] All differences before and after were statistically significant at *P*<.05; 2-tailed *t* test for paired samples.

[b] PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

[c] *t* test (2-tailed) for independent samples.

Discussion

Principal Findings

This RCT demonstrated that a brief educational intervention conducted on the web about SRs significantly increased knowledge of SRs in the target population. To the best of our knowledge, this is the first trial conducted for this purpose. Relatively successful learning models about EBM have been reported in the literature [15,29-31], but we could not find any publications on the effectiveness of educational interventions focused exclusively on knowledge of SRs.

Comparison With Prior Work

The participants from the educational intervention group (162/327, 49.5%), who were presented with a new educational intervention designed for this study, needed an average of 21 minutes to go through the entire interface. The interface included multiple sections beyond educational intervention: pre- and postintervention questionnaires and the evaluation of 4 scientific abstracts. However, the web-based platform used for this study did not allow for the measurement of the time spent on specific items or pages in the interface. Thus, we cannot know how long the participants read the educational texts prepared for the educational intervention and PRISMA checklist groups. However, if we consider the time to read and answer the questions, the participants probably needed 15 minutes or less to read the educational intervention itself. Such an intervention is very short. Therefore, the intervention should be suitable for health professionals who usually state that their lack of time is a major obstacle to practicing EBM [32-34] and implementing the EBM curriculum during education [35].

Initially, participants in both groups rated their knowledge of SRs with a median grade of 3 out of 5. This is comparable with the self-assessed knowledge of EBP among nurses evaluated in the study conducted by Munroe et al [6]. In that study, only 3% of the nurses said that they were very familiar with EBP [6].

Three-quarters of the participants (educational intervention group: 124/162, 76.5%; PRISMA checklist group: 123/165, 74.5%) stated that they had read SRs. We were surprised with the result that 17.3% (28/162) and 18.2% (30/165) of the participants in the educational intervention and PRISMA checklist groups, respectively, stated that they had participated in developing a SR, which is a high percentage [6]. SR methodology is very complex. Thus, it is questionable whether the students have actually participated in the development of SRs in such large numbers. Health students may have participated, for example, in translating Cochrane's plain language summaries into Croatian [36,37]. However, without the possibility of further clarifying what the participants really meant, it is not possible to discuss this topic in further detail. In the study by Olsson et al [38], which focused on nursing PhD programs and candidatures, in the analyzed 135 nursing dissertations made according to the Scandinavian model of integrated research, only 5 published SRs were found (ie, only 4% of the included nurses—dissertation authors—participated in developing an SR). This number is much lower than the percentage of our participants who stated that they had participated in the production of an SR, and our students were not PhD students but Master's-level students.

The primary outcome of this study was the difference in the percentage of correct answers collected from the educational intervention and PRISMA checklist groups when answering questions evaluating knowledge of SRs on the postintervention questionnaire after the participants had read the educational materials. After the training, the educational intervention group had 23% more correct answers than the PRISMA checklist group (ie, the size of the effect expressed in RR was 1.23). In addition, comparing the pre- and postintervention questionnaire results in the educational intervention group, there were significantly more correct answers on the postintervention questionnaire than on the preintervention questionnaire, with an RR of 1.40. The RR of correct answers comparing the pre- and postintervention questionnaires in the PRISMA checklist group was 1.17.

An RCT by Sánchez-Mendiola et al [15] showed a significant effect of EBM education on the final knowledge of EBM among medical students, with a 25.9% increase in correct answers in the knowledge test about EBM [15]. This is comparable with our primary outcome results. However, it should be emphasized that their intervention was very different in terms of content and duration. Sánchez-Mendiola et al [15] tested an EBM course with 14 two-hour weekly sessions during 1 semester. The course was a formal part of the medical school curriculum; it was delivered by 6 experienced professors and included different content compared with ours. Their course covered 15 topics, including clinical decision-making, uncertainty and probability in medicine, the Bayes theorem, and clinical guidelines [15].

Rohwer et al [18] evaluated the effectiveness of e-learning in improving EBHC in a Campbell SR. The study included 24 trials, of which 20 were RCTs and 4 were observational studies, with a total of 3825 participants including physicians, nurses, physiotherapists, physician assistants, and educators at all levels of education. It demonstrated that, compared with nonlearning, pure e-learning improved EBHC knowledge and skills with similar outcomes to face-to-face learning for any observed primary outcome.

In 2021, needs assessments and expectations regarding EBP knowledge acquisition and training activities were explored among frontline health care providers, including postgraduate medical and nursing students who were working or living in China. The results indicated that the respondents expressed a high need for education on evidence quality appraisal, interpretation of SRs or meta-analyses, and knowledge translation [39]. However, it may not be sufficient to only strive for the improvement of knowledge among the targeted individuals. Nursing education at the undergraduate level is starting to teach the process of research integration through EBP implementation with active learning strategies, which is endorsed by the students [40].

To advance the knowledge and application of evidence in daily practice, ultimately, health institutions will also need to recognize the need to foster such topics [41].

Our educational intervention, implemented via a web tool, is particularly suitable in the current time of the COVID-19 pandemic. Owing to containment measures, many parts of the world have switched to web-based education during the COVID-19 pandemic. Bond et al [42] published a living systematic mapping review on August 30, 2021, calling the web-based teaching experience during the pandemic the "first global online semester." Although such teaching was initially seen as a distance learning response to emergency remote teaching [43], the educational experience gained during a pandemic is very valuable for evaluating the distance learning experience. Our study provides a further test of a remotely delivered educational intervention targeting students and health care workers.

Many studies have evaluated experiences with virtual continuing medical education during the pandemic [44-48]. The SR education evaluated in our study could be incorporated into continuing medical education programs for health care professionals. Owing to the short format and the possibility of distance learning, such education could be of interest to health care professionals who want to learn more about the basics of SRs.

In addition to showing the efficacy of our newly designed educational intervention, our study also indicated areas where the target group of participants significantly lacked knowledge. Independent of the group, in the pre- and postintervention questionnaires, the smallest number of correct answers was to questions related to the concept of meta-analysis and questions about graphical representations of meta-analyses (funnel plot and forest plot). Very modest improvements were observed in those questions in the postintervention questionnaire. In a study on knowledge of the basic methodological components of SRs conducted by Puljak and Sapunar [23] among the directors of postgraduate programs at European universities, only 31% of the participants answered correctly that an SR does not necessarily contain a meta-analysis.

There were few correct answers to questions about graphical presentations of meta-analyses. In the educational intervention group, before the training, 20% of the participants correctly answered what a funnel plot and a forest plot represented. In the PRISMA checklist group, only 8% of the participants correctly answered what a funnel plot represented, and 16% correctly answered what a forest plot depicted. In the postintervention questionnaire, in the educational intervention group, 40.7% (66/162) to 49.4% (80/162) of the participants correctly answered the question about the use of the forest plot or funnel plot, whereas, in the PRISMA checklist group, only a fifth (37/165, 22.4%) to a quarter (44/165, 26.7%) of the participants correctly answered these questions. However, after the educational intervention, more than half of the participants did not know the correct answer to the questions regarding the graphical representations of meta-analyses.

Poor knowledge of graphical representations in meta-analyses has been described elsewhere. A survey conducted on psychologists in Italy found that less than a fifth of psychologists estimated that they had sufficient knowledge of the forest plot [49]. Less than 15% of psychologists stated that they had sufficient knowledge of the funnel plot [49]. A survey conducted on psychologists in Spain showed that only approximately 10% of the participants said that they had satisfactory knowledge of the forest plot. Only 7% of the participants said that they had satisfactory knowledge of the funnel plot [50]. Only 10% of PhD program directors accurately recognized the purpose of funnel plots, and 11.3% recognized the purpose of the forest plot [23]. Poor knowledge of graphical representations of meta-analyses may be the best indicator of generally poor knowledge of SR and meta-analysis methodology.

In this study, we also included a practical knowledge test that involved the recognition of journal abstracts of SRs after the intervention. The accuracy of journal abstract recognition was 22% to 72% in the educational intervention group and 46% to 74% in the PRISMA checklist group, without a statistically significant difference between the groups. This was the final test of understanding and pragmatic application of the knowledge acquired in our trial. We found that the educational intervention did not significantly affect the recognition of abstracts of SRs. It is possible that a time lag or longer systematic learning is needed for the acquired knowledge of SRs to influence the practical application of the knowledge itself. It is also possible that it is necessary to further adjust the educational intervention to enable the practical application of the acquired knowledge.

In the postintervention questionnaire, the participants were asked which sources of information they would use in searching for an answer to a question from their clinical practice, and most participants from both groups opted for scientific literature (educational intervention group: 120/162, 74.1%; PRISMA checklist group: 132/165, 80%) or SRs (educational intervention group: 135/162, 83.3%; PRISMA checklist group: 136/165, 82.4%). Compared with the results of a study conducted by Sánchez-Mendiola et al [15] on medical students in Mexico, in our study, a significantly higher percentage of students chose to search for answers in SRs and scientific literature. In the study by Sánchez-Mendiola et al [15], most participants from the group that attended EBM classes stated that, in solving a certain health problem, they looked for answers in review articles or the Cochrane Library only occasionally, whereas very often they would look for the answer to a health problem in textbooks or search engines or they would ask their teachers. Nevertheless, the number of students who would seek answers in the Cochrane Library or scientific articles was higher than in groups of students who did not attend classes on EBM [15]. Evidently, education about EBM—or, in our case, about SRs—has the express intention to use these data sources more often in solving clinical problems.

Strengths and Limitations

The strength of this study is the appropriate sample size and a high number of fully completed questionnaires, with three-quarters of the participants (327/420, 77.9%) completing the questionnaire in full and almost equal numbers of unfinished questionnaires in the educational intervention and PRISMA checklist groups, allowing for comparable results. Furthermore, we tested the practical application of the acquired knowledge of SRs by asking the students to recognize summaries of SRs or narrative reviews.

A limitation of this study is a highly homogeneous sample that does not allow for significant analyses by sociodemographic subgroups. We acknowledge that there is a potential self-selection aspect in our final sample. We do not have data about nonparticipants among the eligible students and, theoretically, there could be some differences between responders and nonresponders. However, this is an inherent problem of any trial—the eligible participants are invited to take part, and they can choose whether they want to participate.

The study was conducted in only 1 country but in multiple institutions across the country. We did not measure the time spent in the intervention; some participants may not have spent much time reading the text. Furthermore, we measured the outcomes immediately after the reading of the educational texts. Such short-term follow-up does not allow for monitoring of the long-term retention of knowledge of SRs among the participants. Longer-term research will make it possible to verify the long-term effectiveness of the intervention on the knowledge of the target group.

Finally, we would like to note that, in this manuscript, when referring to the studies of other researchers, we used the terms EBM, EBP, and EBHC as they were reported in those manuscripts.

Conclusions

A short web-based educational intervention about SRs is an effective tool for short-term improvement of knowledge of SRs among health care studies students, most of whom were employed as health care professionals. This education can be further studied, modified, and used in the continuing medical education of health care professionals.

Acknowledgments

This study was conducted as a part of the Professionalism in Health: Decision-making in Practice and Science project (IP-2019-04-4882) funded by the Croatian Science Foundation, whose principal investigator is Professor Ana Marušić, MD, PhD. This study was part of the PhD thesis of the first author, MKM. The thesis was written and defended in the Croatian language at the University of Split School of Medicine in Croatia. The authors are grateful to the thesis committee members, professors Ozren Polasek, Zarko Alfirevic, and Damir Roje, for their constructive feedback on this study.

Data Availability

The raw data from this study were published on the Open Science Framework website [21] and made publicly available.

Authors' Contributions

MKM, MC, AM, DS, TPP, IB, RT, and LP contributed to study design. MKM, IB, RT, S Malisa, MN, KI, DA, NS, SZ, S Miksic, DC, and LP contributed to data collection and analysis. MKM, IB, DC, and LP wrote the first draft of the manuscript. MKM, MC, AM, DS, TPP, IB, RT, S Malisa, MN, KI, DA, NS, SZ, S Miksic, DC, and LP critically revised the manuscript. MKM, MC, AM, DS, TPP, IB, RT, S Malisa, MN, KI, DA, NS, SZ, S Miksic, DC, and LP approved the final version of the manuscript.

Conflicts of Interest

Multiple authors of this study are members of Cochrane Croatia (AM, TPP, IB, RT, and LP), but this was not an official research project of the global Cochrane organization.

Editorial Notice

This randomized study was retrospectively registered. The authors published the protocol for the trial [21] before it commenced instead of registering it with a trial registry. As the trial does not involve a study of the cause-and-effect relationship between a health-related intervention and a health outcome, it does not meet the definition of a clinical trial. The editor granted an exception from ICMJE (International Committee of Medical Journal Editors) rules mandating prospective registration of randomized trials because the risk of bias appears low and because the publication of a protocol is equivalent to trial registration and meets all of the purposes listed by the ICMJE regarding transparency.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 87 KB - jmir_v24i8e37000_app1.pdf]

Multimedia Appendix 2
Text of the first email and reminders to invite participants to a randomized controlled trial.
[DOCX File , 14 KB - jmir_v24i8e37000_app2.docx]

Multimedia Appendix 3
Educational intervention.
[DOCX File , 218 KB - jmir_v24i8e37000_app3.docx]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 checklist (divided into 11 sections for the purpose of the study).
[DOCX File , 17 KB - jmir_v24i8e37000_app4.docx]

Multimedia Appendix 5
The text of the pre- and postintervention questionnaires.
[DOCX File , 19 KB - jmir_v24i8e37000_app5.docx]

Multimedia Appendix 6
Four abstracts selected for assessment.
[DOCX File , 19 KB - jmir_v24i8e37000_app6.docx]

References

1. Puljak L. The difference between evidence-based medicine, evidence-based (clinical) practice, and evidence-based health care. J Clin Epidemiol 2022 Feb;142:311-312. [doi: 10.1016/j.jclinepi.2021.11.015] [Medline: 34780980]
2. Seshia SS, Young GB. The evidence-based medicine paradigm: where are we 20 years later? Part 1. Can J Neurol Sci 2013 Jul 23;40(4):465-474. [doi: 10.1017/s0317167100014542] [Medline: 23786728]
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72. [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
4. Novak K, Mirić D, Jurin A, Vukojević K, Aljinović J, Carić A, et al. Awareness and use of evidence-based medicine databases and Cochrane Library among physicians in Croatia. Croat Med J 2010 Apr;51(2):157-164 [FREE Full text] [doi: 10.3325/cmj.2010.51.157] [Medline: 20401959]

5. Ulvenes LV, Aasland O, Nylenna M, Kristiansen IS. Norwegian physicians' knowledge of and opinions about evidence-based medicine: cross-sectional study. PLoS One 2009 Nov 13;4(11):e7828 [FREE Full text] [doi: 10.1371/journal.pone.0007828] [Medline: 19915708]
6. Munroe D, Duffy P, Fisher C. Nurse knowledge, skills, and attitudes related to evidence-based practice: before and after organizational supports. Medsurg Nurs 2008 Feb;17(1):55-60. [Medline: 18429543]
7. Usher-Smith J, Harrison H, Dennison R, Kelly S, Jones R, Kuhn I, et al. Developing training and opportunities for students in systematic reviews. Med Educ 2021 Nov;55(11):1333-1334 [FREE Full text] [doi: 10.1111/medu.14649] [Medline: 34467543]
8. Fares M, Alahdab F, Alsaied T. Systematic reviews and meta-analyses for cardiology fellows. Congenit Heart Dis 2016 Jul 27;11(4):369-371. [doi: 10.1111/chd.12377] [Medline: 27232196]
9. Vrdoljak D, Petric D, Diminić Lisica I, Kranjčević K, Došen Janković S, Delija I, et al. Knowledge and attitudes towards evidence-based medicine of mentors in general practice can be influenced by using medical students as academic detailers. Eur J Gen Pract 2015 Sep 04;21(3):170-175. [doi: 10.3109/13814788.2015.1046124] [Medline: 26108154]
10. Balajić K, Barac-Latas V, Drenjancević I, Ostojić M, Fabijanić D, Puljak L. Influence of a vertical subject on research in biomedicine and activities of The Cochrane Collaboration branch on medical students' knowledge and attitudes toward evidence-based medicine. Croat Med J 2012 Aug;53(4):367-373 [FREE Full text] [doi: 10.3325/cmj.2012.53.367] [Medline: 22911530]
11. Glass GF, Tan H, Chan E. Effect of an evidence-based poster on the knowledge of delirium and its prevention in student nurses: a quasi-experimental study. Contemp Nurse 2021 Dec 09;57(6):462-471. [doi: 10.1080/10376178.2022.2044872] [Medline: 35184674]
12. Varnell G, Haas B, Duke G, Hudson K. Effect of an educational intervention on attitudes toward and implementation of evidence-based practice. Worldviews Evid Based Nurs 2008;5(4):172-181. [doi: 10.1111/j.1741-6787.2008.00124.x] [Medline: 19076918]
13. Stevenson K, Lewis M, Hay E. Do physiotherapists' attitudes towards evidence-based practice change as a result of an evidence-based educational programme? J Eval Clin Pract 2004 May;10(2):207-217. [doi: 10.1111/j.1365-2753.2003.00479.x] [Medline: 15189387]
14. Shuval K, Berkovits E, Netzer D, Hekselman I, Linn S, Brezis M, et al. Evaluating the impact of an evidence-based medicine educational intervention on primary care doctors' attitudes, knowledge and clinical behaviour: a controlled trial and before and after study. J Eval Clin Pract 2007 Aug;13(4):581-598. [doi: 10.1111/j.1365-2753.2007.00859.x] [Medline: 17683300]
15. Sánchez-Mendiola M, Kieffer-Escobar LF, Marín-Beltrán S, Downing SM, Schwartz A. Teaching of evidence-based medicine to medical students in Mexico: a randomized controlled trial. BMC Med Educ 2012 Nov 06;12:107 [FREE Full text] [doi: 10.1186/1472-6920-12-107] [Medline: 23131115]
16. Ngim CF, Ibrahim H, Abdullah N, Lai NM, Tan RK, Ng CS, et al. A web-based educational intervention module to improve knowledge and attitudes towards thalassaemia prevention in Malaysian young adults. Med J Malaysia 2019 Jun;74(3):219-225 [FREE Full text] [Medline: 31256177]
17. Vinokur AD, Merion RM, Couper MP, Jones EG, Dong Y. Educational web-based intervention for high school students to increase knowledge and promote positive attitudes toward organ donation. Health Educ Behav 2006 Dec 31;33(6):773-786. [doi: 10.1177/1090198106288596] [Medline: 16923836]
18. Rohwer A, Motaze N, Rehfuess E, Young T. E ‐ learning of evidence ‐ based health care (EBHC) to increase EBHC competencies in healthcare professionals: a systematic review. Campbell Syst Rev 2017 Mar 02;13(1):1-147 [FREE Full text] [doi: 10.4073/csr.2017.4]
19. Cuschieri S. The CONSORT statement. Saudi J Anaesth 2019 Apr;13(Suppl 1):S27-S30 [FREE Full text] [doi: 10.4103/sja.SJA_559_18] [Medline: 30930716]
20. Phillips AC, Lewis LK, McEvoy MP, Galipeau J, Glasziou P, Moher D, et al. Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET). BMC Med Educ 2016 Sep 06;16(1):237 [FREE Full text] [doi: 10.1186/s12909-016-0759-1] [Medline: 27599967]
21. Martinić MK, Puljak L, Čivljak M, Marusic A, Sapunar D, Peričić TP, et al. Online educational intervention to improve knowledge about systematic reviews among health science professionals – a randomized controlled trial. OSF. 2022 Aug 02. URL: https://osf.io/x2mf5/ [accessed 2022-06-24]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
23. Puljak L, Sapunar D. Acceptance of a systematic review as a thesis: survey of biomedical doctoral programs in Europe. Syst Rev 2017 Dec 12;6(1):253 [FREE Full text] [doi: 10.1186/s13643-017-0653-x] [Medline: 29233170]
24. Krnic Martinic M, Pieper D, Glatt A, Puljak L. Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks. BMC Med Res Methodol 2019 Nov 04;19(1):203 [FREE Full text] [doi: 10.1186/s12874-019-0855-0] [Medline: 31684874]
25. Sharma R, Lakhani R, Rimmer J, Hopkins C. Surgical interventions for chronic rhinosinusitis with nasal polyps. Cochrane Database Syst Rev 2014 Nov 20;11(11):CD006990. [doi: 10.1002/14651858.CD006990.pub2] [Medline: 25410644]

26. Lissiman E, Bhasale A, Cohen M. Garlic for the common cold. Cochrane Database Syst Rev 2012 Mar 14;11(3):CD006206. [doi: 10.1002/14651858.CD006206.pub3] [Medline: 22419312]
27. Bogduk N, Dreyfuss P, Govind J. A narrative review of lumbar medial branch neurotomy for the treatment of back pain. Pain Med 2009 Sep 01;10(6):1035-1045. [doi: 10.1111/j.1526-4637.2009.00692.x] [Medline: 19694977]
28. Lam SK, Kwong EW, Hung MS, Pang SM, Chiang VC. Nurses' preparedness for infectious disease outbreaks: a literature review and narrative synthesis of qualitative evidence. J Clin Nurs 2018 Apr 11;27(7-8):e1244-e1255. [doi: 10.1111/jocn.14210] [Medline: 29193397]
29. Flores-Mateo G, Argimon JM. Evidence based practice in postgraduate healthcare education: a systematic review. BMC Health Serv Res 2007 Jul 26;7:119 [FREE Full text] [doi: 10.1186/1472-6963-7-119] [Medline: 17655743]
30. Horsley T, Hyde C, Santesso N, Parkes J, Milne R, Stewart R. Teaching critical appraisal skills in healthcare settings. Cochrane Database Syst Rev 2011 Nov 09(11):CD001270 [FREE Full text] [doi: 10.1002/14651858.CD001270.pub2] [Medline: 22071800]
31. Khan KS, Coomarasamy A. A hierarchy of effective teaching and learning to acquire competence in evidenced-based medicine. BMC Med Educ 2006 Dec 15;6:59 [FREE Full text] [doi: 10.1186/1472-6920-6-59] [Medline: 17173690]
32. Baig M, Sayedalamin Z, Almouteri O, Algarni M, Allam H. Perceptions, perceived barriers, and practices of physicians' towards evidence-based medicine. Pak J Med Sci 2016 Dec 31;32(1):49-54 [FREE Full text] [doi: 10.12669/pjms.321.8841] [Medline: 27022344]
33. Risahmawati RR, Emura SS, Nishi TT, Koizumi SS. Japanese resident physicians' attitudes, knowledge, and perceived barriers on the practice of evidence based medicine: a survey. BMC Res Notes 2011 Sep 28;4(1):374 [FREE Full text] [doi: 10.1186/1756-0500-4-374] [Medline: 21955346]
34. Adeodu A, Agius R, Madan I. Attitudes and barriers to evidence-based guidelines among UK occupational physicians. Occup Med (Lond) 2009 Dec 24;59(8):586-592. [doi: 10.1093/occmed/kqp121] [Medline: 19704031]
35. Halalau A, Holmes B, Rogers-Snyr A, Donisan T, Nielsen E, Cerqueira TL, et al. Evidence-based medicine curricula and barriers for physicians in training: a scoping review. Int J Med Educ 2021 May 28;12:101-124 [FREE Full text] [doi: 10.5116/ijme.6097.ccc0] [Medline: 34053914]
36. Puljak L. Using social media for knowledge translation, promotion of evidence-based medicine and high-quality information on health. J Evid Based Med 2016 Feb;9(1):4-7. [doi: 10.1111/jebm.12175] [Medline: 26372327]
37. Jakus D, Behmen D, Buljan I, Marušić A, Puljak L. Efficacy of reminders for increasing volunteer engagement in translating Cochrane plain language summaries: a pilot randomised controlled trial. BMJ Evid Based Med 2021 Apr 07;26(2):49-50. [doi: 10.1136/bmjebm-2020-111378] [Medline: 32636205]
38. Olsson C, Ringnér A, Borglin G. Including systematic reviews in PhD programmes and candidatures in nursing - 'Hobson's choice'? Nurse Educ Pract 2014 Mar;14(2):102-105. [doi: 10.1016/j.nepr.2014.01.005] [Medline: 24480093]
39. Chau J, Chien W, Liu X, Hu Y, Jin Y. Needs assessment and expectations regarding evidence-based practice knowledge acquisition and training activities: a cross-sectional study of healthcare personnel in China. Int J Nurs Sci 2022 Jan;9(1):100-106 [FREE Full text] [doi: 10.1016/j.ijnss.2021.11.001] [Medline: 35079611]
40. Ruppel KJ, Bone BA. Teaching evidence-based practice: knowledge to implementation in a BSN program, part 2. Worldviews Evid Based Nurs 2022 Jan 12 (forthcoming). [doi: 10.1111/wvn.12556] [Medline: 35020257]
41. Sorice V, Neal A. Nurses perceived institutional double standards in education and application of evidence-based practice. Evid Based Nurs 2022 Apr 25:ebnurs-2022-103515 (forthcoming). [doi: 10.1136/ebnurs-2022-103515] [Medline: 35470177]
42. Bond M, Bedenlier S, Marín VI, Händel M. Emergency remote teaching in higher education: mapping the first global online semester. Int J Educ Technol High Educ 2021 Aug 30;18(1):50 [FREE Full text] [doi: 10.1186/s41239-021-00282-x] [Medline: 34778537]
43. Hodges C, Moore S, Lockee B, Trust T, Bond A. The difference between emergency remote teaching and online learning. Educause Review. 2020. URL: https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning [accessed 2022-06-22]
44. Weber C, Ntasumbumuyange D, Ngoga E, Bazzett-Matabele L, Francis J, Paley P, et al. Continuing medical education during COVID-19: virtual training for gynecologic oncology management in Rwanda. Int J Gynecol Cancer 2021 Aug 20;31(8):1184-1185 [FREE Full text] [doi: 10.1136/ijgc-2021-002476] [Medline: 34016700]
45. Burgos L, Gil Ramirez A, Utengen A, Thamman R. Use of Twitter during COVID-19 pandemic: an opportunity for continuing medical education in cardiology. Medicina (B Aires) 2020;80 Suppl 6:122-123 [FREE Full text] [Medline: 33481748]
46. Kisilevsky E, Margolin E, Kohly RP. Access, an unintended consequence of virtual continuing medical education during COVID-19: a department's experience at the University of Toronto. Can J Ophthalmol 2021 Feb;56(1):e18-e19 [FREE Full text] [doi: 10.1016/j.jcjo.2020.10.002] [Medline: 33069676]
47. Wong A, Vohra R, Kopec K, Brooke N, Stolbach A. The importance of continuing medical education during the COVID-19 pandemic: the Global Educational Toxicology Uniting Project (GETUP). J Med Toxicol 2020 Jul 04;16(3):340-341 [FREE Full text] [doi: 10.1007/s13181-020-00788-2] [Medline: 32500282]

48. Kanneganti A, Lim KM, Chan GM, Choo S, Choolani M, Ismail-Pratt I, et al. Pedagogy in a pandemic - COVID-19 and virtual continuing medical education (vCME) in obstetrics and gynecology. Acta Obstet Gynecol Scand 2020 Jun 17;99(6):692-695 [FREE Full text] [doi: 10.1111/aogs.13885] [Medline: 32418212]
49. Badenes-Ribera L, Frias-Navarro D, Iotti NO, Bonilla-Campos A, Longobardi C. Perceived statistical knowledge level and self-reported statistical practice among academic psychologists. Front Psychol 2018;9:996 [FREE Full text] [doi: 10.3389/fpsyg.2018.00996] [Medline: 29988476]
50. Badenes-Ribera L, Frias-Navarro D, Pascual-Soler M, Monterde-I-Bort H. Knowledge level of effect size statistics, confidence intervals and meta-analysis in Spanish academic psychologists. Psicothema 2016 Nov;28(4):448-456. [doi: 10.7334/psicothema2016.24] [Medline: 27776615]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
EBHC: evidence-based health care
EBM: evidence-based medicine
EBP: evidence-based practice
GREET: Guideline for Reporting Evidence-based Practice Educational Interventions and Teaching
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: relative risk
SR: systematic review

Edited by A Mavragani; submitted 02.02.22; peer-reviewed by S Bhattacharjya, K Kalavani, N Khan; comments to author 20.04.22; revised version received 10.05.22; accepted 13.05.22; published 25.08.22.

Please cite as:
Krnic Martinic M, Čivljak M, Marušić A, Sapunar D, Poklepović Peričić T, Buljan I, Tokalić R, Mališa S, Neuberg M, Ivanišević K, Aranza D, Skitarelić N, Zoranić S, Mikšić Š, Čavić D, Puljak L
Web-Based Educational Intervention to Improve Knowledge of Systematic Reviews Among Health Science Professionals: Randomized Controlled Trial
J Med Internet Res 2022;24(8):e37000
URL: https://www.jmir.org/2022/8/e37000
doi:10.2196/37000
PMID:36006686

Review

The Role of Serious Video Games in the Treatment of Disordered Eating Behaviors: Systematic Review

Wymann S W Tang[1*]; Tricia J Y Ng[1*], MBBS; Joseph Z A Wong[1*], MBBS; Cyrus S H Ho[2], MBBS, MRCP, FAMS

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Cyrus S H Ho, MBBS, MRCP, FAMS
Department of Psychological Medicine
Yong Loo Lin School of Medicine
National University of Singapore
1E Kent Ridge Road, NUHS Tower Block
Singapore, 119228
Singapore
Phone: 65 67727831
Email: pcmhsh@nus.edu.sg

Abstract

Background: Eating disorders and other forms of disordered eating cause significant complications and comorbidities in patients. However, full remission with current standard treatment remains low. Challenges to treatment include underdiagnosis and high dropout rates, as well as difficulties in addressing underlying emotion dysregulation, poor impulse control, and personality traits. Serious video games (SVGs), which have the advantages of being highly engaging and accessible, may be potential tools for delivering various forms of treatment in addressing the underlying psychopathology of disordered eating.

Objective: This review aims to provide an overview of the possible mechanisms by which SVGs may affect the clinical course of disordered eating, while evaluating the outcomes of studies that have assessed the role of SVGs in the treatment of disordered eating.

Methods: A systematic search was performed on PubMed, PsycINFO, and Embase, using keywords related to SVGs, disordered eating, and eating disorders. A narrative synthesis was subsequently carried out.

Results: In total, 2151 papers were identified, of which 11 (0.51%) were included. Of these 11 studies, 10 (91%) were randomized controlled trials, and 1 (9%) was a quasi-experimental study. The types of SVG interventions varied across the studies and targeted different mechanisms of disordered eating, ranging from addressing problem-solving and emotion regulation skills to neurocognitive training for inhibitory control. Most (10/11, 91%) of the studies showed some benefit of the SVGs in improving certain physical, behavioral, or psychological outcomes related to disordered eating. Some (4/11, 36%) of the studies also showed encouraging evidence of the retention of these benefits at follow-up.

Conclusions: The studies included in this review provide collective evidence to suggest the various roles SVGs can play in plugging potential gaps in conventional therapy. Nonetheless, challenges exist in designing these games to prevent potential pitfalls, such as excessive stress arising from the SVGs themselves or potential gaming addiction. Further studies will also be required to assess the long-term benefits of SVGs as well as explore their potential preventive, and not just curative, effects on disordered eating.

(J Med Internet Res 2022;24(8):e39527) doi:10.2196/39527

KEYWORDS

serious video games; serious games; video games; gamification; digital health; eHealth; mobile health; mHealth; disordered eating; eating disorders

Introduction

Background

At least 9% of the world's population is affected by eating disorders [1], with adolescents and young adults being the most likely to be diagnosed with eating disorders [2,3]. Besides their implications for mental health, eating disorders also cause multisystemic medical complications [4]. *Disordered eating* is a term that encompasses eating disorders that were formally defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, such as anorexia nervosa (AN), bulimia nervosa (BN), and binge eating disorder (BED) [5]. However, besides the aforementioned eating disorders, disordered eating may also refer to pathological eating behaviors, including restricting, bingeing, purging, or other compensatory behaviors, without fulfilling the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, criteria for eating disorders [6]. Disordered eating attitudes can be driven by underlying body dissatisfaction or body image concerns (for restricting and purging), as well as an impairment of inhibitory control (for bingeing) [7,8].

Treatment of eating disorders hinges on early recognition and intervention, which is associated with better response to therapy as well as long-term outcomes [9]. According to guidelines from the National Institute for Health and Care Excellence, treatment of eating disorders such as AN, BN, and BED generally involves multidisciplinary effort, including psychoeducation, dietary rehabilitation, and monitoring and treatment of physical complications, as well as psychological treatment. Although eating disorder–focused cognitive behavioral therapy (CBT) is commonly perceived to be the first-line treatment and has the strongest and most rapid effects, evidence of its efficacy has been scarce for persistent AN as well as AN in adolescents [10]. Of note, a review published in 2002 revealed that only 46.8% of the patients with AN reached full recovery, 33.5% improved, and in 20.8% the disorder became chronic [11]. Treatment is also often limited by high rates of patients dropping out, ranging from 20.2% to 49.6% [12]. Mortality rates of these illnesses are still high, with those who received inpatient treatment for AN having more than 5 times increased mortality risk [13].

Challenges to the treatment of disordered eating are multifold. To start with, disordered eating is often underdiagnosed and undertreated because of a lack of awareness or feelings of shame associated with its diagnosis [14-16]. Obstacles to treating disordered eating can then be broadly considered in terms of patient- and clinician-related factors. Clinician-related factors that prevent use of evidence-based therapies include lack of training and individual beliefs regarding the effectiveness of certain forms of therapy, as well as over- or undervaluing certain elements of therapy [17]. There may also be difficulties in establishing a therapeutic alliance with the patient who is required to become vulnerable and give up some sense of control during the course of treatment [18]. Patient-related factors for poor treatment response can be examined through the underlying psychopathology of eating disorders. In patients with AN or BN, for instance, an underlying ego-syntonic pursuit of thinness or body dissatisfaction may be difficult for the patient to give up [18,19]. Furthermore, core features seen in disordered eating, such as the lack of impulse control, poor emotion regulation [20], and high reward-seeking behavior [21], are often difficult to address even with established forms of psychotherapy [22,23]. These comorbidities can also in turn affect motivation and compliance to psychological treatment [18]. In addition, traits of narcissistic, borderline, obsessive-compulsive, and avoidant personalities, which may be common in patients with disordered eating, may also negate treatment adherence and effectiveness [18,24]. The limitations of current conventional treatment as well as difficulties arising from the aforementioned factors necessitate a consideration of alternative treatment options or treatment options complementary to existing ones.

Technology is increasingly harnessed in the treatment of psychiatric conditions [25]. One example is the use of virtual reality (VR) to target clinical features of eating disorders such as binge eating, cravings, and body dissatisfaction through VR-mediated cue exposure and reference frame shifting [26]. According to a recent meta-analysis, VR-enhanced CBT has been shown to display better efficacy than CBT alone in reducing the frequency of binges and situation-induced body dissatisfaction [27]. VR has also been shown to be a feasible intervention to improve inhibitory control and thereby reduce binge eating episodes [28]. Internet-based CBT is another innovation that circumvents certain limitations of traditional CBT, such as high costs, long waiting time, and perceived stigma associated with seeking help for psychiatric disorders [29]. However, the challenges of high dropout rates and poor compliance remain [30].

This review is interested in the possible role of serious video games (SVGs) in addressing disordered eating. Games are by definition an activity in which "independent decision makers seek to achieve their objectives in a limiting context" [31]. Serious games can be simply viewed as games with a *serious* objective, often for education, vocational training, or problem-solving [32]. Nonetheless, the line between serious and entertainment games is sometimes blurred because certain entertainment games are sometimes repurposed for an educational purpose as well [33]. Serious games have been used in various fields, ranging from education and military applications to interpersonal communications training [32]. A key feature of these serious games is that besides delivering knowledge or skills to the player, they stimulate an environment through narrative story, gameplay, or encounters that is safe and controlled for users to be able to practice new learned skills or behaviors [34,35]. In today's landscape, serious games are increasingly available on digitized platforms, ranging from mobile apps [36] to VR devices [35].

Serious games have had some early success in treating a range of psychiatric conditions and have been deemed to be highly feasible and acceptable to both patients and clinicians [34]. Psychiatric conditions in which serious games have been used to address symptoms include depression [37] and addiction problems such as substance abuse [38] and internet addiction [39]. It is noteworthy that these psychiatric conditions happen to be comorbidities commonly associated with disordered eating. One example of serious games being used in treating psychiatric conditions is Smart, Positive, Active, Realistic, X-Factor Thoughts (SPARX), a participative game based on CBT that targets depression in adolescents. The treatment outcomes of SPARX were shown to be comparable with those of conventional treatment despite its being a predominantly self-guided resource [40,41]. Dropout rates for this intervention were also low at approximately 9%, suggesting that such games may have an advantage in engaging patients belonging to the adolescent age group [41].

Serious games as a treatment can be highly engaging [42]. It has been shown that being appropriately challenged has positive effects on both engagement and learning, with the challenge of the game being a strong predictor of learning outcomes [43]. In serious games, players can advance through individualized game difficulty levels and be constantly challenged, allowing for personal growth [35]. By contrast, this may be difficult to replicate in traditional CBT. Furthermore, the potential of serious games in improving impulse control [44] and emotion regulation [45], which are core features in the psychopathology of disordered eating, would be one of their key foreseeable advantages. However, current studies examining the therapeutic effects of serious games on persons with disordered eating are sporadic. Hence, there is a need to systematically consolidate and critique such available studies to understand the effectiveness of serious games as a therapeutic medium for disordered eating behaviors.

Objectives

This study aimed to answer the question regarding the effectiveness of SVGs in reducing disordered eating behaviors and addressing their underlying psychopathology. The review will provide an overview of the possible mechanisms by which SVGs can affect the clinical course of disordered eating, while evaluating the effectiveness of SVGs and their potential role in complementing current treatment options for disordered eating.

Methods

Search Strategy

The systematic review and meta-analysis were reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Three databases (PubMed, Embase, and PsycINFO) were last searched on March 29, 2022, with no restrictions on publication dates. The following search terms were applied in the search strategy, with the use of relevant controlled vocabulary such as Medical Subject Headings, Emtree, and PsycINFO Thesaurus terms: *((serious gaming) OR (game) OR (computer-assisted therapy) OR (gamification) OR (gaming simulation) OR (video game) OR (applied game) OR (mobile game) OR (gamified application) OR (digital game)) AND ((eating disorder) OR (anorexia) OR (bulimia) OR (binge eating) OR (impulsive eating) OR (body image) OR (body dissatisfaction) OR (self-control) OR (inhibitory control)).*

Selection of Articles

The inclusion criteria were as follows: studies (1) needed to be peer-reviewed randomized controlled trials (RCTs) or quasi-experimental design studies; (2) involved interventions delivered on a digital platform with gaming elements; (3) assessed the efficacy of SVGs in terms of their therapeutic benefit for eating disorders, disordered eating behaviors, and underlying body image concerns or inhibitory control; (4) involved populations prone to disordered eating; and (5) were in English.

Studies assessing the effects of gamified food-related inhibitory control training (ICT) were included because of evidence suggesting that interventions targeting food-related impulsivity have the potential to reduce binge eating frequency as well as address food cravings [46,47]. This is because disordered eating behaviors such as overeating have been linked to lack of inhibitory control, which is an underlying trait seen in eating disorders [8,48]. Studies examining the use of video games originally designed for entertainment purposes and not for their therapeutic effect were also included.

Studies were excluded if (1) the intervention was related to web-based CBT, web-based counseling, or guided self-help without elements of gamification; and (2) it involved VR-mediated interventions. Studies with VR-mediated interventions were excluded to isolate the effect of non–VR-related gamification on disordered eating as much as possible because of the extensive evidence of the efficacy of VR-mediated interventions [26-28].

The selection of articles was conducted by 3 authors (JZAW, TJYN, and WSWT). The selection was performed over 2 phases. In the first phase, articles were screened depending upon their relevance to this review based on their title and abstract. Shortlisted articles from the first phase of screening subsequently underwent full-text assessments for their eligibility to be included in this review. Disagreements on the selection process were resolved by discussion among the aforementioned authors as well as consultation with the senior author (CSHH).

Review and Quality Assessment

The quality of each included study was independently evaluated by 2 separate authors (TJYN and WSWT) with any disagreements resolved through discussion. The RCTs were assessed using the Cochrane Risk-of-Bias 2 tool, whereas the noncontrolled experimental study was assessed using the Joanna Briggs Institute Critical Appraisal Checklist for Quasi-Experimental Studies [49,50].

Data Extraction and Evaluation

After finalizing the selection of studies, the following information was extracted from the papers: (1) publication details (eg, title, authors, and country), (2) details of the SVG intervention (eg, type of platform, game objectives, and gameplay), (3) details of the studied population (eg, population type, sample size, and gender ratio), and (4) type of outcomes measured and the respective results. The studies were then categorized by the posited mechanisms by which the SVG intervention affected disordered eating.

The interventions used in the included studies were collectively assessed by adapting a framework developed by Murray et al [51] for evaluating the potential benefits of digital health interventions. The following factors were considered: (1) the accessibility of SVGs to populations with disordered eating, (2) causal explanations for how SVGs can treat disordered eating, (3) the key components required for SVGs to affect positive outcomes on disordered eating, (4) how target populations should be supported in the use of SVGs to treat or prevent disordered eating, (5) the possible harms of SVGs and the likelihood of their risks, (6) the costs of using SVGs incurred by users and the health system, and (7) the overall utility of SVGs. The evaluation of the SVGs was aided by a framework developed by Liverpool et al [52] that highlights key components contributing to engagement in digital mental health interventions among young people. This involves assessing the SVGs for intervention-specific factors, including suitability, usability, and acceptability, as well as user-specific factors, including motivation, opportunity, and capability [52].

Results

Search Results

Our search strategy yielded 2916 articles. Hand searching did not uncover other relevant studies. From the 2916 articles, 765 (26.23%) duplicates were removed. Of the remaining 2151 articles, we excluded 2108 (98%) after title and abstract screening because they did not include serious gaming as an intervention, or they focused on a disease unrelated to eating disorders. After full-text screening of the 43 remaining articles, 32 (74%) were excluded. Reasons for the exclusion of articles after full-text screening include not reporting outcomes related to eating disorders, assessing SVGs that were not run on digital platforms, and interventions not having gamified elements. Thus, of the initial 2916 articles identified, 11 (0.38%) were included in the systematic review. The PRISMA flowchart is presented in Figure 1. The main outcomes of the included articles are summarized in Table 1, whereas the details of the SVG intervention of each study are summarized in Table 2. The characteristics of the included studies are presented in Multimedia Appendix 1 [53-63], and the assessments of the risk of bias are presented in Multimedia Appendix 2 [54-63] and Multimedia Appendix 3 [53]. Of the 11 included studies, 10 (91%) were deemed to be at a low overall risk of bias, with the exception of 1 (9%) study, regarding which concerns were raised over the suitability of the outcome measures.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing the selection of the studies.

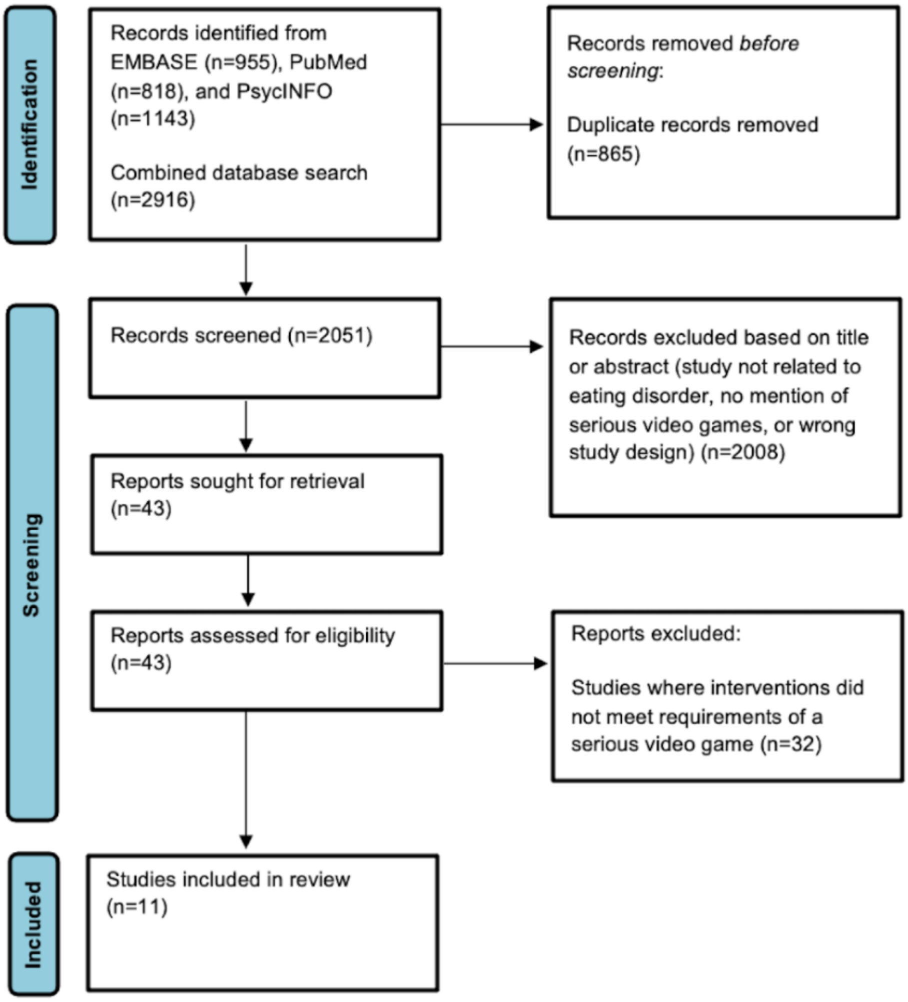


Table 1. Characteristics and outcomes of the included studies.

Serious video game category and title	Study, year	Outcome measures (physical, behavioral, and psychological)	Main outcomes
Serious video games for emotion regulation skills			
The Use of Videogames as Complementary Therapeutic Tool for Cognitive Behavioral Therapy in Bulimia Nervosa Patients	Fernandez-Aranda et al [53], 2015	• Frequency of bingeing and purging • Dropout rates • Eating disorder psychopathology • Anxiety and anger • Remission rate (partial or complete)	• Intervention group achieved nonstatistically significant higher rates of total remission than the control group (50% vs 28%, respectively; *P*=.22) • Intervention group had lower treatment attrition rates than the control group (20% vs 44%, respectively; Cohen d=0.54) and displayed improvements in emotion regulation, whereas the control group showed persisting emotion dysregulation
Serious video games for body image concerns			
An App-Based Blended Intervention to Reduce Body Dissatisfaction: A Randomized Controlled Pilot Study	Kollei et al [54], 2017	• Eating disorder psychopathology • Depressive symptoms • Body dissatisfaction	• Intervention group showed significantly greater reduction in body dissatisfaction (Cohen d=–0.62; *P*=.001) and a medium-sized effect in reduction of eating disorder symptoms (Cohen d=–0.46; *P*=.007) which persisted at 1-month follow-up • No significant effect on depressive symptoms was noted
A Brief Mobile Evaluative Conditioning App to Reduce Body Dissatisfaction? A Pilot Study in University Women	Kosinski [55], 2019	• Depressive symptoms • Eating disorder psychopathology • Self-esteem • Body dissatisfaction and drive for thinness	• No significant reduction of body dissatisfaction between the evaluative conditioning and control conditions was noted, but body dissatisfaction fell across conditions with a small effect size (*r*=0.27; *P*<.005). Similar patterns presented for the drive for thinness (*r*=0.67; *P*<.001) and self-esteem (*r*=0.29; *P*<.05) • No statistically significant effects were observed for bulimia and restraint scores
An Interactive Training Programme to Treat Body Image Disturbance	Gledhill et al [56], 2017	• Eating disorder psychopathology • Self-esteem • Body size perception and body image concerns	• Study 1 • The intervention succeeded in shifting the thin-fat categorical boundary for individuals with body size concerns, as well as improved eating restraint (day 14 difference z score=0.92, 95% CI 0.33 to 1.51; *P*=.003), body weight (day 14 difference z score=1.15, 95% CI 0.58 to 1.72; *P*<.001), and body shape concerns (day 14 difference z score=1.04, 95% CI 0.45 to 1.64; *P*=.001) • Study 2 • The intervention succeeded in shifting the thin-fat categorical boundary significantly in participants with anorexia nervosa. Eating disorder symptoms also improved for at least a month (Eating Disorder Examination Questionnaire score day 1 vs day 30 difference z score=0.74, 95% CI 0.20 to 1.28; *P*=.008) • The degree of body size category boundary shift was significantly correlated with changes in the eating disorder symptoms

Serious video game category and title	Study, year	Outcome measures (physical, behavioral, and psychological)	Main outcomes
When You Exercise Your Avatar in a Virtual Game: The Role of Avatars' Body Shape and Behavior in Users' Health Behavior	Joo and Kim [57], 2017	• Exercise and eating behavior	• No significant effects of the avatars' lifestyle on the participants' exercising or eating behaviors were observed • There was a significant positive effect of the normal-weight body shape of the avatars on the participants' exercising behaviors (*P*=.02). No difference was noted for players who used the obese avatars
Serious video games for neurocognitive training to influence eating behaviors			
App-Based Food-Specific Inhibitory Control Training as an Adjunct to Treatment as Usual in Binge-Type Eating Disorders: A Feasibility Trial	Keeler et al [58], 2022	• Food valuation on palatability of high- and low-energy–dense foods • Eating disorder psychopathology • Depressive symptoms • Anxiety • Impulsivity	• Intervention did not reduce binge eating frequency but showed greater reduction in eating disorder psychopathology (SES[a]=–0.57, 95% CI –1.12 to –0.03) and valuation of high-energy–dense foods than usual treatment (SES=–0.61, 95% CI –0.99 to –0.24). These effects were lost and reduced, respectively, at 8-week follow-up • At 8 weeks, the intervention group showed greater reduction in food addiction symptoms and lack of perseverance with a small effect size (SES=–0.23, 95% CI –0.81 to 0.34)
Gamified Working Memory Training in Overweight Individuals Reduces Food Intake but Not Body Weight	Dassen et al [59], 2018	• BMI • Food intake and healthy eating • Self-control • Dropout rate • Executive function • Eating disorder psychopathology	• WM[b] training did not result in significant additional weight loss • WM training resulted in a significant reduction in caloric intake after training, especially at high levels of craving • Both groups showed improvements in self-reported emotional eating and self-control
Gender Differences in the Effect of Gamification on Weight Loss During a Daily, Neurocognitive Training Program	Forman et al [60], 2021	• Weight • Enjoyment of game and compliance to treatment • Inhibitory control	• Gamification had a significantly stronger effect on weight loss for men than for women • No significant differences were observed between genders for the effect of gamification on enjoyment, compliance, and impulse control
A Serious Game to Increase Healthy Food Consumption in Overweight or Obese Adults: Randomized Controlled Trial	Blackburne et al [61], 2016	• Eating behavior • Cognitive restraint • Go–No-Go performance	• Inhibitory control improved with the intervention, which was associated with increased consumption of healthy foods and reduced consumption of unhealthy foods. Cognitive restraint also improved
Computerized Neurocognitive Training for Improving Dietary Health and Facilitating Weight Loss	Forman et al [62], 2019	• Weight • Frequency of food consumption • Implicit preference for sweets	• ICT[c]—both gamified and nongamified—were deemed acceptable and feasible • Only participants with higher baseline implicit preference for sweets experienced weight loss benefits from ICT. However, gamification marginally reduced the impact of ICT
Executive Function Training With Game Elements for Obese Children: A Novel Treatment to Enhance Self-regulatory Abilities for Weight-Control	Verbeken et al [63], 2013	• BMI • Treatment feasibility and acceptability • Executive function • Visuospatial WM • Stop-signal task performance	• The intervention showed significant effects in WM and meta-cognition and displayed significant improvements in weight loss maintenance at 8 weeks, although the effect was lost at 12 weeks • No significant effects were observed for inhibition and the stop-signal task

[a] SES: standardized between-group effect sizes.

[b] WM: working memory.

[c] ICT: inhibitory control training.

Table 2. Characteristics of the serious video gaming interventions.

Game category and study	Game title	Serious game genre	Platform	Objectives	Gameplay
Serious video games for emotion regulation skills					
Fernandez-Aranda et al [53], 2015	PlayMancer: Islands	Goal oriented and problem-solving	PC	To increase emotion–self-control skills and self-control over users’ general urgency to act	• Players are immersed in the setting of an island and are required to overcome challenges by achieving therapeutic targets • Biosensors and a camera that continuously tracks the emotional state of the player are used to monitor physiological changes in response to the players’ emotional state
Serious video games for body image concerns					
Kollei et al [54], 2017	Mindtastic Body Dissatisfaction app	Cognition and brain training	Mobile app	Approach-avoidance training to foster approach of functional stimuli and avoidance dysfunctional stimuli	• Players are shown (1) pictorial stimuli of their own bodies as well as that of their ideal bodies and (2) positive and negative body-related statements. They are required to pull the positive statements and pictures of themselves toward themselves and swipe away the negative statements and idealized pictures
Kosinski [55], 2019	Executive conditioning app	Cognition and brain training	Mobile app	Evaluative conditioning	• The player’s photographs are taken to act as conditioned stimuli. Positive photographs that elicit a positive affective response and do not correspond to feminine ideals were used as unconditioned stimuli. Players are shown 3 conditioned stimuli and unconditioned stimuli pairings at the start and are required during the game to pick out their conditioned stimuli and unconditioned stimuli pair as quickly as possible
Gledhill et al [56], 2017	Perceptual training with two-alternative forced-choice decisions	Cognition and brain training	PC	Evaluative conditioning	• Participants are presented with a series of computer-generated imagery images of women’s bodies and trained to judge the respective body size. Feedback was given to the participants on whether their responses were accurate. “Inflationary” feedback was given with the intent to shift their categorical boundary of a “fat” body shape by 2 body shape variations higher
Joo and Kim [57], 2017	The Sims 4	Goal oriented and problem-solving	PC	To increase emotion– self - control skills and reduce general impulsive behaviors	• A web-based life simulation game, The Sims 4, was used • Players were assigned to either a normal weight or obese avatar. Players were then instructed to operate their avatars in a healthy (exercise and fresh foods) or unhealthy lifestyle setting (sedentary lifestyle and unhealthy foods)
Serious video games for neurocognitive training to influence eating behaviors					

Game category and study	Game title	Serious game genre	Platform	Objectives	Gameplay
Keeler et al [58], 2022	FoodT, an inhibitory control training app	Cognition and brain training	Mobile app	To increase inhibitory control	• Players are presented with pictorial stimuli consisting of high-energy foods, low-energy foods, and filler items, accompanied by "Go," or "No Go" cues. Participants are required to tap on the "Go" items and avoid tapping on the "No Go" items
Dassen et al [59], 2018	WM[a] training	Cognition and brain training	Tablet computer or PC	Psychoeducation and WM training	• Each session comprises 3 WM tasks in the setting of a restaurant involving visuospatial memory, backward digit span, and object memory
Forman et al [60], 2021	Go–No-Go training	Cognition and brain training	PC	To increase inhibitory control	• Players are presented with pictorial stimuli consisting of healthy and unhealthy foods, accompanied by "Go," and "No Go" cues, respectively. Participants are required to tap on the "Go" items and avoid tapping on the "No Go" items
Blackburne et al [61], 2016	"NoGo," a Go–No-Go inhibitory control training app	Cognition and brain training	Mobile app	To increase inhibitory control	• Players are shown stimuli of healthy and unhealthy foods. Each game consists of (1) Go–No-Go trials where the reaction timer starts counting down next to the image after it is shown and (2) stop trials where the timer counts down while the images change between categories
Forman et al [62], 2019	Go–No-Go training	Cognition and brain training	PC	To increase inhibitory control	• The gamified inhibitory control training involved the task of moving in a grocery store as quickly as possible while choosing the correct foods. This required players to respond to frequently presented stimuli (healthy foods such as fruit and vegetables) and inhibit their responses to nonfrequent stimuli (high-sugar food)
Verbeken et al [63], 2013	"Braingame Brian," an executive function training game	Cognition and brain training	PC	To increase executive function (inhibitory control and WM)	• The game is set in a game world with a storyline where the character, Brian, is required to complete tasks involving (1) WM training where the player has to reproduce correctly a random sequence of rectangles lighting up and (2) inhibitory control training in the setting of a factory, including both go trials and stop trials

[a] WM: working memory.

Physical Outcomes

Of the 11 studies, 4 (36%) measured the effects of the gamified interventions on physical outcomes such as weight or BMI. All interventions were used in the context of treating overeating in individuals who were overweight [59,60,62,63]. Among the 4 studies, only 2 (50%)—the gamified ICT in the study by Forman et al [60] and the executive function (EF) training in the study by Verbeken et al [63]—reported statistically significant reduction in BMI after the serious game intervention. This effect, achieved in the study through an intervention centered on ICT was noted to be more significant in men than in women [60]. However, in another study by Forman et al [62], it was reported that gamification slightly reduced weight loss benefits compared with normal ICT [62].

Of note, the other studies in which a healthy increase in BMI would have been an ideal outcome in the context of undereating did not measure weight or BMI as a study outcome likely because of the short duration of their interventions.

Behavioral Outcomes

In terms of behavioral outcomes, the studies reported a variety of outcomes such as frequency of binges and purges; food intake; enjoyment in using, and compliance to, the intervention; dropout rates; nonverbal communication; and drive for exercise. However, the reporting on the type of behavioral outcomes among the studies was not consistent.

Varying outcomes have been obtained in terms of eating habits and attitudes toward food. CBT coupled with a role-playing problem-solving SVG achieved neither a statistically significant increase in the reduction of bingeing or purging episodes nor an increase in the rates of total or partial clinical remission of BN [53]. In addition, an avatar-based SVG based on social cognitive theory was unable to influence the short-term eating behaviors of participants [57].

However, the study by Dassen et al [59] showed that gamified working memory (WM) training helped in reducing caloric intake, especially at high levels of craving [59]. In a different vein, the study by Blackburne et al [61] showed that a gamified ICT app improved inhibitory control that was associated with increased consumption of healthy foods and reduced consumption of unhealthy foods. Similarly, an app-based food-specific ICT in the study by Keeler et al [58] showed small-sized effects on a greater reduction in food addiction and medium-sized effects in participants' valuation on the palatability of high-energy–dense food. However, the latter effect was diminished at later follow-up. The intervention also did not show greater reduction in binge eating frequency than usual treatment [58].

Dropout or attrition rates were measured in 27% (3/11) of the studies. The study by Fernandez-Aranda et al [53] showed that conventional CBT coupled with an SVG (*PlayMancer:Islands*) resulted in lower treatment attrition rates than CBT alone (20% vs 44%, respectively) [53]. In a different setting, the gamified WM training in the study by Dassen et al [59] demonstrated a participant dropout rate of 23% (21/91) across the interventional training and sham training groups using the same game. Gamified ICT in the study conducted by Forman et al [62] demonstrated an attrition rate of 14.8%, which did not differ significantly from the attrition rate in the nongamified ICT group.

Psychological Outcomes

Overview

Across the studies included in the review, a variety of instruments were used to assess the features of eating disorders and their psychiatric comorbidities. Commonly applied questionnaires in the field of eating disorders, namely the Eating Disorder Inventory-2, Symptom Checklist–Revised, State-Trait Anxiety Index, and State-Trait Anger Expression Inventory-2, were used. In some (2/11, 18%) of the studies, throughout the duration of the treatment, patients kept a daily food and purging diary. Assessments were made before and after group therapy. The frequency of binges and purges was reported in the study by Fernandez-Aranda et al [53]; however, no statistically significant differences were found between the 2 clinical groups.

Eating Disorder Psychopathology

Of the 11 studies, 6 (55%) reported on participant scores on the eating disorder questionnaires [53-56,58,59]. Of these 6 studies, 4 (67%) displayed statistically significant improvements in eating disorder symptoms [54,56,58,59], although these effects were not always maintained at follow-up after a longer period [54,58].

Body Dissatisfaction and Body Size Perception

As it is an important psychopathological factor driving disordered eating, body dissatisfaction was reported on by 18% (2/11) of the studies, involving 113 participants in total. Both studies involved university students: Kollei et al [54] focused on students with significant body dissatisfaction, whereas Kosinski [55] focused on female students. Both interventions were based on approach-avoidance training aimed at fostering the avoidance of dysfunction stimuli and approach of functional stimuli. This was done by pairing participants' photographs with positively conditioned stimuli. In both studies, participants in the intervention group reported significant reductions in body dissatisfaction, although the study by Kosinski [55] did not display a significant difference compared with the control group where neutral stimuli were used.

Although the study by Gledhill et al [56] did not measure body dissatisfaction directly as an outcome, it showed that a gamified perceptual training paradigm with inflationary feedback was successful in modifying perceptions on body size (measured by their thin-fat categorical boundary) in both individuals with body size concerns and patients with AN.

Gender Differences in Outcomes

Forman et al [60] assessed the associations between gender differences (between men and women) and the effectiveness of a gamified neurocognitive intervention program on weight loss and inhibitory control. Although gamification significantly enhanced weight loss in men more than in women, the association between overall effectiveness on inhibitory control and gender differences was unclear. The study also assessed participant enjoyment of the daily neurocognitive training program. Although there was no significant difference due to gender in the effect of enjoyment, women generally reported, with a small effect, higher enjoyment scores. By contrast, gamification did not seem to enhance men's enjoyment of the intervention. The study also examined compliance to the treatment regimen with gamification. Although statistically nonsignificant, compliance to treatment regimen as a result of gamification improved in men but decreased in women [60].

Discussion

Principal Findings

The results of the review show an overview of the possible ways that SVGs or gamification of CBT interventions can effect change in the symptoms of disordered eating or its underlying psychopathology. These effects can be classified into physical, behavioral, and psychological outcomes. Among the papers that reported physical outcomes of BMI or weight change, it was found that the SVGs and gamified interventions potentially had some effect in improving weight. However, it may not be clear whether this effect can be truly attributed to the inclusion of gamification elements or whether it would still be retained in the nongamified intervention of the same type [62]. The duration of the interventions was also likely to be too short to identify the long-term effects of these SVGs on the participants' weight. Behaviorally speaking, there were mixed results on the effect of the SVGs on desired changes in eating behavior. However, it is noted that gamified versions of neurocognitive training such as WM or ICT seemed to report more positive behavioral changes than the role-playing or avatar-based games. Psychologically speaking, it was observed that SVGs of different types such as those targeting body dissatisfaction, inhibitory control, or WM were all able to produce improvements in eating disorder symptoms or psychopathology [54,56,58,59]. This informs developers of future SVGs for disordered eating on potential game mechanisms and types.

Key Components of SVGs and Their Mechanisms in Treating Disordered Eating Behaviors

The studies (n=11) included in this review comprised 11 different SVGs. As seen in Table 2, the characteristics of the SVGs examined in this review were varied and addressed different aspects of disordered eating with different mechanisms. Of the 11 SVGs, 2 (18%) were gamified adaptations of approach-avoidance and evaluative conditioning training based on the cognitive behavioral model of body satisfaction, aimed to foster approach of functional stimuli and avoidance of dysfunctional stimuli in body dissatisfaction [54,55]. In a similar vein, the SVG in the study by Gledhill et al [56] was also based on cognitive training and worked by calibrating participants' perception of categorical definitions of body shape. Of the 11 SVGs, 4 (36%) aimed to improve inhibitory control regarding unhealthy foods through ICT with *Go–No-Go* (GNG) food stimuli [58,60-62], whereas in the study by Fernandez-Aranda et al [53], the SVG was an immersive role-playing game that required players to solve problems by using emotion regulation skills and achieving impulse control. Interestingly, Dassen et al [59] and Verbeken et al [63] recognized the role of WM and EF, respectively, in behavioral self-regulation and sought to assess the effect of gamified WM and EF training on food intake [59,63,64]. Finally, it is noted from the paper by Joo and Kim [57] that commercially available simulation games, such as *TheSims 4*, may potentially be helpful in influencing real-life health behaviors of users through their avatars' body shape and health behaviors.

The paper by Fernandez-Aranda et al [53] was the only one that examined the effects of an SVG and CBT on eating disorders, specifically BN. The study examined the role the SVG was able to play in addressing the limitations of CBT in treating the underlying traits of eating disorders (namely emotion regulation and impulsivity) while harnessing its gamified elements to ensure compliance and increased accessibility. *PlayMancer:Islands* is an SVG specifically designed for this purpose. The gameplay is unique in that the facial expressions and physiological markers such as heart rate and respiratory rate were monitored as responses to the players' emotional states. Undesirable emotional states would then correspond to increased difficulty in completing tasks. Through the game, participants are required to learn and demonstrate emotion–self-control skills. Although the intervention had limited effect on clinical remission of BN, it was successful in addressing eating disorder psychopathology and in reducing treatment dropout rates. This is corroborated by case studies and case series, which, although not included in this review, showed positive effects of *PlayMancer: Islands* on the intervention group participants' impulse control, emotion regulation, anxiety, and novelty seeking, as well as various physiological measures, including a functional magnetic resonance imaging scan comparing engagement of brain areas, compared with those of healthy controls [65-67]. The case study by Giner-Bartolomé et al [66] even reported changes in eating behavior, such as the number of bingeing episodes. Although SVGs such as *PlayMancer: Islands* cannot be considered an alternative to replace conventional therapy, it is likely that they can be used effectively to supplement the limitations of CBT in addressing emotion regulation and personality traits underlying eating disorders [24]. Nonetheless, stronger evidence for this intervention may be required in studies with a larger sample size and longer duration of follow-up. It will also be worth assessing the effect of *PlayMancer: Islands* on eating disorders other than BN with common psychopathological features, such as AN or BED. The feasibility and accessibility of the SVG intervention would also be a future practical consideration, given the use of biosensors and facial recognition technology.

It is interesting to note that the majority (10/11, 91%) of the other SVGs reviewed differ from *PlayMancer: Islands* in that these other games seek to provide cognitive response training to modify and address the underlying perceptions that drive disordered eating as opposed to the goal-oriented and problem-solving nature of gameplay used in *PlayMancer: Islands*. Of note, body dissatisfaction, ICT, and WM were the main concepts targeted by the SVGs studied.

The studies by Kollei et al [54] and Kosinski [55] both showed significant effects of SVGs in improving symptoms of body dissatisfaction. The Mindtastic Body Dissatisfaction app in the study by Kollei et al [54] used approach-avoidance training where participants are conditioned to perform approach actions (swiping toward themselves) and avoidance actions (swiping away) in response to specific stimuli such as positive and negative body-related statements, respectively. This approach has previously been shown to modify biases toward food and alcohol and in reducing their consumption [68-70]. By contrast, the mobile app intervention in the study by Kosinski [55] used evaluative conditioning, a form of Pavlovian conditioning, where a change in behavior or response is induced when conditioned stimuli (participants' own photographs) are paired to unconditioned stimuli (positive body image photographs). The evaluative conditioning approach is supported by the work of Aspen et al [71], whose study showed that pairing participants' bodies with positive social stimuli and pairing other bodies with neutral stimuli resulted in improvements in body shape concerns and self-esteem, as well as reduced food restriction. Kollei et al [54] and Kosinski [55] collectively show the potential of gamified mobile app interventions in providing remote cognitive training to modulate adaptive body image attitudes, which can be seen to be highly feasible and accessible. As both studies examined the effects of the interventions on university students, it would be useful to understand the generalizability of these effects on other demographics, as well as persons formally diagnosed with eating disorders. Given the use of these mobile app interventions in university students not diagnosed with eating disorders, it may also be interesting to assess the potential role of SVGs in preventing eating disorders in populations considered to be at high risk who have not yet been formally diagnosed. Studies with a longer follow-up duration may be helpful to inform whether such games can reduce the incidence of eating disorder diagnoses.

Commonly compared with approach-avoidance training, GNG training is another form of motor response training aimed at changing behaviors. GNG training was the modality that was gamified in the *FoodT* app in the study by Keeler et al [58]; this app, when used by patients with BN and BED, showed a reduction in eating disorder pathology and diminished the perceived palatability of high-energy–dense foods. GNG training was also used in the intervention in the studies by Forman et al [60,62] and Blackburne et al [61]. Although used in different populations, the effect size on eating disorder pathology of the *FoodT* app was similar to that of the Mindtastic Body Dissatisfaction app in the study by Kollei et al [54]. Although the meta-analyses conducted suggested that GNG training may be more effective than approach-avoidance training in influencing food behaviors [72,73], it is difficult to compare the usefulness of both SVGs because they address different psychopathological drivers of disordered eating (poor inhibitory control and body dissatisfaction). The *FoodT* app also lends support to the feasibility of cognitive training for disordered eating through gamified mobile apps, and it may be interesting to see whether a combination of features from these cognitive training apps can lead to better outcomes.

Accessibility and Feasibility of Using SVGs

In addition to the benefits SVGs may have in the treatment of disordered eating, they may have other advantages. Compared with conventional treatment, SVGs are likely to be more accessible to patients with disordered eating. First, there is a high degree of correspondence between the demographics of video game players and patients with disordered eating; 38% of the video game players in the United States fell within the age range of 18 to 34 years, whereas 20% were aged <18 years [74]. It is hence likely that SVGs will appeal more to patients with disordered eating. Furthermore, with more SVGs using mobile apps as a platform coupled with increasing smart device ownership, SVGs have the potential to reach more persons with disordered eating. Finally, SVGs have relatively low barriers to use and also offer users the option of being by themselves, which may help circumvent issues of trust with the therapist or personality-related comorbidities such as avoidance in patients with disordered eating [18,75]. Hence, these factors may address certain challenges in the treatment of disordered eating such as low treatment uptake rates or high dropout rates.

Further Tailoring of SVGs and Supporting Their Use

The study by Forman et al [60] compares gender differences in the effect of gamification on weight loss. Eating disorders tend to affect women more than men; however, disordered eating can affect men as well. In general, most studies have samples that consist predominantly of women. The study provides a different perspective on gender differences in the effect of SVGs. Most video game users tend to be men [76], and men are reported to be more motivated by gaming elements than women, which is supported by Hassouneh and Brengman [77], whose study indicates that the effects of SVGs on women and men can differ. The study also found that compliance to SVG interventions tends to be higher in men than in women. In addition, it was reported that the gamification elements had more positive outcomes for men. This may inform future recommendations on the potential use of SVGs for the treatment of disordered eating based on gender, as well as expectations regarding their subsequent clinical effectiveness. However, the SVG assessed by Forman et al [60] used a gamified GNG training modality; hence, it is not clear whether the trend noted in this study would apply to other SVG genres. As discussed by Forman et al [60], it is plausible that the inclusion of other gaming elements such as more elaborate backstories, incentives, and components that require collaboration with other participants in these SVGs may contribute to their appeal to women [78-81]. Such features may serve to enhance user motivation as well as foster a sense of connection with peers facing similar struggles with disordered eating. This may further promote user engagement with the SVGs.

Although not explicitly addressed by the included studies, it is important to consider the concerns over privacy and confidentiality that users may have regarding such SVGs, especially in the context of treating disordered eating behaviors. A review by Borghouts et al [82] showed that concerns over the safety and privacy of disclosed information can act as a barrier to engagement in such digital mental health interventions [82]. On the flipside, assurance of anonymity in the use of the digital intervention encouraged engagement [52,82]. Hence, in the design and implementation of such SVGs, it would be necessary to make sure that appropriate safeguards are in place to protect sensitive user data. It has also been suggested that the use of trusted brand names and transparency as well as evidence of credibility would serve to increase engagement of young users with digital mental health interventions [52].

Possible Risks and Adverse Outcomes

Despite the many benefits an incorporation of SVGs may confer to the treatment of disordered eating, it is important to note potential pitfalls where SVGs may complicate treatment. One such instance would be the situation where the inclusion of serious gamification elements such as player scores may inadvertently cause additional stress or cognitive load because of the participant's drive to do well in the game and distract participants from the therapeutic intentions of the game [83,84]. This could also potentially cause further insult to the self-esteem of the participant or further diminish confidence and motivation in complying with treatment. This is supported by the study by Mekler et al [85], which suggests that gamification elements such as scores and leader boards may function more as an extrinsic source of motivation that increases performance quantity but do not have a significant relationship with competence or intrinsic motivation; for example, players may attempt the game multiple times for the simple object of passing a particular level while not necessarily improving in competency. This may be apparent in the study by Forman et al [62] where gamified ICT actually reduced the weight loss benefits derived from conventional ICT. This was posited by the authors to possibly be due to distraction from the core stimuli by gaming elements such as visuals, music, and sound effects, which may have reduced prepotent reward response, hence affecting inhibitory response.

Nonetheless, among the other SVGs assessed in this review, scoring systems and leader boards are not key elements and may not necessarily cause the aforementioned effect. Furthermore, games such as *PlayMancer: Islands*, which uses biosensors to track physiological and facial expression responses to emotional states, will be able to pick up states of distress in participants as well as obtain an objective measure of progress [53]. Hence, as more research is being conducted to assess the effects of gamification elements on motivation, careful design and implementation of gamification will be crucial.

Given the relatively short duration of the implementation of the SVGs studied in this review, it may be a concern that the effects of gamification such as the participants' enjoyability, compliance, and improvements in psychopathology may be due in part to the novelty of the gamification experience [86,87]. This implies that the usefulness or effectiveness of an SVG can be lost over time as the novelty effect of gamification wears off. This may not be a significant problem if the gamified intervention is meant to be used only in the short term, but it will have to be taken into account if the serious game is expected to be part of a patient's long-term treatment. As discussed by Hamari and Kovisto [88], the addition of social networking and community elements to the gamification process may be helpful in enhancing effectiveness as well as the participants' willingness to use the service. These effects are positively related to positive recognition and reciprocity. Hence, it is plausible that implementation of SVGs in the setting of support groups for disordered eating could enhance and prolong their benefits, while promoting a sense of connectedness.

There may also be concerns over SVGs fueling other problems such as gaming addiction. This is because impulsivity traits such as urgency and lack of perseverance are both associated with eating disorders such as binge eating as well as with problems of addiction [89]. At the same time, problematic internet use has been shown to act as a predictor of eating disorders in a meta-analysis [90]. However, none of the SVGs (n=11) assessed in the review have reported addiction as an adverse effect. This is likely to be due to the short duration of the intervention, making addiction unlikely. However, this may have to be taken into consideration if the SVGs are implemented for a longer duration, and countermeasures such as additional supervision or imposing appropriate limits on the use of the games may be warranted.

Overall Utility and User Engagement

Taking the aforementioned considerations into account, although most of the outcomes measured in the individual papers were different, they collectively provide evidence to suggest the various roles SVGs can play in plugging potential gaps left by conventional therapy. Fitting the aforementioned factors into the framework developed by Liverpool et al [52] to evaluate user engagement, SVGs fare well in intervention-specific factors such as suitability, usability, and acceptability. In terms of person-specific influencing factors, young users likely have ample capability to engage in the use of SVGs. Furthermore, the novelty of SVGs also contributes to good user motivation, although its sustainability requires further examination. Opportunities for the adoption of SVGs may also be limited by confidentiality concerns as well as a lack of a sense of connectedness with others.

Strengths and Limitations

The strength of this study is that it provides a broad overview of the various ways in which gamification may be imbued in different types of interventions to address different psychopathological and clinical aspects of disordered eating; for instance, SVGs that address body dissatisfaction, which may be more prevalent in AN, and those that target inhibitory control, which may be associated more strongly with overeating, BED, or BN [7,48]. Furthermore, this study provides insight into the effectiveness of SVGs not only in eating disorders but also in disordered eating, which encompasses pathological eating behaviors beyond the strict diagnoses of eating disorders.

The limitations of this study are that only a small number of studies were included for analysis, providing limited amount of data and evidence for the assessment of the effectiveness of SVGs in disordered eating. Furthermore, the outcomes of the studies were varied, making it difficult to compare effect sizes among them. In addition, given the small sample sizes in the studies, there is a higher risk of bias and potential lack of representativeness. It was also observed that not all the included studies addressed populations who had received formal diagnoses of disordered eating but instead included participants from the general population or those who had other characteristics such as obesity, which may not be accurate proxies for disordered eating. Hence, it may be difficult to extrapolate the effects of the studied interventions to apply in populations with formal diagnoses of disordered eating behaviors. However, the studies in the review do show the promise that SVGs hold for effecting change in eating behaviors. Hence, it may be helpful for more studies to validate this effect on populations with disordered eating. It is also noted that the SVGs reviewed in this paper targeted a predominantly young population such as university students, with a large proportion of participants being women. As such, more evidence may be required to advance knowledge on the effectiveness of SVGs in disordered eating in other age groups as well as in men, given the knowledge that eating disorders in men also face significant underdiagnosis and undertreatment [15].

Further Studies and Recommendations

We have identified some gaps in current research that could be addressed in subsequent studies. First, more RCTs need to be conducted to expand the evidence pool. The World Health Organization's global strategy on digital health includes the recommendation that research is important in ensuring safety and accountability in the implementation of digital health interventions [91]. Hence, it is important to solidify our understanding of the efficacy of SVGs as a treatment intervention for disordered eating. Pending studies, such as the one pertaining to the Self-help, Integrated, and Gamified Mobile-Phone Application intervention for patients who were overweight with maladaptive eating behaviors, will add to our current understanding of gamified mobile apps for disordered eating [92]. Studies investigating the effectiveness of SVGs for disordered eating in different socioeconomic backgrounds, levels of computer literacy, genders, and ages may also be helpful.

As mentioned earlier, the longest study in this review was held only over 8 weeks. Hence, the long-term impact of the interventions and the retention of its beneficial effects might not be well understood and can be explored further. Learning about the long-term impact of SVGs in the context of disordered eating may then also provide clues to their potential role not only in treatment but also in prevention of disordered eating in persons with high-risk factors.

Some of the SVGs discussed were computer applications. However, in recent years, computer sales have been declining, and other devices such as smartphones are increasingly being used [93]. As such, it will be prudent for future studies to study games supported on such platforms, which promise greater accessibility.

Conclusions

In conclusion, this review supports the effectiveness of SVGs in complementing the current standards of care in treating disordered eating. Although more research is needed to gain a better understanding of the long-term effects of SVGs on disordered eating and their effectiveness across different demographic groups, SVGs have shown that they are able to provide measurable short-term benefits by addressing the underlying psychopathological processes that drive behaviors of disordered eating. Nonetheless, challenges exist in developing effective SVGs for disordered eating such that while having to keep the games appropriately challenging and engaging, caution has to be taken in the game design to prevent potential adverse outcomes.

Authors' Contributions

All authors had access to the data and a role in writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of included studies.
[DOCX File , 19 KB - jmir_v24i8e39527_app1.docx]

Multimedia Appendix 2
Risk-of-bias assessment according to the revised Cochrane Risk-of-Bias 2 tool for randomized controlled trials.
[PNG File , 163 KB - jmir_v24i8e39527_app2.png]

Multimedia Appendix 3
Risk-of-bias assessment according to the Joanna Briggs Institute Critical Appraisal Checklist for Quasi-Experimental Studies.
[DOCX File , 16 KB - jmir_v24i8e39527_app3.docx]

References

1. Arcelus J, Mitchell AJ, Wales J, Nielsen S. Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. Arch Gen Psychiatry 2011 Jul;68(7):724-731. [doi: 10.1001/archgenpsychiatry.2011.74] [Medline: 21727255]
2. Litmanen J, Fröjd S, Marttunen M, Isomaa R, Kaltiala-Heino R. Are eating disorders and their symptoms increasing in prevalence among adolescent population? Nord J Psychiatry 2017 Jan;71(1):61-66. [doi: 10.1080/08039488.2016.1224272] [Medline: 27626363]
3. Javaras KN, Runfola CD, Thornton LM, Agerbo E, Birgegård A, Norring C, et al. Sex- and age-specific incidence of healthcare-register-recorded eating disorders in the complete swedish 1979-2001 birth cohort. Int J Eat Disord 2015 Dec;48(8):1070-1081 [FREE Full text] [doi: 10.1002/eat.22467] [Medline: 26769444]
4. Westmoreland P, Krantz MJ, Mehler PS. Medical complications of anorexia nervosa and bulimia. Am J Med 2016 Jan;129(1):30-37. [doi: 10.1016/j.amjmed.2015.06.031] [Medline: 26169883]

5. Erzegovesi S, Bellodi L. Eating disorders. CNS Spectr 2016 Aug;21(4):304-309. [doi: 10.1017/S1092852916000304] [Medline: 27319605]
6. Reba-Harrelson L, Von Holle A, Hamer RM, Swann R, Reyes ML, Bulik CM. Patterns and prevalence of disordered eating and weight control behaviors in women ages 25-45. Eat Weight Disord 2009 Dec;14(4):e190-e198 [FREE Full text] [doi: 10.1007/BF03325116] [Medline: 20179405]
7. Aparicio-Martinez P, Perea-Moreno AJ, Martinez-Jimenez MP, Redel-Macías MD, Pagliari C, Vaquero-Abellan M. Social media, thin-ideal, body dissatisfaction and disordered eating attitudes: an exploratory analysis. Int J Environ Res Public Health 2019 Oct 29;16(21):4177 [FREE Full text] [doi: 10.3390/ijerph16214177] [Medline: 31671857]
8. Wu M, Hartmann M, Skunde M, Herzog W, Friederich H. Inhibitory control in bulimic-type eating disorders: a systematic review and meta-analysis. PLoS One 2013 Dec 31;8(12):e83412 [FREE Full text] [doi: 10.1371/journal.pone.0083412] [Medline: 24391763]
9. Treasure J, Russell G. The case for early intervention in anorexia nervosa: theoretical exploration of maintaining factors. Br J Psychiatry 2011 Jul;199(1):5-7. [doi: 10.1192/bjp.bp.110.087585] [Medline: 21719874]
10. Agras WS, Bohon C. Cognitive behavioral therapy for the eating disorders. Annu Rev Clin Psychol 2021 May 07;17:417-438. [doi: 10.1146/annurev-clinpsy-081219-110907] [Medline: 33962536]
11. Steinhausen HC. The outcome of anorexia nervosa in the 20th century. Am J Psychiatry 2002 Aug;159(8):1284-1293. [doi: 10.1176/appi.ajp.159.8.1284] [Medline: 12153817]
12. Wallier J, Vibert S, Berthoz S, Huas C, Hubert T, Godart N. Dropout from inpatient treatment for anorexia nervosa: critical review of the literature. Int J Eat Disord 2009 Nov;42(7):636-647. [doi: 10.1002/eat.20609] [Medline: 19208386]
13. van Hoeken D, Hoek HW. Review of the burden of eating disorders: mortality, disability, costs, quality of life, and family burden. Curr Opin Psychiatry 2020 Nov;33(6):521-527 [FREE Full text] [doi: 10.1097/YCO.0000000000000641] [Medline: 32796186]
14. Hudson JI, Hiripi E, Pope Jr HG, Kessler RC. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. Biol Psychiatry 2007 Feb 01;61(3):348-358 [FREE Full text] [doi: 10.1016/j.biopsych.2006.03.040] [Medline: 16815322]
15. Strother E, Lemberg R, Stanford SC, Turberville D. Eating disorders in men: underdiagnosed, undertreated, and misunderstood. Eat Disord 2012;20(5):346-355 [FREE Full text] [doi: 10.1080/10640266.2012.715512] [Medline: 22985232]
16. Ogg EC, Millar HR, Pusztai EE, Thom AS. General practice consultation patterns preceding diagnosis of eating disorders. Int J Eat Disord 1997 Jul;22(1):89-93. [doi: 10.1002/(sici)1098-108x(199707)22:1<89::aid-eat12>3.0.co;2-d] [Medline: 9140741]
17. Waller G. Treatment protocols for eating disorders: clinicians' attitudes, concerns, adherence and difficulties delivering evidence-based psychological interventions. Curr Psychiatry Rep 2016 Apr;18(4):36 [FREE Full text] [doi: 10.1007/s11920-016-0679-0] [Medline: 26893234]
18. Kaplan AS, Garfinkel PE. Difficulties in treating patients with eating disorders: a review of patient and clinician variables. Can J Psychiatry 1999 Sep;44(7):665-670. [doi: 10.1177/070674379904400703] [Medline: 10500867]
19. Sharpe H, Griffiths S, Choo TH, Eisenberg ME, Mitchison D, Wall M, et al. The relative importance of dissatisfaction, overvaluation and preoccupation with weight and shape for predicting onset of disordered eating behaviors and depressive symptoms over 15 years. Int J Eat Disord 2018 Oct;51(10):1168-1175 [FREE Full text] [doi: 10.1002/eat.22936] [Medline: 30194690]
20. Racine SE, Horvath SA. Emotion dysregulation across the spectrum of pathological eating: comparisons among women with binge eating, overeating, and loss of control eating. Eat Disord 2018;26(1):13-25. [doi: 10.1080/10640266.2018.1418381] [Medline: 29384463]
21. Jappe LM, Frank GK, Shott ME, Rollin MD, Pryor T, Hagman JO, et al. Heightened sensitivity to reward and punishment in anorexia nervosa. Int J Eat Disord 2011 May;44(4):317-324 [FREE Full text] [doi: 10.1002/eat.20815] [Medline: 21472750]
22. Fernàndez-Aranda F, Álvarez-Moya EM, Martínez-Viana C, Sànchez I, Granero R, Penelo E, et al. Predictors of early change in bulimia nervosa after a brief psychoeducational therapy. Appetite 2009 Jun;52(3):805-808. [doi: 10.1016/j.appet.2009.03.013] [Medline: 19501787]
23. Klump KL, Strober M, Bulik CM, Thornton L, Johnson C, Devlin B, et al. Personality characteristics of women before and after recovery from an eating disorder. Psychol Med 2004 Nov;34(8):1407-1418. [doi: 10.1017/s0033291704002442] [Medline: 15724872]
24. Agüera Z, Krug I, Sánchez I, Granero R, Penelo E, Peñas-Lledó E, et al. Personality changes in bulimia nervosa after a cognitive behaviour therapy. Eur Eat Disord Rev 2012 Sep;20(5):379-385. [doi: 10.1002/erv.2163] [Medline: 22367847]
25. Hirschtritt ME, Insel TR. Digital technologies in psychiatry: present and future. Focus (Am Psychiatr Publ) 2018 Jul;16(3):251-258 [FREE Full text] [doi: 10.1176/appi.focus.20180001] [Medline: 31975919]
26. Riva G, Malighetti C, Serino S. Virtual reality in the treatment of eating disorders. Clin Psychol Psychother 2021 May;28(3):477-488 [FREE Full text] [doi: 10.1002/cpp.2622] [Medline: 34048622]
27. Low TL, Ho R, Ho C, Tam W. The efficacy of virtual reality in the treatment of binge-purging eating disorders: a meta-analysis. Eur Eat Disord Rev 2021 Jan;29(1):52-59. [doi: 10.1002/erv.2804] [Medline: 33180339]

28. Manasse SM, Lampe EW, Juarascio AS, Zhu J, Forman EM. Using virtual reality to train inhibitory control and reduce binge eating: a proof-of-concept study. Appetite 2021 Feb 01;157:104988 [FREE Full text] [doi: 10.1016/j.appet.2020.104988] [Medline: 33049341]
29. Webb CA, Rosso IM, Rauch SL. Internet-based cognitive-behavioral therapy for depression: current progress and future directions. Harv Rev Psychiatry 2017;25(3):114-122 [FREE Full text] [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
30. Ter Huurne ED, de Haan HA, Postel MG, DeJong CA, VanDerNagel JE, van der Palen J. Long-term effectiveness of Web-based cognitive behavioral therapy for patients with eating disorders. Eat Weight Disord 2021 Apr;26(3):911-919. [doi: 10.1007/s40519-020-00929-0] [Medline: 32449152]
31. Abt CC. Serious Games. Lanham, MD, USA: University Press of America; 1987.
32. Laamarti F, Eid M, El Saddik A. An overview of serious games. Int J Comput Games Technol 2014 Oct;2014(3):1-15. [doi: 10.1155/2014/358152]
33. Djaouti D, Alvarez J, Jessel JP. Classifying serious games: the G/P/S model. In: Felicia P, editor. Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. Hershey, PA, USA: IGI Global; 2011.
34. Fitzgerald M, Ratcliffe G. Serious games, gamification, and serious mental illness: a scoping review. Psychiatr Serv 2020 Feb 01;71(2):170-183. [doi: 10.1176/appi.ps.201800567] [Medline: 31640521]
35. Ferguson C, van den Broek EL, van Oostendorp H. On the role of interaction mode and story structure in virtual reality serious games. Comput Educ 2020 Jan;143(C):103671. [doi: 10.1016/j.compedu.2019.103671]
36. Gaspar JD, Lage EM, Silva FJ, Mineiro É, Oliveira IJ, Oliveira I, et al. A mobile serious game about the pandemic (COVID-19 - did you know?): design and evaluation study. JMIR Serious Games 2020 Dec 22;8(4):e25226 [FREE Full text] [doi: 10.2196/25226] [Medline: 33301416]
37. Fleming TM, Cheek C, Merry SN, Thabrew H, Bridgman H, Stasiak K, et al. Serious games for the treatment or prevention of depression: a systematic review. Revista Psicopatol Psicol Clín 2014;19(3):227-242. [doi: 10.5944/rppc.vol.19.num.3.2014.13904]
38. Boendermaker WJ, Prins PJ, Wiers RW. Prevention in addiction: using serious games to (re)train cognition in Dutch adolescents. In: Proceedings of the 1st Joint International Conference on Serious Games. 2015 Presented at: JCSG '15; June 3-4, 2015; Huddersfield, UK p. 173-178. [doi: 10.1007/978-3-319-19126-3_15]
39. Kim JY, Bae JH. Analysis of serious games for preventing internet gaming addiction. Int J Inf Technol Manag 2018;17(1/2):62-70. [doi: 10.1504/ijitm.2018.10010508]
40. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
41. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
42. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS One 2017 Mar 31;12(3):e0173403 [FREE Full text] [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
43. Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J, Edwards T. Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. Comput Human Behav 2016 Jan;54:170-179. [doi: 10.1016/j.chb.2015.07.045]
44. Crepaldi M, Colombo V, Mottura S, Baldassini D, Sacco M, Cancer A, et al. The use of a serious game to assess inhibition mechanisms in children. Front Comput Sci 2020 Aug 25;2:1-11. [doi: 10.3389/fcomp.2020.00034]
45. Mena-Moreno T, Fernández-Aranda F, Granero R, Munguía L, Steward T, López-González H, et al. A serious game to improve emotion regulation in treatment-seeking individuals with gambling disorder: a usability study. Front Psychol 2021 Mar 4;12:621953 [FREE Full text] [doi: 10.3389/fpsyg.2021.621953] [Medline: 33746839]
46. İnce B, Schlatter J, Max S, Plewnia C, Zipfel S, Giel KE, et al. Can we change binge eating behaviour by interventions addressing food-related impulsivity? A systematic review. J Eat Disord 2021 Mar 18;9(1):38 [FREE Full text] [doi: 10.1186/s40337-021-00384-x] [Medline: 33736708]
47. Giel KE, Speer E, Schag K, Leehr EJ, Zipfel S. Effects of a food-specific inhibition training in individuals with binge eating disorder-findings from a randomized controlled proof-of-concept study. Eat Weight Disord 2017 Jun;22(2):345-351. [doi: 10.1007/s40519-017-0371-3] [Medline: 28271453]
48. Ames SL, Kisbu-Sakarya Y, Reynolds KD, Boyle S, Cappelli C, Cox MG, et al. Inhibitory control effects in adolescent binge eating and consumption of sugar-sweetened beverages and snacks. Appetite 2014 Oct;81:180-192 [FREE Full text] [doi: 10.1016/j.appet.2014.06.013] [Medline: 24949566]
49. 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]

50. Checklist for quasi-experimental studies (non-randomised experimental studies). Joanna Briggs Institute. 2020. URL: https://jbi.global/sites/default/files/2020-08/Checklist_for_Quasi-Experimental_Appraisal_Tool.pdf [accessed 2022-04-19]
51. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
52. Liverpool S, Mota CP, Sales CM, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
53. Fernandez-Aranda F, Jimenez-Murcia S, Santamaría JJ, Giner-Bartolomé C, Mestre-Bach G, Granero R, et al. The use of videogames as complementary therapeutic tool for cognitive behavioral therapy in bulimia nervosa patients. Cyberpsychol Behav Soc Netw 2015 Dec;18(12):744-751. [doi: 10.1089/cyber.2015.0265] [Medline: 26583754]
54. Kollei I, Lukas CA, Loeber S, Berking M. An app-based blended intervention to reduce body dissatisfaction: a randomized controlled pilot study. J Consult Clin Psychol 2017 Nov;85(11):1104-1108. [doi: 10.1037/ccp0000246] [Medline: 28857576]
55. Kosinski T. A brief mobile evaluative conditioning app to reduce body dissatisfaction? A pilot study in university women. Front Psychol 2019 Nov 21;10:2594 [FREE Full text] [doi: 10.3389/fpsyg.2019.02594] [Medline: 31824380]
56. Gledhill LJ, Cornelissen KK, Cornelissen PL, Penton-Voak IS, Munafò MR, Tovée MJ. An interactive training programme to treat body image disturbance. Br J Health Psychol 2017 Feb;22(1):60-76 [FREE Full text] [doi: 10.1111/bjhp.12217] [Medline: 27813208]
57. Joo YK, Kim K. When you exercise your avatar in a virtual game: the role of avatars' body shape and behavior in users' health behavior. Interact Comput 2017 May;29(3):455-466. [doi: 10.1093/iwc/iwx003]
58. Keeler JL, Chami R, Cardi V, Hodsoll J, Bonin E, MacDonald P, et al. App-based food-specific inhibitory control training as an adjunct to treatment as usual in binge-type eating disorders: a feasibility trial. Appetite 2022 Jan 01;168:105788 [FREE Full text] [doi: 10.1016/j.appet.2021.105788] [Medline: 34728250]
59. Dassen FC, Houben K, Van Breukelen GJ, Jansen A. Gamified working memory training in overweight individuals reduces food intake but not body weight. Appetite 2018 May 01;124:89-98. [doi: 10.1016/j.appet.2017.05.009] [Medline: 28479405]
60. Forman EM, Manasse SM, Dallal DH, Crochiere RJ, Berry MP, Butryn ML, et al. Gender differences in the effect of gamification on weight loss during a daily, neurocognitive training program. Transl Behav Med 2021 Apr 26;11(4):1015-1022 [FREE Full text] [doi: 10.1093/tbm/ibaa091] [Medline: 32945860]
61. Blackburne T, Rodriguez A, Johnstone SJ. A serious game to increase healthy food consumption in overweight or obese adults: randomized controlled trial. JMIR Serious Games 2016 Jul 13;4(2):e10 [FREE Full text] [doi: 10.2196/games.5708] [Medline: 27417192]
62. Forman EM, Manasse SM, Dallal DH, Crochiere RJ, Loyka CM, Butryn ML, et al. Computerized neurocognitive training for improving dietary health and facilitating weight loss. J Behav Med 2019 Dec;42(6):1029-1040 [FREE Full text] [doi: 10.1007/s10865-019-00024-5] [Medline: 30891657]
63. Verbeken S, Braet C, Goossens L, van der Oord S. Executive function training with game elements for obese children: a novel treatment to enhance self-regulatory abilities for weight-control. Behav Res Ther 2013 Jun;51(6):290-299. [doi: 10.1016/j.brat.2013.02.006] [Medline: 23524063]
64. Hofmann W, Gschwendner T, Friese M, Wiers RW, Schmitt M. Working memory capacity and self-regulatory behavior: toward an individual differences perspective on behavior determination by automatic versus controlled processes. J Pers Soc Psychol 2008 Oct;95(4):962-977. [doi: 10.1037/a0012705] [Medline: 18808271]
65. Fagundo AB, Santamaría JJ, Forcano L, Giner-Bartolomé C, Jiménez-Murcia S, Sánchez I, et al. Video game therapy for emotional regulation and impulsivity control in a series of treated cases with bulimia nervosa. Eur Eat Disord Rev 2013 Nov;21(6):493-499. [doi: 10.1002/erv.2259] [Medline: 24092598]
66. Giner-Bartolomé C, Fagundo AB, Sánchez I, Jiménez-Murcia S, Santamaría JJ, Ladouceur R, et al. Can an intervention based on a serious videogame prior to cognitive behavioral therapy be helpful in bulimia nervosa? A clinical case study. Front Psychol 2015 Jul 14;6:982 [FREE Full text] [doi: 10.3389/fpsyg.2015.00982] [Medline: 26236261]
67. Fagundo AB, Via E, Sánchez I, Jiménez-Murcia S, Forcano L, Soriano-Mas C, et al. Physiological and brain activity after a combined cognitive behavioral treatment plus video game therapy for emotional regulation in bulimia nervosa: a case report. J Med Internet Res 2014 Aug 12;16(8):e183 [FREE Full text] [doi: 10.2196/jmir.3243] [Medline: 25116416]
68. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? Dev Cogn Neurosci 2013 Apr;4:38-51 [FREE Full text] [doi: 10.1016/j.dcn.2012.11.002] [Medline: 23218805]
69. Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. Psychol Sci 2011 Apr;22(4):490-497 [FREE Full text] [doi: 10.1177/0956797611400615] [Medline: 21389338]
70. Schumacher SE, Kemps E, Tiggemann M. Bias modification training can alter approach bias and chocolate consumption. Appetite 2016 Jan 01;96:219-224. [doi: 10.1016/j.appet.2015.09.014] [Medline: 26375357]
71. Aspen V, Martijn C, Alleva JM, Nagel J, Perret C, Purvis C, et al. Decreasing body dissatisfaction using a brief conditioning intervention. Behav Res Ther 2015 Jun;69:93-99. [doi: 10.1016/j.brat.2015.04.003] [Medline: 25912670]

72. Yang Y, Shields GS, Wu Q, Liu Y, Chen H, Guo C. Cognitive training on eating behaviour and weight loss: a meta-analysis and systematic review. Obes Rev 2019 Nov;20(11):1628-1641. [doi: 10.1111/obr.12916] [Medline: 31353774]
73. Aulbach MB, Knittle K, Haukkala A. Implicit process interventions in eating behaviour: a meta-analysis examining mediators and moderators. Health Psychol Rev 2019 Jun;13(2):179-208. [doi: 10.1080/17437199.2019.1571933] [Medline: 30676235]
74. Distribution of video gamers in the United States in 2021, by age group. Statista. 2021. URL: https://www.statista.com/statistics/189582/age-of-us-video-game-players/ [accessed 2022-03-22]
75. Griffiths MD, Meredith A. Videogame addiction and its treatment. J Contemp Psychother 2009 May 12;39(4):247-253. [doi: 10.1007/s10879-009-9118-4]
76. Hainey T, Westera W, Connolly TM, Boyle L, Baxter G, Beeby RB, et al. Students' attitudes toward playing games and using games in education: comparing Scotland and the Netherlands. Comput Educ 2013 Nov;69:474-484. [doi: 10.1016/j.compedu.2013.07.023]
77. Hassouneh D, Brengman M. A motivation-based typology of social virtual world users. Comput Human Behav 2014 Apr;33:330-338. [doi: 10.1016/j.chb.2013.08.012]
78. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CA, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: the STEP UP randomized clinical trial. JAMA Intern Med 2019 Dec 01;179(12):1624-1632 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]
79. Fortunato M, Harrison J, Oon AL, Small D, Hilbert V, Rareshide C, et al. Remotely monitored gamification and social incentives to improve glycemic control among adults with uncontrolled type 2 diabetes (iDiabetes): protocol for a randomized controlled trial. JMIR Res Protoc 2019 Nov 20;8(11):e14180 [FREE Full text] [doi: 10.2196/14180] [Medline: 31746765]
80. Jakicic JM, Rogers RJ. Gamification and social incentives increase physical activity. Nat Rev Endocrinol 2020 Jan;16(1):10-12. [doi: 10.1038/s41574-019-0281-8] [Medline: 31654042]
81. Kurtzman GW, Day SC, Small DS, Lynch M, Zhu J, Wang W, et al. Social incentives and gamification to promote weight loss: the LOSE IT randomized, controlled trial. J Gen Intern Med 2018 Oct;33(10):1669-1675 [FREE Full text] [doi: 10.1007/s11606-018-4552-1] [Medline: 30003481]
82. 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]
83. Katz B, Jaeggi S, Buschkuehl M, Stegman A, Shah P. Differential effect of motivational features on training improvements in school-based cognitive training. Front Hum Neurosci 2014 Apr 24;8:242 [FREE Full text] [doi: 10.3389/fnhum.2014.00242] [Medline: 24795603]
84. 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]
85. Mekler ED, Brühlmann F, Tuch AN, Opwis K. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Comput Human Behav 2017 Jun;71(C):525-534. [doi: 10.1016/j.chb.2015.08.048]
86. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- a literature review of empirical studies on gamification. In: Proceedings of the 47th Hawaii International Conference on System Sciences. 2014 Presented at: HICSS '14; January 6-9, 2014; Waikoloa, HI, USA p. 3025-3034. [doi: 10.1109/hicss.2014.377]
87. Koivisto J, Hamari J. Demographic differences in perceived benefits from gamification. Comput Human Behav 2014 Jun;35:179-188. [doi: 10.1016/j.chb.2014.03.007]
88. Hamari J, Koivisto J. “Working out for likes”: an empirical study on social influence in exercise gamification. Comput Human Behav 2015 Sep;50:333-347. [doi: 10.1016/j.chb.2015.04.018]
89. Rømer Thomsen K, Callesen MB, Hesse M, Kvamme TL, Pedersen MM, Pedersen MU, et al. Impulsivity traits and addiction-related behaviors in youth. J Behav Addict 2018 Jun 01;7(2):317-330 [FREE Full text] [doi: 10.1556/2006.7.2018.22] [Medline: 29642723]
90. Hinojo-Lucena FJ, Aznar-Díaz I, Cáceres-Reche MP, Trujillo-Torres JM, Romero-Rodríguez JM. Problematic Internet use as a predictor of eating disorders in students: a systematic review and meta-analysis study. Nutrients 2019 Sep 09;11(9):2151 [FREE Full text] [doi: 10.3390/nu11092151] [Medline: 31505749]
91. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2022-08-19]
92. Podina IR, Fodor LA, Cosmoiu A, Boian R. An evidence-based gamified mHealth intervention for overweight young adults with maladaptive eating habits: study protocol for a randomized controlled trial. Trials 2017 Dec 12;18(1):592 [FREE Full text] [doi: 10.1186/s13063-017-2340-6] [Medline: 29233169]
93. Alsop T. PCs, tablets, ultra mobiles, mobile phones global shipments forecast 2013-2020. Statista. 2020. URL: https://www.statista.com/statistics/265878/global-shipments-of-pcs-tablets-ultra-mobiles-mobile-phones/ [accessed 2022-03-22]

Abbreviations

AN: anorexia nervosa
BED: binge eating disorder
BN: bulimia nervosa
CBT: cognitive behavioral therapy
EF: executive function
GNG: Go–No-Go
ICT: inhibitory control training
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SPARX: Smart, Positive, Active, Realistic, X-Factor Thoughts
SVG: serious video game
VR: virtual reality
WM: working memory

Edited by R Kukafka, G Eysenbach; submitted 13.05.22; peer-reviewed by H Mehdizadeh, J Gaspar, M Babashahi; comments to author 25.06.22; revised version received 18.07.22; accepted 29.07.22; published 29.08.22.

Please cite as:
Tang WSW, Ng TJY, Wong JZA, Ho CSH
The Role of Serious Video Games in the Treatment of Disordered Eating Behaviors: Systematic Review
J Med Internet Res 2022;24(8):e39527
URL: https://www.jmir.org/2022/8/e39527
doi:10.2196/39527
PMID:36036967

Original Paper

Effects of Substance Use, Recovery, and Non–Drug-Related Online Community Participation on the Risk of a Use Episode During Remission From Opioid Use Disorder: Longitudinal Observational Study

Elahe Naserianhanzaei[1], PhD; Miriam Koschate-Reis[2,3], PhD

[1]Department of Political Sciences, Faculty of Humanities, Arts and Social Sciences, University of Exeter, Exeter, United Kingdom
[2]Department of Psychology, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom
[3]Institute for Data Science and Artificial Intelligence, University of Exeter, Exeter, United Kingdom

Corresponding Author:
Elahe Naserianhanzaei, PhD
Department of Political Sciences, Faculty of Humanities, Arts and Social Sciences
University of Exeter
Amory Building, Rennes Drive
Exeter, EX4 4RJ
United Kingdom
Phone: 44 1392 72 5122
Email: e.naserianhanzaei@exeter.ac.uk

Abstract

Background: Opioid addiction is currently one of the most pressing public health issues. Despite several treatment options for opioid addiction, the recurrence of use episodes during remission remains high. Research indicates that meaningful membership in various social groups underpins the successful transition from addiction to long-term remission. However, much of the current literature focuses on online peer-support groups for individuals in remission from substance use, sometimes also called recovery groups, a term we will use in line with the terminology used by the online community we studied. In contrast, online group memberships that promote substance use and groups that are unrelated to substance use and remission (non–drug-related groups) are rarely studied.

Objective: This study aims to understand whether engagement with a variety of Reddit subforums (subreddits) provides those in remission from opioid use disorder (OUD) with social capital, thereby reducing their risk of a use episode over several years. More specifically, it aims to examine the different effects of engagement with substance use, recovery, and non–drug-related subreddits.

Methods: A data set of 457 individuals in remission from OUD who posted their remission start date on Reddit was collected, of whom 219 (47.9%) indicated at least one use episode during the remission period. Using a Cox proportional hazards model, the effects of the number of non–drug-related, recovery, and substance use subreddits an individual had engaged with on the risk of a use episode were tested. Group engagement was assessed both in terms of the absolute number of subreddits and as a proportion of the total number of subreddits in which an individual had posted.

Results: Engagement with a larger number of non–drug-related online communities reduced the likelihood of a use episode irrespective of the number of posts and comments made in these forums. This was true for both the absolute number of non–drug-related communities ($P<.001$) and the proportion of communities with which a person engaged ($P<.001$). The findings were less conclusive for recovery support and substance use groups; although participating in more recovery support subreddits reduced the risk of a use episode ($P<.001$), being part of a higher proportion of recovery support groups relative to other subreddits increased the risk ($P=.01$). A higher proportion of substance use subreddits marginally increased the risk of a use episode ($P=.06$); however, the absolute number of substance use subreddits significantly reduced the risk of a use episode ($P=.002$).

Conclusions: Our work indicates that even minimal regular engagement with several non–drug-related online forums may provide those in remission from OUD with an opportunity to grow their social capital and reduce the risk of a use episode over several years.

(J Med Internet Res 2022;24(8):e36555) doi:10.2196/36555

KEYWORDS
online communities; opioid addiction; recovery capital; social identity; Reddit; social media

Introduction

Background

In recent years, research on remission from addiction has shown the importance of social groups [1,2] and recovery capital [3] more widely. This trend builds on recent findings in the health and well-being literature, which suggest that the joining of new groups can act as a *social cure* [4]. The social cure hypothesis states that a higher number of memberships in social groups is associated with better mental and physical health outcomes, better resilience, and higher well-being. With regard to substance use disorders, much of the literature focuses on the benefits of peer-support groups for remission. Overall, this literature suggests that highly structured, peer-led support groups, both offline and online, support remission [5-10]. Few studies have examined the impact of social groups beyond such peer-support groups, and these tend to focus on a small number of offline groups or individuals (eg, family, friends, and coworkers) [11,12], thereby overlooking the potential of *online communities* as social capital for managing the remission process.

Understanding the impact of online community memberships beyond the direct effects of online peer-support groups can help tailor support for those with limited access to offline social networks. Access to a face-to-face support network can be limited by geographical location (eg, rural areas), difficulties with mobility (eg, lack of transport and failure of groups to accommodate mobility impairments), or significant caring responsibilities (eg, childcare availability) [13,14]. In some cases, the stigma surrounding addiction can also make it harder for those in remission to access offline help [15]. More recently, lockdowns during the COVID-19 pandemic have severely limited face-to-face interactions for those in remission, not only with peer-support groups but also with a wider support network of family, friends, and health care professionals [16]. Therefore, it is not surprising that online communities play an increasingly important role in providing social support, advice, and information [17,18] and may act as an additional source of social capital during remission.

Social Groups and Recovery

Overview

Research on the social cure provides mounting evidence that membership in several social groups increases health, well-being, and resilience [4,19]. For instance, research shows that a higher number of self-reported group memberships is associated with lower levels of smoking and drinking [20]. A recent study of online communities found that the number of subreddits an individual posted in and the evenness of participation in these different online communities reduced the risk of use episodes over several years for those in remission from opioid use disorder (OUD) [21]. Being part of various social groups is thought to provide members with psychological and physical resources such as a sense of connectedness, meaning, purpose, and worth as a member of a positively valued group [4]. Furthermore, groups also provide direct and indirect social support, a sense of personal control through the group's ability to affect change, and the social power and agency that the opportunity for collective action brings [4,22]. Similarly, recovery capital can be built by developing and strengthening links with those who are in recovery (*bonding*) and the wider community that may provide support (*bridging*) [3,23,24].

The Social Identity Model of Recovery (SIMOR) [1] integrates these 2 strands of research and suggests that meaningful membership in various social groups underpins the successful transition from addiction to long-term recovery. Rather than focusing on a few strong interpersonal bonds, the SIMOR suggests that being a part of several social groups is among the key factors associated with remission from substance use. Group membership becomes part of an individual's self-concept—their social identity—once the individual sees themselves, and is seen by others, as belonging to a particular social group or category [25,26]. When an individual identifies themselves as a group member, as well as begins to feel part of the group, group norms become internalized and guide attitudes, emotions, and behavior [27].

On the basis of this approach, being a member of a community that endorses substance use can be expected to increase substance use as this behavior is a normative expression of this particular group membership. In contrast, the model suggests that membership in recovery support groups shifts the identity away from an *addiction* identity toward a *recovery* identity. A *recovery* identity is thought to promote remission, such as a reduction in substance use or abstinence, through its norms and by providing (social) resources. In the longer term, gaining or regaining membership in groups that are not associated with substance use or remission (ie, non–drug-related groups) should help the individual build a social identity that increases health and well-being and is resilient to life changes [28,29]. Hence, the SIMOR conceptualizes remission from substance use disorder as a long-term transformation of social relationships and, correspondingly, the social self. The self is transformed from an *addiction* identity to a *recovery* identity toward a range of non–drug-related social identities (eg, parent, employee, and volunteer) that the individual gains or regains during their recovery journey [1].

Recovery Communities

Joining and participating in traditional support groups, such as Alcoholics Anonymous (AA) or Narcotics Anonymous, have shown significant promise in assisting individuals with a substance use disorder in maintaining their abstinence [30] by providing an encouraging and supportive community and by facilitating programs for addiction management and remission. An AA intervention designed to change an individual's social network—away from network members encouraging substance use toward abstinence-supporting network members—found positive effects on behavioral and attitudinal support for abstinence [5,31].

Similarly, positive effects on sustained remission have been found for online peer-support networks. For example, research found that engagement in online recovery support communities on Reddit reduced the risk of recurring use episodes in those who were in remission from smoking or from alcohol use disorders [8]. Similarly, a qualitative study showed that participation in the online recovery support group *Soberistas* was related to an offline commitment to changing drinking behaviors [32]. Studies on recovery from OUD show that online recovery groups are supportive [33] and promote remission despite a high chance of a recurring use episode among group members [34].

Active participation has been shown to be an important component [8]. For instance, a recent study found that those who participated more on a Facebook page dedicated to supporting those with a substance use disorder in the early stages of remission were more likely to remain in the program [9]. Here, the level of participation was assessed through the number of received *likes* and increased use of the word *we*.

Non–Drug-Related Communities

Further evidence from the field of addiction highlights the important role of a wider social network beyond recovery support groups. Much of this literature examines the effect of peer support on the decision to "quit" substance use and start therapy [35,36] or the reverse effect of the decision to quit on the composition of the social network [37]. However, a few studies have shown that social network composition affects subsequent substance use.

A longitudinal social network study across 32 years found that alcohol consumption tends to follow the behavior of individuals in a person’s social network [38]. These effects were mostly driven by interpersonal relationships with family members and close friends rather than groups (eg, coworkers or neighbors). The study also did not differentiate between problematic alcohol consumption and general alcohol consumption.

Furthermore, a self-report study with residents in a therapeutic community found that a higher proportion of non–drug-related group memberships decreased substance use at the 6-month follow-up [11]. In this study, the number of group memberships and categorization into non–drug-related (or *low-risk*) and substance use (or *high-risk*) groups were based on a mapping exercise. Participants were asked to group their social relations into different categories (eg, family, friends, coworkers, and recovery peer groups). Participants also indicated the number of people in each *group* who regularly used substances. Groups in which most individuals used substances regularly were then labeled as *high-risk* groups and those in which most of the members were abstinent, in remission, or whose drug use was unknown to the participant were then labeled as *low-risk* groups, with more mixed groups remaining uncategorized. As a result, recovery support groups were not considered in their own right but were classified as high risk, low risk, or uncategorized based on the perceived substance use of its members and not based on the norms of the group. This method also does not clearly differentiate between interpersonal relations (eg, family and friends) and group membership (eg, work team). Surprisingly, participants indicated only a very small number of *group memberships* (median 4). This suggests that small groups were mostly considered by participants rather than wider social categories (eg, parent and Christian) or shared interest groups (eg, volunteer and rugby fan) that form an individual’s social identity [27].

In summary, current research suggests that face-to-face interactions with individuals and groups who are not engaging in substance use are more likely to support remission than interactions with individuals who are known to continue to use substances. However, there appears to be very little, if any, research examining the effects of a wide variety of (online) communities that are concerned with non–drug-related interests, such as video games, literature, sports, and politics. Instead, current literature focuses on small groups and interpersonal social networks.

Substance Use Communities

In line with the SIMOR, a few studies indicate that a higher proportion of individuals engaging in substance use—or substance use groups—in an individual’s network tends to be related to higher substance use [11,38], although others have found no statistically significant effect [39]. However, these studies also indicate difficulties in neatly categorizing a group as a *substance use group* as opposed to a *non–drug-related group*. The use versus nonuse binary has been challenged by several researchers (eg, [40,41]) as adding a burden on those in remission. For instance, qualitative research suggests that severing ties with groups that continue to engage in substance use can result in the loss of trusted and emotionally significant relationships, particularly for young people [40]. Importantly, those who continue to engage in substance use do not necessarily seek to undermine remission in others and may provide information and emotional support. Similarly, the suggestion that the "bad company" of substance use groups undermines remission has been criticized as being based on relatively sparse empirical evidence [41]. Specifically, for online communities, we are not aware of any research examining the effect of participation in communities that actively promote or endorse substance use on the risk of recurring use episodes for those in remission.

In summary, there appears to be some empirical support for the positive effects of recovery support groups and offline non–drug-related groups on remission from substance use disorders. However, evidence is surprisingly sparse regarding the effects of groups that promote substance use and non–drug-related online interest groups. Looking closely at the research also reveals a mix of approaches in determining the type of group—often revealing a focus on the perceived behavior of known individuals rather than the norms that the wider social group is promoting. It is also unclear from the current literature whether it is close interpersonal network members (eg, family, friends, and coworkers) or interactions with wider communities (eg, parents, volunteers, sports fans, and Christians) that support remission. Finally, the current literature is limited to separately examining the effects of substance use, recovery, and non–drug-related groups.

Aim and Hypotheses

This study aimed to extend a previous study [21] on the effects of multiple online group memberships on remission from OUD. More specifically, this study tested the effects of substance use, recovery, and non–drug-related online communities separately in terms of absolute numbers of memberships and as a proportion of all memberships. Differentiating between the effects of different types of communities is both theoretically and practically important. In addition to testing the wider social cure hypothesis, this approach will test the SIMOR, which suggests that substance use groups affect remission differently from recovery and nonusing groups. Although several studies have tested parts of the SIMOR, only one study that we are aware of has tested all 3 types of groups together [11], and none has done so in an online environment. Understanding how different types of online communities affect remission also has practical implications by allowing those in remission to make more informed choices about their online community engagement. Furthermore, by examining data across several years, the study also sheds light on the longitudinal effects that these 3 different types of social groups have on the risk of a use episode over time, thereby considerably extending the knowledge about the long-term effects of social group membership.

In line with the SIMOR, we expected that a higher number (and proportion) of online community memberships actively supporting recovery will decrease the risk of a use episode over time (*hypothesis 1*). Similarly, a higher number (and proportion) of non–drug-related online groups were predicted to reduce the risk of a use episode (*hypothesis 2*). In contrast, a higher number (and proportion) of memberships in online communities that advocate substance use were expected to increase the risk of a use episode in those in remission from OUD (*hypothesis 3*).

Methods

Data Selection

The online forum platform Reddit provides a unique opportunity to study the effects of membership and activity patterns in various online communities on the risk of recurring use episodes [8,18,21]. Reddit is a public social news and discussion website where user-created content is organized into topic-based boards called *subreddits*. It accommodates >100,000 different subreddits where content is shared within a community of interest, such as about politics, business, parenting, medical conditions, sports, literature, music, video games, and life choices. This abundance of social groups allows us to simultaneously examine the effects of substance use, recovery, and non–drug-related online communities on the risk of a use episode during remission.

The subreddit r/OpiatesRecovery has >31,000 members and has provided individuals who wish to recover from OUD with recovery information and peer support since 2012. Some members make statements about their remission status on a regular basis by announcing the number of days they have not been using opioids. In combination with statements announcing a use episode, this provides us with information about the point in the recovery journey when a community member had a recurring use episode. Furthermore, anonymous but unique user IDs allow us to gather data on community members' activities across the entire Reddit platform, providing behavioral information on engagement with online communities. Using behavioral information, problems with memory bias and differences in the definition of what is meant by a *group membership* can be circumvented. This is particularly important in light of the recent finding that online interactions are not spontaneously self-reported as social contact [42].

Ethics Approval and Privacy Considerations

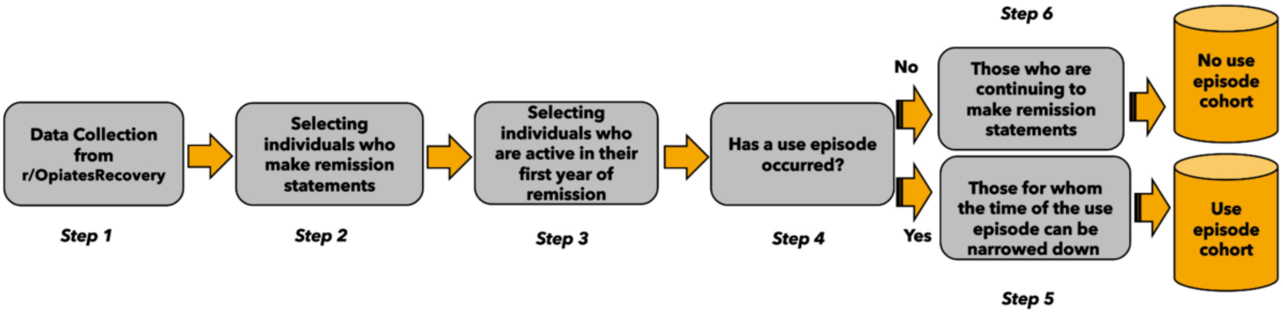
Before work commenced, the study received ethics approval from the University of Exeter's institutional review board (eCLESPsy001576), in line with the guidelines of the British Psychological Society and the American Psychological Association. This study used publicly accessible Reddit data for the analysis. Reddit usernames were used to collect quantitative data across different subreddits and were then replaced with anonymous participant numbers in the working data set. In this study, we do not report any user identifiable information to protect user privacy (eg, direct quotations and usernames); instead, we paraphrase quotations to illustrate our method.

Data Collection and Preparation

The steps involved in the data collection and preparation processes are illustrated in Figure 1. The study used publicly accessible data from r/OpiatesRecovery between February 2012 (when the group started) and June 2019 (the month of data collection).

After cleaning the data from adversarial content, bots, and so on, the initial data set contained 295,232 posts and comments from 18,125 individuals (step 1). Among these posts, we identified statements in which individuals announced their remission status (eg, "I'm 21 days clean") and found 2950 instances of remission being announced by 1651 individuals (step 2). As we wished to study the activity of those in remission from OUD from the beginning of their current remission period, individuals whose self-reported time in remission exceeded their time on Reddit by >1 year were excluded from the analyses (eg, an individual who joined in 2015 but said in a post that they have been in remission since 2012). To allow sufficient information about each individual in our data set, we only retained data from those who posted at least 10 posts and comments in any subreddit for a minimum of 3 different months during the first year of remission. This resulted in 1081 individuals (step 3).

Figure 1. Diagram of data collection and preparation for identifying use episode and no use episode cohorts.



To identify which individuals had experienced a use episode, we examined statements commonly used to report such an episode (eg, "I started using again last night"). As we aimed to include as wide a sample as possible, we used inconsistent reporting of abstinence to identify those with a use episode. Where an inconsistent number of days without a use episode was reported (eg, "5 days without using" after announcing "21 days clean" a week earlier), a trained psychologist individually inspected remission status statements and found 51 additional individuals with a likely use episode. On the basis of statements and inconsistent reporting, we identified a *use episode* cohort of 335 individuals (step 4) among the 1081 individuals that regularly announced their remission progress. To ensure that survival analysis can estimate time correctly, we excluded those individuals for whom the time of the use episode could not be narrowed down to a specific month. For individuals whose remission statements only indirectly indicated a use episode, we used the month between the last report of a consistent period without a use episode and the announcement of a new period of remission (step 5). Individuals who had a consistent increase in remission days and did not make any statements about a use episode were grouped into the *no use episode* cohort. To ensure *noninformative* censoring, we only kept data from individuals in the *no use episode* cohort who were still active in r/OpiatesRecovery by the end of our data collection period in June 2019 (step 6). In total, our sample comprised 457 individuals from r/OpiatesRecovery, of whom 219 (47.9%) reported a use episode during remission, and 238 (52.1%) did not report a use episode.

Labeling of Online Communities

To study the impact of distinct types of online groups on recovery progress, we categorized them into 3 types in line with the literature on the SIMOR: substance use, recovery, and non–drug-related groups. We considered subreddits that actively promoted substance use or discussed safe use as *substance use groups* (eg, r/Drugs, r/trees, and r/fentanyl). These also included legal substances such as alcohol consumption and tobacco smoking (eg, r/Cigarette, r/juul, and r/alcohol). Groups that discussed and supported recovery from addiction to any substance were considered recovery support communities (eg, r/narcoticsanonymous, r/quittingsmoking, and r/leaves). Both the description of the subreddit and posts within the community were considered when categorizing. Overall, 2 research assistants labeled the data, with the 2 authors providing guidance and reviewing the labels.

The participants in our data set were part of >4500 subreddits, which makes it impractical to label all of them. Therefore, we used a snowball approach where we started with r/Opiates and r/OpiatesRecovery, found similar subreddits to each of these, and then identified their category based on the content posted on the subreddit. We followed the same procedure for the newly identified subreddits, limiting our search to 4 rounds. We used Sayit [43], an online tool for finding subreddits similar to a specific one based on the number of mutual members. This procedure led to 1247 subreddits. Of these 1247 subreddits, 151 (12.11%) were labeled as substance use communities, 38 (3.05%) were labeled as recovery support communities, and 1058 (84.84%) were labeled as non–drug-related communities. Given that non–drug-related subreddits had a much higher prevalence, although we explicitly searched for subreddits on substance use and recovery support, we tested whether those who had not been labeled could be considered non–drug-related communities. To do so, we randomly selected 500 subreddits that had so far not been labeled and categorized them into substance use, recovery, and non–drug-related groups. Only 3.2% (16/500) were found to be substance use or recovery support groups. Therefore, it was decided that the risk of categorizing the remaining unlabeled forums as non–drug-related groups to the quality of the data was negligible.

To assess the effect of different types of groups on remission outcomes, researchers used 2 different operationalizations: the absolute number of groups of a particular type [20] and the proportion of a particular type of group relative to the total number of groups [11]. Assessing the absolute number of group memberships is in line with much of the social cure research that tends to find that a larger absolute number of group memberships is beneficial to health, mental health, and resilience [4]. In contrast, the SIMOR suggests that recovery from addiction is affected by the balance between substance use, recovery, and non–drug-related groups (ie, the proportion). Here, we use both operationalizations to examine the SIMOR and the wider social cure hypothesis.

For each participant, we counted the number of different subreddits in which they had posted at least once for each month from their self-reported remission start date until the first use episode or the last remission status announcement. On the basis of a labeled list of subreddits, we assessed the number of substance use, recovery, and non–drug-related group memberships for each month. In addition, we calculated the proportion of these group memberships by dividing membership into 1 of 3 categories (substance use, recovery, or non–drug-related) by the total number of group memberships in the same month (eg, number of substance use groups divided by number of all groups an individual has posted in). For each month, we also collected the number of posts and comments that a participant had contributed to the different online communities. Summary statistics for the sample are provided in Table 1, showing that participation in online communities far outstrips the number of groups that individuals who are in remission from a substance use disorder report in the offline world [11].

Table 1. Summary statistics of no use episode and use episode cohorts (N=457).

	All	No use episode	Use episode
Total sample, n (%)	457 (100)	238 (52.1)	219 (47.9)
Total posts, n (%)	237,435 (100)	137,030 (57.7)	100,405 (42.3)
Total subreddits, n (%)	4582 (100)	3513 (76.7)	2268 (49.5)
Posts per individual, median (range)	158 (1-15,805)	157 (1-15,805)	160 (3-7184)
Subreddits per individual, median (range)	14 (1-532)	18 (1-522)	11 (1-532)

Data Analysis

The availability of longitudinal data up to a maximum of 6 years in an individual's remission journey allowed us to predict the risk of a use episode as an outcome variable in a survival analysis (the step-by-step R code for the survival analysis is provided in Multimedia Appendix 1). The absolute number of substance use, recovery, and non–drug-related communities that an individual posted in at least once, as well as the proportion of these communities, served as predictor variables in our models. The number of posts and comments that an individual provided to these communities was statistically controlled for. This allowed us to test whether it is the number of group memberships, or an individual's posting behavior, that is related to the risk of a use episode.

Survival analysis is a type of time-to-event analysis that has been widely adopted when the research interest is a combination of *whether* the event has occurred (binary outcome) and when it has occurred (continuous outcome). It provides unbiased survival estimates by using the information provided by individuals who have experienced the event (here, a use episode), as well as by those who have not (here, no use episode; so-called censored data). To explore the effects of various factors on the time to relapse, we used an extended Cox model [44]. This survival analysis regression method explores the relationship between the event of interest (the use episode in our study) and factors that affect the time at which the event occurs. This allowed us to study how survival probabilities change with changes in the studied factors. Unlike the basic Cox model, the extended version is designed to accommodate time-dependent variables; that is, variables whose values for a given participant may differ over time. Cox modeling does not make any assumptions about the statistical distribution of survival times, unlike most other statistical models, which makes it an appropriate choice for our research problem.

The data used in the study, the activity of individuals in terms of Reddit group membership over time, is provided as a supplementary file in Multimedia Appendix 1.

Results

Preliminary Results

Summary statistics (Table 2) of the number of substance use, recovery, and non–drug-related groups that individuals participated in at least once show that participation in non–drug-related groups is common among individuals in remission from OUD. They also indicate that those in the *use episode* cohort show significantly less engagement with non–drug-related groups than those in the *no use episode* cohort. Interestingly, there were no significant differences in the number of substance use and recovery groups, respectively, between the *no use episode* and the *use episode* cohorts.

Table 2. Summary statistics and significance tests for online community types by use episode and no use episode cohorts.

Communities	No use episode, median (range)	Use episode, median (range)	Significance test	
			U test	*P* value
Substance use	1 (0-10)	0 (0-28)	24,414.00	.20
Recovery	2 (1-11)	2 (1-11)	27,946.50	.16
Non–drug-related	16 (0-507)	8 (0-493)	22,036.00	.004

Absolute Number of Group Memberships

In line with our hypotheses, we tested whether the absolute number of memberships in substance use, recovery, and non–drug-related groups affected the risk of a use episode during OUD remission. In line with SIMOR, we expect recovery and non–drug-related groups to decrease the risk of a use episode over time (hypotheses 1 and 2) and substance use groups to increase the risk of a use episode (hypothesis 3). The number of posts within these communities was controlled statistically. The statistics for the 3 separate survival analyses and the combined analysis are presented in Table 3.

Table 3. Effects of the absolute number of memberships in recovery, non–drug-related, and substance use online forums on the risk of a use episode during opioid use disorder remission.

Variable	*b* (SE)	*P* value	Odds ratio (95% CI)
Separate models			
Recovery models			
Number of posts (recovery)	0.002 (0.0008)	.04	1.0017 (5.61×10^{-5} to 0.003)
Recovery memberships	−0.20 (0.08)	.009	0.8182 (−0.352 to −0.049)
Non–drug-related			
Number of posts (non–drug-related)	0.0004 (0.0002)	.005	1.0004 (0.0001 to 0.0006)
Non–drug-related memberships	−0.05 (0.008)	<.001	0.9499 (−0.066 to −0.036)
Substance use models			
Number of posts (substance use)	0.002 (0.002)	.11	1.0020 (−0.0006 to 0.0046)
Substance use memberships	−0.22 (0.07)	.002	0.8024 (−0.361 to −0.078)
Combined model			
Number of posts (total)	0.0004 (0.0001)	.001	1.0004 (0.0001 to 0.0007)
Recovery memberships	0.02 (0.07)	.73	1.02 (−0.108 to 0.154)
Non–drug-related memberships	−0.06 (0.01)	<.001	0.94 (−0.073 to −0.040)
Substance use memberships	0.08 (0.05)	.12	1.08 (−0.020 to 0.172)

In line with hypothesis 1, survival analysis showed that the number of online recovery groups an individual is a part of is negatively and significantly related to the risk of a use episode (*P*=.01), irrespective of the number of posts. Similarly, the higher the number of memberships in non–drug-related online groups, the lower the risk of a use episode (*P*<.001), supporting hypothesis 2. However, in contrast to hypothesis 3, we found that the higher the absolute number of substance use group memberships, the lower (rather than higher) the risk of a use episode over time. The individual models also showed that the number of posts or comments made in each type of forum tends to increase the risk of a use episode. This effect was statistically significant for recovery and non–drug-related online forums.

When all 3 predictors were included in the same survival analysis, only the effect of the absolute number of non–drug-related online communities remained significant. Membership in non–drug-related groups has significant potential for reducing the risk of a use episode by 6% per additional non–drug-related group that an individual joins. Importantly, this effect is found when controlling for the number of posts or comments made, showing that the effect is not because of more active engagement in online communities. In fact, the total number of posts or comments was significantly and positively related to the occurrence of a use episode during remission, indicating a higher risk for those who contributed more frequently.

Proportion of Group Memberships

Next, we tested the same 3 hypotheses but with the number of group memberships in a particular type of group (substance use, recovery, or non–drug-related) relative to the number of total Reddit communities in which an individual participated (ie, the proportion). Again, we statistically controlled for the total number of posts or comments that contributed to the respective type of group. The statistics for the 3 separate survival models are presented in Table 4.

Survival analysis showed an unexpectedly positive and significant effect of the proportion of recovery group memberships on the risk of relapse. Being part of a higher proportion of online recovery subreddits significantly increases the likelihood of an individual reporting a use episode over time (*P*<.001). This finding does not support hypothesis 1. The model also showed a marginally significant positive effect of the proportion of substance use groups on the risk of a use episode (*P*=.06). In line with hypothesis 3, this finding suggests that a higher proportion of substance use groups may increase the risk of a use episode during remission. In contrast, being part of a higher proportion of non–drug-related groups significantly decreases the risk of a use episode (*P*<.001), in line with hypothesis 2.

Table 4. Effects of the proportion of memberships in recovery, non–drug-related, and substance use online forums on the risk of a use episode during remission.

Variable	*b* (SE)	*P* value	Odds ratio (95% CI)
Recovery models			
Number of posts (recovery)	0.001 (0.001)	.29	1.00 (−0.0008 to 0.0027)
Recovery memberships (%)	1.61 (0.20)	<.001	5.00 (1.219 to 2.002)
Non–drug-related models			
Number of posts (non–drug-related)	−0.001 (0.001)	.30	1.00 (−0.003 to 0.0009)
Non–drug-related memberships (%)	−1.39 (0.26)	<.001	0.25 (−1.900 to −0.879)
Substance use models			
Number of posts (substance use)	−0.004 (0.002)	.38	1.00 (−0.011 to 0.004)
Substance use memberships (%)	1.16 (0.60)	.06	3.20 (−0.062 to 2.387)

Discussion

Principal Findings

Using naturally occurring data from the popular online platform Reddit, we examined the risk of a use episode over several years of remission from OUD. Our study tested predictions by the SIMOR that recovery and non–drug-related group memberships sustain remission, whereas substance use groups undermine it. Here, we specifically tested the SIMOR in an *online* environment to provide empirical evidence for the effects of online community memberships that go beyond membership in a single online recovery support group.

Our findings show that a higher number of memberships in non–drug-related online groups are associated with a lower risk of use episodes during OUD remission. The more online non–drug-related groups an individual recovering from OUD becomes part of, the lower the risk of a use episode over time. This effect persisted when membership in recovery and substance use groups was accounted for. These findings support the SIMOR's focus on building non–drug-related group memberships to sustain remission [1]. They also support predictions by the wider social cure literature, which suggests that social groups deliver health benefits and create resilience [4].

In contrast to much of the literature, we found mixed support for membership in recovery support groups. The results indicate that a higher absolute number of memberships in recovery support groups is associated with a reduction in the risk of a use episode but only when other types of group membership are not controlled for. Furthermore, we found a significant increase in the risk for those with a higher proportion of recovery groups among all their Reddit groups.

We offer 2 explanations for this result. First, engaging with several recovery support forums unrelated to opioid remission may indicate that the individual is dealing with polysubstance dependence or has been using multiple substances to self-medicate health issues (eg, chronic pain) or underlying mental health problems (eg, depression and anxiety). Recovering from multiple substances in addition to opioids may increase the chances of a use episode because of the higher risk to mental and physical health [45]. However, this explanation cannot account for the finding that the absolute number of recovery support group memberships reduced the risk of a use episode (or was unrelated to risk when other group memberships were accounted for). A second possible explanation is that a narrow focus on online recovery support groups may be detrimental to sustained remission when online activities do not include participation in non–drug-related groups. Investing primarily in recovery support groups (ie, having a high proportion of recovery support group memberships relative to other memberships) appears to undermine building a resilient social self that incorporates a variety of social identities derived from valued group membership in different spheres of life. This points toward a need for future research to examine the extent to which membership in (online) recovery support groups may reduce engagement with non–drug-related communities and the effects of this on the risk of recurring use episodes and other remission outcomes. Examining the effects of recovery support group membership in isolation may risk missing wider, potentially detrimental effects.

Furthermore, our analysis uses data across several years, thereby examining a longer period of recovery than is usually investigated in the literature, which tends to focus on 6 months to 1 year after treatment. Therefore, it is possible that our findings reflect the importance of non–drug-related identities that replace a *recovery* identity in the long term, as suggested by the SIMOR [1]. Hence, our work provides the first indication that the recovery journey needs to continue beyond building recovery support group memberships and that online communities can play a part in providing a diverse range of non–drug-related group memberships.

We also found inconsistent support for the proposed detrimental effect of substance use groups on the risk of a use episode during remission. Memberships in online communities that promote substance use did not show the expected effect of increasing the risk of a use episode. Instead, we found a significant negative effect of the absolute number of substance use communities on the risk of a use episode. This effect disappeared when other group memberships were controlled for. However, we also found a marginally significant effect that a higher proportion of substance use community memberships increases the likelihood of a use episode. Together, this inconsistent pattern does not unequivocally support the assumption by the SIMOR that substance use group memberships undermine long-term remission.

This finding is surprising as substance use groups on Reddit share pictures of drugs and paraphernalia. Such pictures have been linked to cravings in individuals with substance use disorder [46]. However, our study did not exclusively focus on substance use groups linked to opioids. Those who participate in substance use communities may seek alternative treatments to avoid opioid use [47]. Furthermore, some researchers have challenged the assumption by the SIMOR, as well as similar models, that substance use groups are necessarily detrimental. For instance, qualitative research has shown that groups (and individuals) associated with continuing substance use can provide significant recovery support through information [48] and trusted and emotionally significant friendships [40]. Future research is needed to provide a clearer, as well as a more differentiated, picture of substance use group memberships and their part in the recovery journey.

Importantly, we controlled for posting activity throughout to test whether the number of group memberships or the activity level drives effects. Interestingly, posting activity was positively related to the risk of a use episode, suggesting that those who posted, or commented, more frequently were more likely to report a use episode. Surprisingly, this effect was significant for both recovery and non–drug-related online communities, indicating that this effect was not because of an increased discussion of drug-related issues. Future research should examine whether posting frequency is an early warning sign of recurring use episodes or whether other variables related to posting activities act as risk factors.

Limitations and Future Research

To test the SIMOR and extend previous research on the effects of online community memberships on the risk of a use episode during OUD remission [21], we categorized subreddits into substance use, recovery, and non–drug-related groups. Our discussion of the effects of recovery groups and substance use groups indicates that such a classification might be too simplistic to capture the effects of specific groups on the risk of recurring use episodes. The supposedly neat categorization into these 3 types of groups by the SIMOR has already been criticized elsewhere [40,41,49]. Although such labeling of groups provides some benefits in terms of public health messaging, it may not adequately reflect the complexities on the ground.

Another limitation is the lack of demographic information and other relevant information in our sample. As a result, we cannot statistically control for potential confounding variables such as previous opioid use severity, treatment uptake, comorbidity, and other relevant variables. For instance, individuals who manage to join more online groups during their remission from OUD may have personal skills (eg, self-confidence) and privileges (eg, free time and unrestricted internet access) that allow them to engage with more online communities.

The sample size of our study was smaller than those of other studies that used computational approaches. However, the sample size is still larger than that of the vast majority of studies in the area of social group effects on remission outcomes. Furthermore, sample size is not an indicator of the validity of the findings. It is increasingly being recognized that an important challenge when using online data for clinical or social research questions is that validity is established across all components: a high-quality sample, valid measures, and appropriate statistical methods. As outlined in the *Methods* section, the sample size was determined by methodologically justified steps to ensure valid conclusions. For instance, rather than including all individuals who posted in r/OpiatesRecovery and did not report a use episode into the *no use episode* cohort, we ensured that we had a clear indication from remission status statements that no use episode had occurred where inconsistent remission status statements were made, and a trained psychologist reviewed the case. Similarly, we invested time in labeling a large number of subreddits not only based on their name or mission statement but also by taking the posts in the subreddits into account.

Importantly, our research found that the activity level in communities was unrelated to the risk of a use episode or increased risk rather than lowering it. However, we do not have data for individuals who are entirely passive members of a community; that is, those who read posts but never or rarely post themselves (sometimes referred to as *lurkers*). The CEO of Reddit, Steve Huffman, estimates that two-thirds of the members may fall into this category [50]. Recent research shows that active participation in face-to-face groups is not an attractive or viable option for everyone [14]. Therefore, understanding whether *lurking* is enough to provide a person with a sense of social identity and group resources (eg, advice and shared experience) is a relevant (although technically challenging) avenue for future research.

Practical Implications

Online communities provide a new methodological way of studying remission from substance use disorders. The public and anonymous nature of online communities allows us to examine longitudinal data of those who are in remission over several years, including data from individuals that may be difficult to recruit for traditional surveys or interview studies (eg, those who do not access therapy through the health care system). Much of the literature on social group effects on remission relies on cross-sectional data, with only a few studies testing longitudinally, rarely for more than 6 months or a year at best. Sample sizes tend to be small (often N=20-50 for smaller, cross-sectional studies, with N=200-350 for large studies) and often stem from therapeutic communities or peer-support groups (eg, AA or Narcotics Anonymous), thereby limiting findings to those who already successfully access formal support. Examining online community behavior, such as participation in different online forums, also allows us to observe real-world behavior rather than having to rely on the memory and subjective definitions of participants in survey studies and interviews.

Online communities have the potential to benefit those who are in remission from substance use disorder by providing relatively easy access to social groups. Our findings suggest that fairly minimal engagement with non–drug-related groups can increase resilience during recovery over a long period. Although online communities are not universally accessible because of the requirement for an internet connection and basic literacy and computer skills, they create fewer barriers than many offline groups [14]. They may also provide a sense of continuity to those who need to move their location.

Importantly, our research indicated that an exclusive or narrow focus on recovery support communities may increase the risk of recurring use episodes. Platforms and mutual help communities that offer online support to those in substance use remission but are not part of a wider platform (eg, Reddit) may wish to flag non–drug-related online communities to their members to enable them to build wider social group memberships and avoid a narrow focus on recovery support groups, particularly in the longer term.

Conclusions

Opioid addiction is one of the most pressing public health issues of the day and was declared a national health emergency by the US government in 2017 [51], with an average of 128 overdose deaths from opioids every day in the United States alone [52]. Worldwide, 118,000 deaths in 2015 were directly associated with OUD [53]. Access to therapy and recovery support groups can prove difficult for several reasons, such as a lack of funding, stigma, and personal circumstances. Online groups, such as forums on popular platforms like Reddit, may provide a lifeline for those who are in remission from OUD. Providing evidence-based support for the use of online groups during OUD remission is an important public health task. Here, we provide the first evidence that online forums that are unrelated to substance use and recovery advice can provide social capital that significantly reduces the risk of a use episode across several years significantly—by as much as 6% per additional non–drug-related online community. Our results also suggest that more research is needed to understand the circumstances under which ties with substance use groups may not pose a risk to sustained remission from OUD. Similarly, further research is needed to understand the circumstances under which a narrow focus on online recovery support groups may be harmful during OUD remissions.

Acknowledgments

This research was supported by funding from the Engineering and Physical Sciences Research Council, the United Kingdom, to the second author (EP/S001409/1 and EP/R013144/1). The authors would like to thank Dr Sam Ellis, Jodie Hall, and Rada Biyukova for their contribution to this study. A related manuscript was published at the Association for Computing Machinery on Human-Computer Interaction in 2021.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection and survival analysis.
[DOCX File , 10 KB - jmir_v24i8e36555_app1.docx]

References

1. Best D, Beckwith M, Haslam C, Alexander Haslam S, Jetten J, Mawson E, et al. Overcoming alcohol and other drug addiction as a process of social identity transition: the social identity model of recovery (SIMOR). Addict Res Theory 2015 Aug 18;24(2):111-123. [doi: 10.3109/16066359.2015.1075980]
2. Frings D, Albery IP. The Social Identity Model of Cessation Maintenance: formulation and initial evidence. Addict Behav 2015 May;44:35-42. [doi: 10.1016/j.addbeh.2014.10.023] [Medline: 25468675]
3. Cloud W, Granfield R. Conceptualizing recovery capital: expansion of a theoretical construct. Subst Use Misuse 2008 Jul 03;43(12-13):1971-1986. [doi: 10.1080/10826080802289762] [Medline: 19016174]
4. Haslam C, Jetten J, Cruwys T, Dingle G, Haslam S. The New Psychology of Health, Unlocking the Social Cure. London: Routledge; 2018.
5. Litt MD, Kadden RM, Kabela-Cormier E, Petry NM. Changing network support for drinking: network support project 2-year follow-up. J Consult Clin Psychol 2009 Apr;77(2):229-242 [FREE Full text] [doi: 10.1037/a0015252] [Medline: 19309183]
6. Timko C, Debenedetti A, Billow R. Intensive referral to 12-Step self-help groups and 6-month substance use disorder outcomes. Addiction 2006 May;101(5):678-688 [FREE Full text] [doi: 10.1111/j.1360-0443.2006.01391.x] [Medline: 16669901]
7. Humphreys K, Barreto NB, Alessi SM, Carroll KM, Crits-Christoph P, Donovan DM, et al. Impact of 12 step mutual help groups on drug use disorder patients across six clinical trials. Drug Alcohol Depend 2020 Oct 01;215:108213 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108213] [Medline: 32801112]
8. Tamersoy A, Chau D, De Choudhury M. Analysis of smoking and drinking relapse in an online community. In: Proceedings of the 2017 International Conference on Digital Health. 2017 Presented at: DH '17: International Conference on Digital Health; Jul 2 - 5, 2017; London United Kingdom. [doi: 10.1145/3079452.3079463]

9. Bliuc A, Best D, Iqbal M, Upton K. Building addiction recovery capital through online participation in a recovery community. Soc Sci Med 2017 Nov;193:110-117. [doi: 10.1016/j.socscimed.2017.09.050] [Medline: 29032310]
10. Bliuc A, Doan T, Best D. Sober social networks: the role of online support groups in recovery from alcohol addiction. J Community Appl Soc Psychol 2018 Nov 12;29(2):121-132. [doi: 10.1002/casp.2388]
11. Beckwith M, Best D, Savic M, Haslam C, Bathish R, Dingle G, et al. Social Identity Mapping in Addiction Recovery (SIM-AR): extension and application of a visual method. Addict Res Theory 2019 Jan 16;27(6):462-471. [doi: 10.1080/16066359.2018.1544623]
12. Bathish R, Best D, Savic M, Beckwith M, Mackenzie J, Lubman DI. “Is it me or should my friends take the credit?” The role of social networks and social identity in recovery from addiction. J Appl Soc Psychol 2017 Jan 20;47(1):35-46. [doi: 10.1111/jasp.12420]
13. Lister JJ, Weaver A, Ellis JD, Himle JA, Ledgerwood DM. A systematic review of rural-specific barriers to medication treatment for opioid use disorder in the United States. Am J Drug Alcohol Abuse 2020 May 03;46(3):273-288. [doi: 10.1080/00952990.2019.1694536] [Medline: 31809217]
14. Stuart A, Stevenson C, Koschate M, Cohen J, Levine M. 'Oh no, not a group!' The factors that lonely or isolated people report as barriers to joining groups for health and well-being. Br J Health Psychol 2022 Feb 24;27(1):179-193. [doi: 10.1111/bjhp.12536] [Medline: 34028949]
15. Hammarlund RA, Crapanzano KA, Luce L, Mulligan LA, Ward KM. Review of the effects of self-stigma and perceived social stigma on the treatment-seeking decisions of individuals with drug- and alcohol-use disorders. Substance Abuse Rehab 2018 Nov;9:115-136. [doi: 10.2147/sar.s183256]
16. Volkow ND. Collision of the COVID-19 and addiction epidemics. Annals Internal Med 2020 Jul 07;173(1):61-62. [doi: 10.7326/m20-1212]
17. Rodham K, McCabe C, Blake D. Seeking support: an interpretative phenomenological analysis of an internet message board for people with Complex Regional Pain Syndrome. Psychol Health 2009 Jul;24(6):619-634. [doi: 10.1080/08870440802563245] [Medline: 20205016]
18. Bunting AM, Frank D, Arshonsky J, Bragg MA, Friedman SR, Krawczyk N. Socially-supportive norms and mutual aid of people who use opioids: an analysis of Reddit during the initial COVID-19 pandemic. Drug Alcohol Depend 2021 May 01;222:108672 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108672] [Medline: 33757708]
19. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010 Jul 27;7(7):e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
20. Sani F. Group identification, social relationships, and health. In: The Social Cure: Identity, Health and Well-Being. Hove, UK: Psychology Press; 2012.
21. Naserianhanzaei E, Koschate-Reis M. Do group memberships online protect addicts in recovery against relapse? Proc ACM Human Comput Interact 2021 Apr 13;5(CSCW1):1-18. [doi: 10.1145/3449142]
22. Greenaway K, Haslam S, Cruwys T, Branscombe N, Ysseldyk R, Heldreth C. From "we" to "me": group identification enhances perceived personal control with consequences for health and well-being. J Pers Soc Psychol 2015 Jul;109(1):53-74 [FREE Full text] [doi: 10.1037/pspi0000019] [Medline: 25938701]
23. Granfield R, Cloud W. Coming Clean, Overcoming Addiction Without Treatment. New York, USA: NYU Press; 1999.
24. Helliwell JF, Putnam RD. The social context of well-being. Philos Trans R Soc Lond B Biol Sci 2004 Sep 29;359(1449):1435-1446 [FREE Full text] [doi: 10.1098/rstb.2004.1522] [Medline: 15347534]
25. Klein O, Spears R, Reicher S. Social identity performance: extending the strategic side of SIDE. Pers Soc Psychol Rev 2007 Feb 29;11(1):28-45. [doi: 10.1177/1088868306294588] [Medline: 18453454]
26. Tajfel H. Human Groups and Social Categories Studies in Social Psychology. Cambridge, UK: Cambridge University Press; 1981.
27. Turner JC, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS. Rediscovering the Social Group: A Self-Categorization Theory. Hoboken, New Jersey: Basil Blackwell; 1987.
28. Steffens NK, Cruwys T, Haslam C, Jetten J, Haslam SA. Social group memberships in retirement are associated with reduced risk of premature death: evidence from a longitudinal cohort study. BMJ Open 2016 Feb 16;6(2):e010164 [FREE Full text] [doi: 10.1136/bmjopen-2015-010164] [Medline: 26883239]
29. Iyer A, Jetten J, Tsivrikos D, Postmes T, Haslam S. The more (and the more compatible) the merrier: multiple group memberships and identity compatibility as predictors of adjustment after life transitions. Br J Soc Psychol 2009 Dec;48(Pt 4):707-733 [FREE Full text] [doi: 10.1348/014466608X397628] [Medline: 19200408]
30. Moos RH, Moos BS. Participation in treatment and Alcoholics Anonymous: a 16-year follow-up of initially untreated individuals. J Clin Psychol 2006 Jun;62(6):735-750 [FREE Full text] [doi: 10.1002/jclp.20259] [Medline: 16538654]
31. Litt MD, Kadden RM, Kabela-Cormier E, Petry N. Changing network support for drinking: initial findings from the Network Support Project. J Consult Clin Psychol 2007;75(4):542-555. [doi: 10.1037/0022-006x.75.4.542]
32. Chambers SE, Canvin K, Baldwin DS, Sinclair JM. Identity in recovery from problematic alcohol use: a qualitative study of online mutual aid. Drug Alcohol Depend 2017 May 01;174:17-22 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.01.009] [Medline: 28282522]

33. D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Escobar Lee K, Cavazos-Rehg PA. Social networking online to recover from opioid use disorder: a study of community interactions. Drug Alcohol Depend 2017 Dec 01;181:5-10 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.09.010] [Medline: 29024875]
34. MacLean D, Gupta S, Lembke A, Manning C, Heer J. Forum77: an analysis of an online health forum dedicated to addiction recovery. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2015 Presented at: CSCW '15: Computer Supported Cooperative Work and Social Computing; Mar 14 - 18, 2015; Vancouver BC Canada. [doi: 10.1145/2675133.2675146]
35. Best D, Groshkova T, Loaring J, Ghufran S, Day E, Taylor A. Comparing the addiction careers of heroin and alcohol users and their self-reported reasons for achieving abstinence. J Groups Addict Recovery 2010 Nov 09;5(3-4):289-305. [doi: 10.1080/1556035x.2010.523364]
36. Haslam C, Best D, A. Dingle G, Staiger PK, Savic M, Bathish R, et al. Social group membership before treatment for substance dependence predicts early identification and engagement with treatment communities. Addict Res Theory 2019 Jan 09;27(5):363-372. [doi: 10.1080/16066359.2018.1537393]
37. Buchanan AS, Latkin CA. Drug use in the social networks of heroin and cocaine users before and after drug cessation. Drug Alcohol Depend 2008 Aug 01;96(3):286-289 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.03.008] [Medline: 18455887]
38. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. Ann Intern Med 2010 Apr 06;152(7):426-33, W141 [FREE Full text] [doi: 10.7326/0003-4819-152-7-201004060-00007] [Medline: 20368648]
39. Havassy BE, Hall SM, Wasserman DA. Social support and relapse: commonalities among alcoholics, opiate users, and cigarette smokers. Addict Behav 1991 Jan;16(5):235-246. [doi: 10.1016/0306-4603(91)90016-b]
40. Herold MD, Søgaard TF. Disturbing the ‘spoiled-unspoiled’ binary: performances of recovering identities in drug-experienced youths’ friendship narratives. Addict Res Theory 2018 Aug 16;27(3):226-234. [doi: 10.1080/16066359.2018.1493462]
41. Fomiatti R, Moore D, Fraser S. Interpellating recovery: the politics of 'identity' in recovery-focused treatment. Int J Drug Policy 2017 Jun;44:174-182. [doi: 10.1016/j.drugpo.2017.04.001] [Medline: 28511843]
42. Keil TF, Koschate M. Variations in subjective definitions of everyday situations as intergroup contact. Br J Soc Psychol 2020 Oct 24;59(4):965-991 [FREE Full text] [doi: 10.1111/bjso.12372] [Medline: 32096272]
43. Visualization of related subreddits. GitHub. URL: https://github.com/anvaka/sayit [accessed 2022-01-04]
44. Kleinbaum D, Klein M. Extension of the Cox proportional hazards model for time-dependent variables. In: Survival Analysis. New York: Springer; 2012.
45. Connor J, Gullo M, White A, Kelly A. Polysubstance use: diagnostic challenges, patterns of use and health. Curr Opin Psychiatry 2014 Jul;27(4):269-275. [doi: 10.1097/YCO.0000000000000069] [Medline: 24852056]
46. Carter BL, Tiffany ST. Meta-analysis of cue-reactivity in addiction research. Addiction 1999 Mar;94(3):327-340 [FREE Full text] [doi: 10.1046/j.1360-0443.1999.9433273.x]
47. Chancellor S, Nitzburg G, Hu A, Zampieri F, De Choudhury M. Discovering alternative treatments for opioid use recovery using social media. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300354]
48. Faulkner-Gurstein R. The social logic of naloxone: peer administration, harm reduction, and the transformation of social policy. Soc Sci Med 2017 May;180:20-27. [doi: 10.1016/j.socscimed.2017.03.013] [Medline: 28315595]
49. Fomiatti R, Moore D, Fraser S. The improvable self: enacting model citizenship and sociality in research on ‘new recovery’. Addict Res Theory 2018 Dec 11;27(6):527-538. [doi: 10.1080/16066359.2018.1544624]
50. Reddit CEO Steve Huffman on the site’s redesign, coming in Q1 2018. Venture Beat. URL: https://venturebeat.com/2017/11/09/reddit-ceo-huffman-on-the-sites-redesign-coming-in-q1-2018/ [accessed 2022-01-04]
51. HHS acting secretary declares public health emergency to address national opioid crisis. Department of Health and Human Services. URL: https://www.hhs.gov/about/news/2017/10/26/hhs-acting-secretary-declares-public-health-emergency-address-national-opioid-crisis.html [accessed 2022-01-04]
52. About underlying cause of death 1999-2018. Centre for Disease Control and Prevention. URL: https://wonder.cdc.gov/ucd-icd10.html [accessed 2022-01-04]
53. Information sheet on opioid overdose. World Health Organisation. URL: https://www.who.int/substance_abuse/information-sheet/en/ [accessed 2022-01-04]

Abbreviations

AA: Alcoholics Anonymous
OUD: opioid use disorder
SIMOR: Social Identity Model of Recovery

Edited by A Mavragani; submitted 19.01.22; peer-reviewed by A Bunting, A Abrantes, M Sillice; comments to author 17.03.22; revised version received 21.04.22; accepted 31.05.22; published 22.08.22.

Please cite as:
Naserianhanzaei E, Koschate-Reis M
Effects of Substance Use, Recovery, and Non–Drug-Related Online Community Participation on the Risk of a Use Episode During Remission From Opioid Use Disorder: Longitudinal Observational Study
J Med Internet Res 2022;24(8):e36555
URL: https://www.jmir.org/2022/8/e36555
doi:10.2196/36555
PMID:35994333

Original Paper

The Internet Knows More Than My Physician: Qualitative Interview Study of People With Rare Diseases and How They Use Online Support Groups

Sadaf Ashtari[1], PhD; Adam Daniel Taylor[1], MSc

California State University, Sacramento, Sacramento, CA, United States

Corresponding Author:
Sadaf Ashtari, PhD
California State University, Sacramento
Tahoe Hall 2086, 6000 J Street
Sacramento, CA, 95819
United States
Phone: 1 9162787102
Email: sadaf.ashtari@csus.edu

Abstract

Background: Patients struggling with rare diseases may face challenges caused by care providers being unfamiliar with their condition. The life span of people with rare diseases may be the same as that of healthy people, but their quality of life is different. Patients with chronic pain are constantly looking for ways to mitigate their pain. Pain killers are not a permanent solution. In addition to the medical and nonmedical costs of rare diseases for both patients and health care providers, there is a need for sustainable sources of information that are available to help with pain and improve their quality of life, with the goal of reducing physician visits and hospital admissions.

Objective: This study investigated the challenges that patients with genetic disorders face in managing their health conditions and finding disease-related information as well as the effect of online peer support groups on pain mitigation and care management.

Methods: Interviews were conducted via Zoom between July 2021 and December 2021. Eligible participants were those who were aged >18 years, had a medical diagnosis of any type of Ehlers-Danlos syndrome (EDS) with chronic pain, and were members of any support group. Participants were recruited through an announcement in the research and survey section of The Ehlers-Danlos Syndrome Society web page. Interviews were analyzed using the framework approach. Data were systematically searched to identify patterns, analyze them, and identify themes. Interview audio files were transcribed and independently coded by two researchers (SA and AT). Through an iterative process, a final coding table was agreed upon by the researchers and used to thematically analyze the data.

Results: We interviewed 30 participants (mean age 37.7, SD 15 years; n=28, 93% were women; n=23, 77% were residing in the United States). Thematic analysis revealed that participants (patients with EDS) were constantly in pain and most of them have not received accurate and timely diagnoses for many years. They expressed their challenges with health care providers regarding diagnosis and treatment, and complained about their providers' lack of support and knowledge. Participants' main sources of information were web-based searches, academic journals, The Ehlers-Danlos Syndrome Society web page, and online peer support groups on Facebook, Reddit, Twitter, and Instagram. Although pain killers, cannabis, and opioids are providing some pain relief, most patients (28/30, 93%) focused on nonmedical approaches, such as hot or ice packs, physical therapy, exercises, massage, mindfulness, and meditation.

Conclusions: This study highlights the information gap between health care providers and patients with genetic disorders. Patients with EDS seek access to information from different web-based sources. To meet the needs of patients with genetic disorders, future interventions via web-based resources for improving the quality of care must be considered by health care professionals and government agencies.

(J Med Internet Res 2022;24(8):e39172) doi:10.2196/39172

KEYWORDS

online peer support group; genetic disorders; pain management; Ehlers-Danlos syndrome; EDS; chronic pain; health care provider; pain mitigation techniques

Introduction

Background

More than 400 million people (30 million Americans) are living with one or more of the 7000 identified rare diseases. Rare diseases encompass many different disorders and symptoms, but approximately 72% of rare diseases are genetic [1]. In this study, our focus was on Ehlers-Danlos syndrome (EDS), a genetic disorder. It is an inherited connective tissue disorder that causes problems in the skin, blood vessels, bones, and other organs. Patients with EDS have to cope with complications such as joint hypermobility or dislocations and damaged skin for their entire life [2]. Similar to other rare diseases, a lack of scientific knowledge and information regarding EDS causes delays in diagnosis. On average, it takes 6 to 8 years for patients to obtain an accurate diagnosis of their disease. There are no Food and Drug Administration–approved drug treatments for >95% of all rare diseases [3]. Therefore, patients with genetic disorders are under considerable pressure to learn about their disease and how they can manage their pain [4]. On the basis of the recent survey results, compared with other diseases such as cancer, patients with genetic disorders did not feel sufficiently supported with issues related to mental health, navigating the health system, physical and daily living, patient care, and sexual needs [5]. The life span of people who are diagnosed with genetic disorders is often the same as that of healthy people, but the quality of their life is usually very different. There are no cures for genetic disorders; therefore, patients cope by mitigating pain and maximizing their quality of life [6].

The availability of health information on the web has increased dramatically over the past decade. A study in 2018 showed that adults in the United States looked on the web for health information 59% more than in 2013 [7]. The study also revealed that, currently, 55% of health care information seekers are relying more on the internet and web-based resources for their health-related information than 5 years ago. More than 67% of American health care information seekers mentioned that they look for health information on social media. The importance of online support groups has been studied among different people [8] with different diseases such as Parkinson disease [9], psychiatric disabilities [10], amyotrophic lateral sclerosis [11], breast cancer [12], chronic diseases [13,14], neuromuscular disorders [15], and alcoholism [16]. Several studies [12,13,17] have discussed the positive impact of online support groups and how patients can benefit from the emotional and community support obtained from such groups. Online support groups are also used for information support. People who faced similar problems can share possible solutions, suggest how to cope with symptoms, and provide information about their disease to other members of the group [11,18].

Objective

This study investigated the approaches followed by patients with EDS to gain information about their disease, find solutions and treatments, and discover pain mitigation techniques. In addition, the effects of online peer support groups on pain mitigation and care management were examined. Information about the challenges faced by patients with genetic disorders with their health care providers is crucial for governments, researchers, health care regulations, and policy makers to be able to improve health care and population health management.

Methods

Overview

This study comprised qualitative, individual, semistructured interviews with patients with EDS to gather perspectives on the care management. Interviews were conducted to investigate the effectiveness of online peer support groups in managing the health conditions of patients with genetic disorders. Transcripts were analyzed using thematic analysis approach, which allows for both inductive and deductive themes to be explored using an iterative, constant comparative coding process.

Ethics Approval

All research activities were approved by the institutional review board of the Sacramento State University (protocol number IRB-19-20-173).

Participants

A total of 30 participants were recruited by advertising the study on The Ehlers-Danlos Syndrome Society web site. Eligible participants were those who were aged >18 years, had a medical diagnosis or a suspected case of any type of EDS with chronic pain, and were members of an online support group.

Recruitment

The recruitment flyer and announcement letter were posted on The Ehlers-Danlos Syndrome Society website. Volunteers for the study were instructed to email the primary investigator to express their interest and learn more about the study. In response to those volunteers, the primary investigator (SA) of the study emailed the e-consent form, study details, time window for the appointment (in PST), and a request for their availability for a 30-minute interview. From July 2021 to December 2021, we received 57 emails and eventually scheduled appointments with 30 (53%) of them. The remaining volunteers (27/57, 47%) did not respond regarding their availability. The interviews were conducted via Zoom, with the participants' permission to record the session.

Data Collection

The semi-instructed interviews were conducted by the first author of the study (SA), using the question guide presented in Multimedia Appendix 1. The discussion began by sharing the experiences of the researcher (SA) as a patient with genetic disorder along with chronic pain and further discussions about the origins of this study. Participants could ask questions at any point during the interview. After explaining the background of the study, participants were asked several demographic questions, leading to questions about their experience with pain and EDS. Participants themselves tended to guide the conversation, and they explained in detail about their journey toward learning more about EDS. The interviewer provided prompts to pursue more detailed discussions on selected topics and help keep the discussions on track.

The main focus of the discussions, along with the research questions, was on (1) pain management techniques, (2) different sources of information about patient's health, (3) use of online support groups in managing one's health condition, and (4) health care providers' roles. The average length of each interview was 45 minutes.

Data Analysis

The interviews were audio-recorded and transcribed verbatim. The research team used inductive thematic approach to develop themes from the interview transcripts. In total, two authors (SA and AT) independently read the 30 transcripts several times to familiarize themselves with the data and identify themes. The 2 authors met to discuss their findings to resolve differences. After this analysis, the authors developed a thematic framework. They also coded all the transcripts using this thematic framework independently. The main topics from the interview guide were used to create the initial deductive codes. Deductive and inductive codes were identified and used, respectively, in our analysis. Then, the authors discussed the results to reach consensus and refine the main themes. NVivo (version 12.2.0; QSR International) was used in this study to organize the transcripts and facilitate analysis.

Results

Overview

Following our template analysis, seven main themes were identified based on patients' experiences of living with genetic disorders and participating in online support groups and their efforts to find solutions for their pain and health conditions. Our themes include (1) patients with EDS disorders are constantly in pain; (2) challenges with health care providers regarding diagnosis and treatment; (3) lack of health care provider support; (4) searching for different sources of information; (5) collection of different pain management techniques; (6) finding disease treatments, lifestyle solutions, and shared experiences in online peer support groups; and (7) changes in health management as a result of participating in online support groups.

Participant Characteristics

A total of 30 English-speaking participants were interviewed in 2021. Participant characteristics are reported in Table 1. Participants had a mean age of 37.7 years. Most participants (28/30, 93%) identified as women, with most of them residing in the United States (23/30, 77%). Our demographics reflect that of the general population with EDS, which is 73.9% women and 26.1% men [19,20]. Most participants (28/30, 93%) were White. In total, >65% of the participants had some college and graduate degree; however, <45% of them were currently working. Owing to chronic pain and health conditions, most of them (17/30, 57%) were temporarily disabled or can no longer work.

Table 1. Participants' characteristics (N=30).

Characteristics	Values, n (%)
Age (years)	
10-21	2 (7)
20-29	4 (13)
30-39	13 (43)
40-49	4 (13)
50-59	4 (13)
60-69	3 (10)
Gender	
Women	28 (93)
Nonbinary	2 (7)
Ethnicity and race	
African American or Black	1 (3)
Pacific Islander	1 (3)
White	28 (93)
Highest level of education	
High school	10 (33)
Bachelor degree	7 (23)
Master degree	10 (33)
PhD	3 (10)
Currently employed	
No	17 (57)
Yes	13 (43)
Country of residence	
International	7 (23)
United States	23 (77)

Patients With EDS Disorders Are Constantly in Pain

Patients with EDS are constantly in pain. They expressed their pain frequency as every day and all the time. The level of pain and its location varied among patients. All quotes are presented verbatim, but to improve readability, some filler words such as *like* and *um* were removed. In addition, the information in the bracket at the end of each quote refers to the interviewee's information that we deidentified and coded. A participant described her pain as follows:

> *Every day, I would say that my normal pain level is probably a three to four that's like my normal, but then whenever I'm having a really bad flare it could be up to like seven sometimes eight.* [H-CA]

The participants also showed their frustration of being in constant pain and their desperation to find a cure or treatment:

> *I am in pain everyday all the time, it's pretty ridiculous.* [E-CA]
>
> *I am in pain every minute, every second, every hour, every day, non stop.* [SH-U]
>
> *I probably suffer about 20 to 30 sublocations and dislocations a day, I mean, I've always been in pain.* [El-U]

Challenges With Health Care Providers Regarding Diagnosis and Treatment

Participants shared their difficulties in receiving timely and precise diagnosis from their health care providers. Most of them (25/30, 83%) declared that their disease remained as a mystery for many years before they were diagnosed with EDS:

> *I've always known there's something different about me, but I was not formally diagnosed until February of this year. I've been regressing pretty steadily since March of 2019, but I did not have a formal diagnosis and so. I go to a lot of different doctors to try different things.* [AM-CO]
>
> *I live in a really rural area as well, so our medical care is pretty limited. You know I don't think there's any doctors here that really too familiar.* [SA-TE]
>
> *I'm so used to managing it on my own. Since I was diagnosed over 100 different doctors specialists in*

> *different areas. Doctors are completely useless, most of the time when I start listing my medical conditions I'll get to the third or fourth one and their eyes glaze over and it's like they just assume.* [SA2-CA]

Several participants (8/30, 27%) were misdiagnosed and had to visit different providers to understand their actual problem:

> *I've been hyper mobile since I was a young child and it got misdiagnosed as various things along the way, until we finally landed on what we now know.* [PA-MI]

> *The diagnosis was only recent probably within the last six months, but up until then everyone thought it was fibromyalgia, but I had been tested for Parkinson's and for multiple sclerosis and for arthritis and every other kind of immunological disease that they could think of.* [CH-NA]

> *It's actually more recent diagnosis. I was mistaken as with lupus for a long time.* [EL-U]

Given that participants had to deal with chronic pain for most of their life, they were eager to learn about their disease and treatment solutions through a trusted source such as a health care provider. Lack of knowledge and awareness among physicians and family members has made the life of patients with EDS more difficult:

> *My biggest pain problem was like my legs, and I remember at being two years old crying on the floor of a gas station and my mom's like come on get up. You know because she didn't believe me, and throughout my whole life I've had these leg pains that would come and go, and just nobody knew what it was you know, and I was 46 years old, when I first got diagnosed.* [SA-SC]

> *I have known about it all my conscious life. I have been working on trying to figure out what this hypermobility means and trying to get doctors and medical professionals help me understand what it means, and it's been a long journey. I did the research myself and looked up medical articles. I took that to my doctor, and I said, "I think I have this" and she read it, looked at it and said, "you know, you are probably right." So, I did it myself, that was in probably 2006, 2007.* [KA-CA]

Lack of Health Care Provider Support

One of the main complaints from participants was the lack of support from their health care providers. Many participants (26/30, 87%) mentioned that their medical team was not supportive and did not trust them when they said they were in pain:

> *I've had issues and it's kind of like, you may have this but it doesn't really matter anyway because it's having a lot of other issues that they couldn't put together, so I've known about it for a long time but been diagnosed for about a year. Because that's one of the biggest things right it's like you have doctors that just don't believe you or they don't believe in EDS. And they'll tell you it's all in your head.* [JU-KE]

> *Just that fear of you know, having a doctor be like Oh, this is all in your head you're just looking for attention you're just a girl Blah Blah Blah like everything that I heard from like age seven. My friend said to do stuff to get doctors to take you seriously.* [JE-BO]

Several participants (10/30, 33%) expressed helplessness in their current situation and desire to find a way to manage their health in a less stressful manner:

> *I distinctly remember being kicked out of Tuft's Ehlers-Danlos because I had been fainting left and right and they're like you should try acupuncture or yoga and we think that this is a psychiatric problem.* [JE-BO]

> *I am scared that it is never going to heal right, and I will never have the strength that I used to have.* [ER-U]

Searching for Different Sources of Information

Participants reported searching different sources for disease information and pain mitigation techniques. Approximately all participants (29/30, 97%) were actively searching the internet to learn about their problem and find solutions to manage their health. Among these resources, online peer support groups hosted by Facebook, Instagram, Twitter, Reddit, and Inspire were the most popular. Each of these platforms has its own tools and methods for distributing information:

> *I find the Facebook group very helpful because they are like people chain, for example one kind of antibiotics that we should not use they shared a name, so I know right now. I'm in local EDS group in Italy.* [DA-IT]

> *The nice thing about doing the Instagram awareness thing was a bunch of people were in my group message but I'm already friendly with them and they are like oh hey I have this too, and let's chat and you know so I'll go to them now and be hey have you tried blah blah blah.* [JE-BO]

> *It's been helpful to be able to talk to other people that have it, too, because they go through what you go through, as opposed to like a medical professional that doesn't go through it and they're trying to help you arguing.* [HA-CA]

> *The subreddit for Ehlers-Danlos and be like hey does anybody know of a thing that's useful for like this type of pain Has anybody had any luck with X or Y or Z.* [JE-BO]

Most participants (25/30, 83%) described their approaches to finding information about their disease. As a first step, they searched symptoms and problems on Google. Most of the time, Google directed them to scientific and scholarly pages from medical journals:

> *I go straight to the internet because that's what my doctors are doing, you know even my doctors don't*

> *know what they're up against so there's really no better.* [CH-NA]
>
> *I'll just Google what the issue is and research that and then I usually bring that to my doctor, and I look more than often and look on PubMed for actual research.* [AM-CO]

Another trusted source of information is The Ehlers-Danlos Syndrome Society website. Participants rely heavily on the disease information found in that website:

> *I also usually go to the Ehlers-Danlos society. They have a lot of information and videos and stuff about EDS so I can usually find stuff there. The Ehlers-Danlos society has an LGBT support group and I go to that one exclusively because it's so much better.* [MI-SE]
>
> *The Ehlers-Danlos society that is a really good place to get information.* [KA-U]

Collection of Different Pain Management Techniques

Participants discussed the nonmedical pain management techniques that they learned through online support groups, web searches, and reading other patients' experiences. Many nonmedical pain mitigation techniques that they use include hot and cold packs, mental therapy and meditation, topical creams, diet, braces, distraction, and salt baths:

> *Ice and heat are my best friend.* [PA-MI]
>
> *My greatest tool for dealing with the pain is mindfulness.* [ME-DC]
>
> *I also have some over the counter muscle rub that's a help, I take a muscle relaxer every day because a lot of the pain that I have is from my muscles being so tense because they're doing the job of my joints.* [LY-NA]
>
> *One thing that I am trying to do is take a look at my own nutrition and figure this out. You know, what is helpful, what should I avoid, and it's been a long journey, but I think I am getting closer. Certain foods are irritating, some fruits make me inflamed and retain water and gain weight. The blood type diet is very helpful.* [KA-CA]

Participants shared their experiences with physical therapy, exercises, dry needling, and acupuncture for reducing their pain. They had both positive and negative experiences with physical therapy. They mentioned that the knowledge of the physical therapist and their familiarity with EDS are necessary for achieving a good result in pain reduction:

> *I'm going to physical therapy for the EDS and that's been helping me a ton in terms of the pain. She started me in the pool which was super helpful.* [JE-BO]
>
> *When I first got diagnosed my doctor sent me for physical therapy. After about a month they said you're not getting any better, so we can't help you so that ended, because at the time they did not know how they can help me.* [SA-SO]
>
> *I get dry needling done. It makes such a difference to my functionality and my pain. It's definitely worth finding the right person to do it.* [SA-AU]

Several participants (4/30, 13%) talked about pain management classes that they attended through their health care provider, but the result was not very satisfactory:

> *I was going to the pain management program. They do a lot of stuff like learning how to meditate and calm your system down. They don't use any kind of painkillers. It's all like the mind and body. The only issue with the pain management program is they're not very educated on people with EDS, and so they were having me do everything that everybody else was doing and that could be very harmful to those of us with EDS.* [HE-CA]

Many participants (28/30, 93%), through trial and error and web-based research, found body positions that gave them less pain and great relief. They tried taping, stretching, moving, sitting, and resting in different positions:

> *I move myself every day. To start the circulation in the blood system, because then I don't get stuck. Sleeping well it's good.* [LO-SW]
>
> *It's very important that I listened to my body and lay down when it says lay down.* [LY-NA]
>
> *What position can I put my body in to quiet it and so slowly but surely, I found things, to do the trick. When I got into where I had to keep moving around...* [SU-NY]

Approximately half of the participants (16/30, 53%) mentioned that they had used different drugs such as opioids and cannabis. However, many of them (10/30, 33%) expressed their concern about addiction or developed drug tolerance and could not receive the same relief in the long term:

> *The topical medical marijuana stuff with cannabidiol in it and essential oils, some of those are amazing.* [SA-CA2]
>
> *I don't take any pain medication I had a very, very bad experience over multiple years of being on progressive levels of opioids.* [SA-CA]
>
> *I was taking opioids when I first got diagnosed. I took those for about three years, and I got off of them two reasons, first, I could tell my body was becoming too used to them, and you know I had to take more and more, but also it actually made my pain worse because you know my body got used to it.* [SA-SO]

Although participants named 23 different nonmedical pain mitigation techniques, pain killers were still on everyone's list. For most of them (20/30, 67%), pain killers and anti-inflammatory medication were not sufficient. They have to use multiple techniques to reduce their pain and frustration. Several pain killers were being used: naltrexone, celecoxib, ibuprofen, meloxicam, carprofen, tylenol, and amitriptyline. They also mentioned using antidepressant medicines such as fluoxetine and diazepam or similar products for sleeping problems caused by their persistent pain.

Finding Pain Solutions, Lifestyle Tips and Tricks, and Community Support Through Online Peer Support Groups

Most interviewees (24/30, 80%) were participating in online peer support groups to find pain management solutions, physician and product referrals, natural remedies, or tips and tricks that can make their life easier. They also benefited from sharing their experiences and hearing other patients' stories:

> *I like when people have recommendations or like what works for them. Different kind of tips and tricks, I guess.* [JE-AL]
>
> *Information about my disorder like you dislocate your shoulder What do you do when your shoulder dislocates do you wear a brace like this do you tape it, how did you tape it yourself, do you use YouTube videos learning it like What do you do you know I mean.* [MI-SE]
>
> *I usually go on them like the Amazon recommendations I find those really helpful, especially when I'm trying to buy furniture. I have asked in the past about assistive devices, so you know it's getting hard for me to turn door handles, anybody got a suggestion.* [SA-AU]

Several participants (7/30, 23%) mentioned their gratitude for being able to help and support other people:

> *I am able to help people because I see these people that are just starting on their journey or they are much younger than me and they've been diagnosed, so I can give them these tips that are all my best practices that you know.* [SA-CA]

Another reason why patients are participating in online peer support groups is the discovery of an empathetic community of people with the same health issues. They stated that being part of the EDS community improves their mental health:

> *It's been helpful to be able to talk to other people that have it, too, because they go through what you go through, as opposed to like a medical professional that doesn't go through it and they're trying to help you.* [HE-CA]
>
> *I was like I'm normal within my subset of people, I'm the majority here, so that was kind of cracking me up because you've been the outcast or that you were the outlier, I guess your whole life.* [SA-CA]

Changes in Health Management as a Result of Participating in Online Support Groups

Approximately all participants (29/30, 97%) reported that their health care management changed as a result of participating in online peer support groups. They agreed that they learned more about their condition and how to handle it after using online support groups. Being part of a group made them feel less alone, reduced their stress, and eventually helped with their pain reduction:

> *You know they're a blessing and a curse I think they're invasive at times and they're a little demanding but they're also incredibly helpful I don't think I would have gotten as far along on my own journey with figuring out what was going on with me, it's definitely helped me to take better care of myself really.* [CH-NA]
>
> *It makes me feel less alone, especially you know, being in you understand, you're in your 30s and you're dealing with all this chronic pain. These groups have also made me feel valid.* [HE-CA]
>
> *They help me remember that I am not a normal person. I cannot do some of the things that other people do, and I do have limits and those are good reminders of that. That oh yeah, I do have limits and oh yeah I shouldn't feel bad that I could not eat pizza with all that cheese and gluten and you know, there is a reason why I am choosing to eat the way I eat, and exercise the way I exercise.* [KA-CA]

They described that their participation in online peer support groups helped them to be more vocal about their needs and start taking better care of themselves:

> *If I didn't have the support group I wouldn't be managing my health, the way that I am.* [JU-KE]
>
> *I'm a lot more proactive I know a lot more. I learned so much.* [LO-UK]
>
> *I've learned to be much more advocate for myself because before you know I found out about it, that I had a rare disorder, it was just you have pain, you know you have to kind of deal with it.* [PA-MI]

Discussion

Principal Findings

Overall, this qualitative study provides strong insights from patients with genetic disorders into the challenges of receiving quality care from health care providers, managing their health condition, and finding pain mitigation techniques.

To the best of our knowledge, this study is among the first to explore the challenges faced by patients with genetic disorders with health care systems and their efforts to find solutions and treatments from the internet via online support groups. The results demonstrate that most patients with genetic disorders are disappointed with the help they receive from their health care providers and actively looking for solutions through different academic and web-based resources. Among all resources, most participants agree that online peer support groups have changed the way they were managing their health previously and helped with care management improvement. The findings from this study can be used to inform future studies on the development of new or improved online support groups. In addition, the findings can be used to reduce the gap of information between patients and health care providers.

Comparison With Previous Studies

This study distinguished the utility of traditional physician-patient relationships versus online support groups in obtaining helpful information for the management of EDS symptoms. The focus of this study was to explore the relationship between people and the medical community when the medical community has limited knowledge of a person's disease. We focused on EDS as a representative of the large community of people with genetic disorders and other rare diseases.

All participants (30/30, 100%) reported experiencing consistent pain, which had major impact on their daily life. Interacting with members of the medical community—general physicians, specialists, chiropractors, and physical therapists—left the study participants dissatisfied with the information they were receiving. From inaccurate diagnoses to lack of knowledge about EDS to recommending treatments that would only worsen the symptoms, trust in health care providers was generally lacking. Owing to frustration, these participants turned to alternative sources of information that they were able to find in online support groups. They found these groups to be helpful places for finding information to manage their conditions. Our findings from this study appear to be consistent with those of previous studies regarding the positive effects of support groups in pain education and the increase in the patient's knowledge about their disease [21-24].

Primarily, study participants were looking for information in 4 categories. First, they wanted to know more about their disease and were interested in any new study that was available in medical journals and publications. Second, they were interested in referrals for effective care providers and products such as braces or topical creams. Third, participants were seeking advice for dealing with everyday challenges such as driving or sleeping comfortably. Finally, participants were looking for emotional support from other people with EDS. Previous studies have demonstrated that patients use support groups to write about their pain and express their feelings. Researchers have found that patients were learning about pain management techniques and reducing their dependence on pharmaceutical pain management tools such as opioids [25-27].

All study participants (30/30, 100%) were very actively engaged in managing their health, and most of them (28/30, 93%) had found that the way they managed their health had been positively affected by participating in online support groups. Participants cited emotional support in using a mobility device, discovery of new bracing techniques, learning self-advocacy in relation to care providers, and generally improved self-reliance as some of the ways online support groups helped them to manage their health.

Study participants mainly resided in the United States, but were also residents of Italy, Canada, Sweden, the United Kingdom, Australia, and the Netherlands. Participants in the United States were slightly more likely to have a positive experience within the health care system, but attitudes toward caregivers and online support groups were similar and independent of country.

Although overall opinions of online support groups were positive, most study participants (24/30, 80%) had significant reservations about them. A common experience was the ineffective ways in which information was presented. Many participants (23/30, 77%) expressed dismay at the amount of information they would have to screen to find the specific information they were looking for. Most participants (25/30, 83%) expressed concern about the veracity of the information they were finding and the lack of citation of a trustworthy source. Many participants (20/30, 67%) found the tone of the support groups to be negative, either owing to self-pity or unwarranted criticism of other group members. Most participants (15/30, 50%) also expressed concern about security and privacy and often preferred to be anonymous. Our findings from this study are compatible with those of previous studies regarding patients' lack of privacy and the need for better information dissemination formats [28,29]. Several other studies have focused on the design and technical aspects of online support groups. They stated that web-based platforms should be more user-friendly and help patients to navigate and access information easily [30,31].

Most study participants indicated that health care professionals did not assist them with or direct them toward the online support groups they ended up using. Health care providers were not aware of the benefits of the information available on these sites. Providers were also not aware of the vast amounts of clinical information that users were compiling on online support groups that would be useful in studying EDS. Given that the economic burden of 379 rare diseases in the United States in 2019 was estimated to be approximately US $1 trillion, health care providers can help to reduce these high medical and nonmedical expenditures by referring patients to online support groups [32]. On the basis of the Centers for Disease Prevention and Control, the total costs of top three chronic diseases such as diabetes [33] (US $327 billion in 2017), cancer [34] (US $157 billion in 2020), and heart disease and stroke [35] (US $214 billion in 2018) are far less than the costs of rare diseases.

Future investigations into online support groups will focus on the ways these sites can be structured to directly meet the needs of patients who are chronically ill.

Limitations

A key limitation of the study was the small sample size and self-selection of participants through one website (The Ehlers-Danlos Syndrome Society). Most participants were women (28/30, 93%), and most were White (28/30, 93%). All the participants (30/30, 100%) were English speakers. In addition, we focused only on one rare disease. A study that includes a broad sample of rare diseases using multiple survey instruments across time can provide valuable insights regarding the role of online support groups in the management of rare diseases. Our findings demonstrated the potential value of online support groups in helping patients in managing their health conditions. Further studies may focus on the existing features of current support groups' platforms, *missing* features, and how developers, with the help of health care providers, can improve their platforms to reduce the spread of misinformation and increase the credibility of contents.

Conclusions

Professionals providing care to patients with EDS and other patients with long-term debilitating conditions should consider online support groups as a source of information for assisting their patients with their conditions. By participating in the same sites as their patients, professionals will be able to provide context to the information being posted, and they will add some trustworthiness to the content. These professionals will also gain insight into the needs of their patients that are not being met in the clinical environment.

Acknowledgments

This study was supported by a grant from Genentech.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview question guide.
[DOCX File , 20 KB - jmir_v24i8e39172_app1.docx]

References

1. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet 2020 Feb 16;28(2):165-173 [FREE Full text] [doi: 10.1038/s41431-019-0508-0] [Medline: 31527858]
2. Ehlers-Danlos syndrome. MedlinePlus. URL: https://medlineplus.gov/genetics/condition/ehlers-danlos-syndrome/ [accessed 2021-07-07]
3. Global Genes homepage. Global Genes. URL: https://globalgenes.org/rare-facts/ [accessed 2021-07-07]
4. Lasker JN, Sogolow ED, Sharim RR. The role of an online community for people with a rare disease: content analysis of messages posted on a primary biliary cirrhosis mailinglist. J Med Internet Res 2005 Mar 31;7(1):e10 [FREE Full text] [doi: 10.2196/jmir.7.1.e10] [Medline: 15829472]
5. Depping MK, Uhlenbusch N, von Kodolitsch Y, Klose HF, Mautner V, Löwe B. Supportive care needs of patients with rare chronic diseases: multi-method, cross-sectional study. Orphanet J Rare Dis 2021 Jan 22;16(1):44 [FREE Full text] [doi: 10.1186/s13023-020-01660-w] [Medline: 33482869]
6. National Center for Advancing Translational Sciences. NIH National Center for Advancing Translational Sciences. URL: https://rarediseases.info.nih.gov/guides/pages/25/how-to-find-a-disease-specialist [accessed 2021-07-14]
7. The great American search for healthcare information. Weber Shandwick AND KRC Research. URL: https://www.webershandwick.com/wp-content/uploads/2018/11/Healthcare-Info-Search-Report.pdf [accessed 2021-06-15]
8. Finding healthcare support in online communities: an exploration of the evolutionefficacy of virtual support groups. Inivan y, ed. In: Handbook on 3D3C Platforms. Cham: Springer; 2016.
9. Attard A, Coulson NS. A thematic analysis of patient communication in Parkinson's disease online support group discussion forums. Comput Human Behav 2012 Mar;28(2):500-506. [doi: 10.1016/j.chb.2011.10.022]
10. Kaplan K, Salzer MS, Solomon P, Brusilovskiy E, Cousounis P. Internet peer support for individuals with psychiatric disabilities: a randomized controlled trial. Soc Sci Med 2011 Jan;72(1):54-62. [doi: 10.1016/j.socscimed.2010.09.037] [Medline: 21112682]
11. Stewart Loane S, Webster CM, D'Alessandro S. Identifying consumer value co-created through social support within online health communities. J Macromarket 2014 May 29;35(3):353-367. [doi: 10.1177/0276146714538055]
12. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group for women with primary breast cancer. Cancer 2003 Mar 01;97(5):1164-1173 [FREE Full text] [doi: 10.1002/cncr.11174] [Medline: 12599221]
13. Yao T, Zheng Q, Fan X. The impact of online social support on patients' quality of life and the moderating role of social exclusion. J Serv Res 2015 May 08;18(3):369-383. [doi: 10.1177/1094670515583271]
14. Kanter E, Bevan JL, Dorros SM. The use of online support groups to seek information about chronic illness: applying the theory of motivated information management. Commun Q 2018 Nov 07;67(1):100-121. [doi: 10.1080/01463373.2018.1539406]
15. 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]
16. Chuang K, Yang C. Social support in online healthcare social networking. core.ac.uk. URL: https://core.ac.uk/download/pdf/4823833.pdf [accessed 2021-06-11]
17. Winefield HR. Support provision and emotional work in an internet support group for cancer patients. Patient Educ Couns 2006 Aug;62(2):193-197. [doi: 10.1016/j.pec.2005.07.002] [Medline: 16139466]
18. McCormack A. Individuals with eating disorders and the use of online support groups as a form of social support. Comput Inform Nurs 2010;28(1):12-19 [FREE Full text] [doi: 10.1097/NCN.0b013e3181c04b06] [Medline: 19940616]

19. Kulas Søborg ML, Leganger J, Quitzau Mortensen L, Rosenberg J, Burcharth J. Establishment and baseline characteristics of a nationwide Danish cohort of patients with Ehlers-Danlos syndrome. Rheumatology (Oxford) 2017 May 01;56(5):763-767. [doi: 10.1093/rheumatology/kew478] [Medline: 28077691]
20. Demmler JC, Atkinson MD, Reinhold EJ, Choy E, Lyons RA, Brophy ST. Diagnosed prevalence of Ehlers-Danlos syndrome and hypermobility spectrum disorder in Wales, UK: a national electronic cohort study and case-control comparison. BMJ Open 2019 Nov 04;9(11):e031365 [FREE Full text] [doi: 10.1136/bmjopen-2019-031365] [Medline: 31685485]
21. Harrison D, Reszel J, Dagg B, Aubertin C, Bueno M, Dunn S, et al. Pain management during newborn screening: using YouTube to disseminate effective pain management strategies. J Perinat Neonatal Nurs 2017;31(2):172-177. [doi: 10.1097/JPN.0000000000000255] [Medline: 28437309]
22. Harrison D, Sampson M, Reszel J, Abdulla K, Barrowman N, Cumber J, et al. Too many crying babies: a systematic review of pain management practices during immunizations on YouTube. BMC Pediatr 2014 May 29;14:134 [FREE Full text] [doi: 10.1186/1471-2431-14-134] [Medline: 24885559]
23. Meldrum S, Savarimuthu BT, Licorish S, Tahir A, Bosu M, Jayakaran P. Is knee pain information on YouTube videos perceived to be helpful? An analysis of user comments and implications for dissemination on social media. Digit Health 2017;3:2055207617698908 [FREE Full text] [doi: 10.1177/2055207617698908] [Medline: 29942583]
24. Ingadottir B, Blondal K, Thue D, Zoega S, Thylen I, Jaarsma T. Development, usability, and efficacy of a serious game to help patients learn about pain management after surgery: an evaluation study. JMIR Serious Games 2017 May 10;5(2):e10 [FREE Full text] [doi: 10.2196/games.6894] [Medline: 28490419]
25. Willis E. Patients' self-efficacy within online health communities: facilitating chronic disease self-management behaviors through peer education. Health Commun 2016;31(3):299-307. [doi: 10.1080/10410236.2014.950019] [Medline: 26325224]
26. Graves RL, Goldshear J, Perrone J, Ungar L, Klinger E, Meisel ZF, et al. Patient narratives in Yelp reviews offer insight into opioid experiences and the challenges of pain management. Pain Manag 2018 Mar 01;8(2):95-104. [doi: 10.2217/pmt-2017-0050] [Medline: 29451418]
27. Berman RL, Iris MA, Bode R, Drengenberg C. The effectiveness of an online mind-body intervention for older adults with chronic pain. J Pain 2009 Jan;10(1):68-79. [doi: 10.1016/j.jpain.2008.07.006] [Medline: 18774342]
28. Henderson EM, Keogh E, Eccleston C. Why go online when you have pain? A qualitative analysis of teenagers' use of the Internet for pain management advice. Child Care Health Dev 2014 Jul 12;40(4):572-579. [doi: 10.1111/cch.12072] [Medline: 23663232]
29. des Bordes JK, Gonzalez E, Lopez-Olivo MA, Shethia M, Nayak P, Suarez-Almazor ME. Assessing information needs and use of online resources for disease self-management in patients with rheumatoid arthritis: a qualitative study. Clin Rheumatol 2018 Jul 2;37(7):1791-1797. [doi: 10.1007/s10067-018-4047-9] [Medline: 29500557]
30. Nielsen M, Jull G, Hodges PW. Information needs of people with low back pain for an online resource: a qualitative study of consumer views. Disabil Rehabil 2014;36(13):1085-1091. [doi: 10.3109/09638288.2013.829532] [Medline: 24001260]
31. Merolli M, Gray K, Martin-Sanchez F. Therapeutic affordances of social media: emergent themes from a global online survey of people with chronic pain. J Med Internet Res 2014 Dec 22;16(12):e284 [FREE Full text] [doi: 10.2196/jmir.3494] [Medline: 25533453]
32. The National Economic Burden of Rare Disease Study. Everylife Foundation for Rare Disease. URL: https://everylifefoundation.org/burden-study/ [accessed 2022-08-01]
33. American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. Diabetes Care 2018 May;41(5):917-928 [FREE Full text] [doi: 10.2337/dci18-0007] [Medline: 29567642]
34. Cancer prevalence and cost of care. NIH National Cancer Institute. URL: https://surveillance.cancer.gov/research/cost-projections.html [accessed 2021-07-08]
35. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics-2018 update: a report from the American heart association. Circulation 2018 Mar 20;137(12):e67-492 [FREE Full text] [doi: 10.1161/CIR.0000000000000558] [Medline: 29386200]

Abbreviations

EDS: Ehlers-Danlos syndrome

Edited by R Kukafka; submitted 02.05.22; peer-reviewed by B Samhan, X Xiong; comments to author 21.05.22; revised version received 31.05.22; accepted 19.07.22; published 25.08.22.

Please cite as:
Ashtari S, Taylor AD
The Internet Knows More Than My Physician: Qualitative Interview Study of People With Rare Diseases and How They Use Online Support Groups
J Med Internet Res 2022;24(8):e39172
URL: https://www.jmir.org/2022/8/e39172
doi:10.2196/39172
PMID:36006679

Original Paper

Identifying Influences in Patient Decision-making Processes in Online Health Communities: Data Science Approach

Mingda Li[1], PhD; Jinhe Shi[1], PhD; Yi Chen[2], PhD

[1]Ying Wu College of Computing, New Jersey Institute of Technology, Newark, NJ, United States

[2]Martin Tuchman School of Management, New Jersey Institute of Technology, Newark, NJ, United States

Corresponding Author:
Yi Chen, PhD
Martin Tuchman School of Management
New Jersey Institute of Technology
184-198 Central Ave
Newark, NJ, 07102
United States
Phone: 1 973 596 6302
Email: yi.chen@njit.edu

Abstract

Background: In recent years, an increasing number of users have joined online health communities (OHCs) to obtain information and seek support. Patients often look for information and suggestions to support their health care decision-making. It is important to understand patient decision-making processes and identify the influences that patients receive from OHCs.

Objective: We aimed to identify the posts in discussion threads that have influence on users who seek help in their decision-making.

Methods: We proposed a definition of influence relationship of posts in discussion threads. We then developed a framework and a deep learning model for identifying influence relationships. We leveraged the state-of-the-art text relevance measurement methods to generate sparse feature vectors to present text relevance. We modeled the probability of question and action presence in a post as dense features. We then used deep learning techniques to combine the sparse and dense features to learn the influence relationships.

Results: We evaluated the proposed techniques on discussion threads from a popular cancer survivor OHC. The empirical evaluation demonstrated the effectiveness of our approach.

Conclusions: It is feasible to identify influence relationships in OHCs. Using the proposed techniques, a significant number of discussions on an OHC were identified to have had influence. Such discussions are more likely to affect user decision-making processes and engage users' participation in OHCs. Studies on those discussions can help improve information quality, user engagement, and user experience.

(J Med Internet Res 2022;24(8):e30634) doi:10.2196/30634

KEYWORDS

influence relationship; decision-making threads; online health communities; patient engagement; deep learning; text relevance measurement

Introduction

Background

In recent years, online health communities (OHCs) such as the Cancer Survivors Network (CSN), MedHelp, DoctorLounge, WebMD, and Health-boards message boards have become one of the most important resources that patients leverage [1]. An OHC is defined as an asynchronous web-based message board system for patients that contains multiple message boards, each of which typically focuses on 1 disease. OHCs provide a web-based channel that enables information exchange, facilitates communication, and provides support to patients and caregivers [2-4]. They are especially valuable for patients with chronic diseases to learn about their conditions and seek social support [5,6].

Empowering and supporting patients to make informed health care decisions is a key component of patient-centered health care and is a social, economic, and technical necessity [7,8]. A lot of patients seek information and advice on OHCs. Existing work has found that nearly half of the threads in a breast cancer forum [9] are related to patient decision-making [1]. Studies have also shown that patients are often influenced by web-based sources and social media in their health care decision-making [10,11].

Objectives

The goal of this study was to identify the influence relationship of posts in discussion threads related to health care decision-making. Specifically, we defined the influence relationships and identified post replies that influenced the initial author, who had questions posted on OHCs.

The outcomes of this study are important for health care professionals to help patients make informed decisions for several reasons. First, analyzing the writing style and pattern of posts that have influence may help explain why they have influence and provide insights to health care professionals on effective communication with patients. Second, if the information provided by posts that have an influence is not accurate, it will mislead patients. It is important to check the information quality in such posts to improve the quality of influence. Furthermore, a patient who has questions but does not receive any replies that have an influence may need further help.

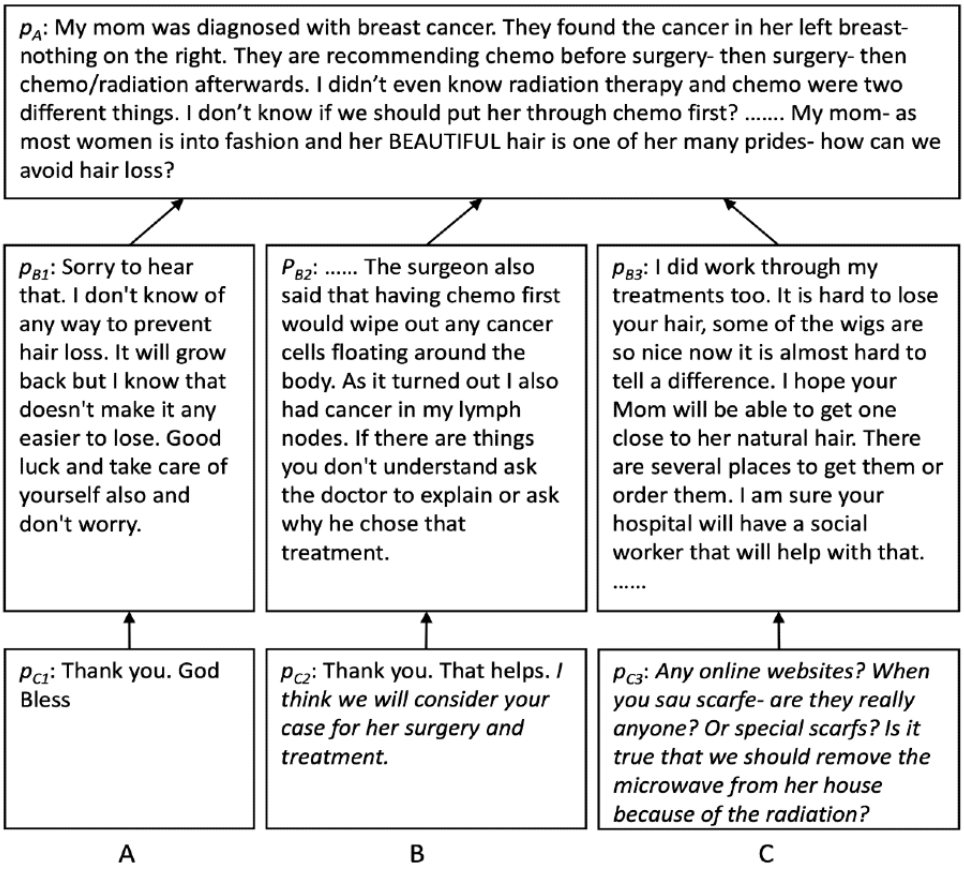
Literature Review

There is a lot of research conducted on OHC analysis, although with limited study on identifying influence relationships of posts. Several studies have been conducted on analyzing the reciprocal patterns between users' replies in discussion forums [12-14]. There is also work on analyzing the patterns between post views and post replies [15]. Many studies have been conducted on identifying influential users in a community [16-20]. In those applications, a post, blog, or tweet typically expresses an opinion of the author, and the replies are considered as an indication of being influenced by the opinion of the original post. That is, the reply relationship is considered as an influence relationship. The focus is on judging the influential power of an author based on activeness of post writing [21] and social network features [17,18] such as PageRank-like algorithms or clustering algorithms.

Finding influence relationships among posts in discussion forums is different from finding influential users and requires different techniques. In an OHC, the initial author of a thread typically expresses a question, not an opinion. The influence happens when a reply to the question affects the initial author. There are only 2 existing studies that consider the influence of the replier on the initial author [21,22]. This influence is identified when the sentiment of the initial author is changed to be similar to that of the replier. However, this definition may not be accurate.

Let us look at an example of a discussion thread related to patient decision-making, shown in Figure 1. An OHC user initialized a thread asking for advice on whether to have chemotherapy before surgery for her mother's treatment plan in post p_A.

Figure 1. Example of a discussion thread.



p_A: My mom was diagnosed with breast cancer. They found the cancer in her left breast- nothing on the right. They are recommending chemo before surgery- then surgery- then chemo/radiation afterwards. I didn't even know radiation therapy and chemo were two different things. I don't know if we should put her through chemo first? .... My mom- as most women is into fashion and her BEAUTIFUL hair is one of her many prides- how can we avoid hair loss?

p_{B1}: Sorry to hear that. I don't know of any way to prevent hair loss. It will grow back but I know that doesn't make it any easier to lose. Good luck and take care of yourself also and don't worry.

P_{B2}: ...... The surgeon also said that having chemo first would wipe out any cancer cells floating around the body. As it turned out I also had cancer in my lymph nodes. If there are things you don't understand ask the doctor to explain or ask why he chose that treatment.

p_{B3}: I did work through my treatments too. It is hard to lose your hair, some of the wigs are so nice now it is almost hard to tell a difference. I hope your Mom will be able to get one close to her natural hair. There are several places to get them or order them. I am sure your hospital will have a social worker that will help with that. ......

p_{C1}: Thank you. God Bless

p_{C2}: *Thank you. That helps. I think we will consider your case for her surgery and treatment.*

p_{C3}: *Any online websites? When you sau scarfe- are they really anyone? Or special scarfs? Is it true that we should remove the microwave from her house because of the radiation?*

In Figure 1A, a user replied by comforting her in post p_{B1}. The reply was not informative. Even though the initial author expressed gratefulness to the author of post p_{B1}, with sentiment changing to be positive in post p_{C1}, she was not influenced by post p_{B1}. Indeed, studies show that 75% to 85% of CSN forum participants change their sentiment in a positive direction through web-based interactions with other community members [23]. A change in sentiment is not necessarily an indicator of being influenced.

In contrast, in Figure 1B, a user shared her experience in a similar situation suggesting to have chemotherapy before a surgery in post p_{B2}. The initial author expressed her gratitude and indicated that she would consider this suggestion in determining her mother's treatment plan (the sentences in italics) in p_{C2}, showing her being influenced.

Contribution

Instead of considering sentiment changes, we propose using questions or future actions on relevant replies as an indicator of being influenced, as illustrated in the aforementioned example. There are 2 major challenges in identifying influence relationships. First, we need to define influence relationships of posts. We examined the semantics of post content to define influence relationships. Unlike influential users, who are defined by network features in the existing work [16-20], text content is the key to determine whether posts have influence. Second, it is hard to identify influence relationships. Unlike typical text classification problems, influence relationships involve multiple posts with reply relationships rather than a single paragraph of text. In addition, influence is an abstract concept. It is challenging to extract relevant features to capture the influence patterns considering both content and the reply relationship.

This study makes novel contributions to identifying influence relationships in discussion threads in OHCs related to patient decision-making. Specifically, (1) we defined the influence relationship between the posts based on the semantics of the post content, (2) an extensible deep learning model that extracts and combines both sparse and dense features was proposed to identify the influence relationships in OHC decision-making threads, and (3) the proposed model achieved good performance in identifying influence relationships in empirical evaluation.

Methods

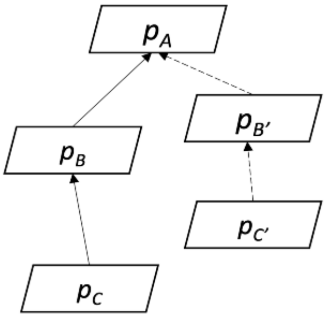
In this section, we first model the OHC data and define the influence relationship in discussion threads. We then propose a deep learning–based model to identify the influence relationships.

Problem Definition

Definition of Discussion Threads

Figure 2 presents an overview of the OHC data structure. We modeled an OHC as a set of discussion threads $T = \{t_1, t_2,..., t_n\}$. Each thread t_i is composed of a set of posts and a function R that represents the reply relationship. For example, Figure 2 illustrates a thread that contains a set of 5 posts $\{p_A, p_B, p_C, p_{B'}, p_{C'}\}$. One of the reply relationships, $R(p_B) = p_A$, represents that post p_B replies to post p_A. Each post p_i consists of a sequence of sentences $p_i = \{s_1, s_2,..., s_l\}$. Each post has an author. We denoted the author relationship using a function U. $U(p_i)$ represents the author of post p_i. Note that a post only has a single author; however, an author may write ≥0 posts in a thread. We used p_A to present the first post of a thread and named it the *initial post*. The author of the *initial post*, $U(p_A)$, is referred to as the *initial author* of the thread.

Figure 2. Data structure of an online health community.



Existing work [1] has studied the thread discussions in OHCs and identified that a subset of threads is related to *patient decision-making*. Such a thread is characterized by questions in the initial post and replies with suggestions of options. Techniques have been developed to identify decision-making threads in OHCs.

In this paper, we study how to identify the cases where the initial author of a decision-making thread is influenced by a reply post. Note that our study is general to any thread discussions related to decision-making. The definition and identification of decision-making threads can be handled using the approach developed in existing work [1] or other approaches. In the rest of this paper, we use *threads* to refer to decision-making threads for simplicity. The defined influence relationship may not be applicable to discussion threads that are not related to decision-making, such as discussion threads for casual communication or experience-sharing threads providing social support.

Definition of Relationships

Overview

Before introducing the definition of *influence relationships*, we first introduce relationships. A relationship is defined on a triple of posts in a thread with reply relationships: an initial post, a reply to the initial post, and the initial author's subsequent reply.

Definition 1 (Relationship)

We define the relationship among three posts p_A, p_B, and p_C, in a thread as $r_i = (p_A, p_B,$ and $p_C)$, where post p_A is the initial post of the thread, post p_B replies to p_A, post p_C replies to p_B, and the authors of p_A and p_C are the same person. That is, $R(p_B) = p_A$, $R(p_C) = p_B$, and $U(p_A) = U(p_C)$.

We used $r_i = (p_A, p_B, p_C)$ to denote the relationship among p_A, p_B, and p_C. Note that there are many such relationships in a thread, and we considered all such triples. For instance, Figure 2 shows a thread with 2 relationships, $r_1 = (p_A, p_B, p_C)$ and $r_2 = (p_A, p_{B'}, p_{C'})$.

Also, note that existing work on identifying influential users [16-20] does not consider the relationships among post triples but only considers the reply relationship between 2 posts.

Definition of Influence Relationships

Intuition

Now, we discuss how to define *influence relationships* on relationship (p_A, p_B, p_C), where post p_B has an influence on the initial author $U(p_A)$.

First, intuitively, if post p_B influences the initial author $U(p_A)$, then the content of these 3 posts must be relevant.

Second, we referred to the definition of *influence* in Merriam-Webster [24]—"to affect or alter by indirect or intangible means"—and the reaction of *being influenced* is to *sway* rather than being convinced. If the initial author considers the suggestion given in post p_B, even if she eventually does not take the suggestion, she is considered to have been influenced by post p_B. On the basis of this definition, we observed 2 indications that the initial author, $U(p_A)$, was influenced by p_B.

An observation of being influenced is that the initial author may ask questions in p_C based on the suggestions in p_B. Curiosity is a motivator for learning and influential in decision-making [25]. An existing study [26] used a statistically large sample of learning forum posts to investigate whether student participation in the forum could be influenced. They observed that students who were influenced by others' interesting answers were more likely to ask follow-up questions. This indicates that asking further questions is a sign of being influenced. The same pattern also exists in OHCs. Let us look at the example in Figure 1C. The initial author expressed concerns about hair loss in p_A. Another user replied in post p_{B3} suggesting the use of wigs. The initial author then replied in post p_{C3} with questions (the sentences in italics) for more details about the suggestion given in post p_{B3}. These questions indicate that the initial author was thinking about the suggestion given in post p_B; that is, being influenced.

The second indication that the initial author was influenced by a post p_B is that she expressed her intention to take action in post p_C. Adjei et al [27] found that member-to-member communication in web-based brand communities greatly influenced the members' future purchase behavior. Similarly, the communication through discussion threads in OHCs may also affect the initial author's future actions. Let us look at the example in Figure 1B again. For the treatment question asked in p_A, a forum user shared her experience and discussed the treatment in post p_{B2}. The initial author then replied with a planned action (the sentence in italics) in p_{C2}. The intention of future action based on the communications in the thread is an indicator of the influence relationship.

On the basis of these observations, we define influence relationships in decision-making threads in the following section.

Definition 2 (Influence Relationship)

A relationship $r_i = (p_A, p_B, p_C)$ is considered as an influence relationship—that is, $U(p_A)$ is influenced by p_B—if and only if the following conditions are met: (1) the content of p_B is relevant to post p_A, (2) the content of p_C is relevant to post p_B, and (3) p_C contains questions or indicates future actions.

To identify influence relationships, we modeled it as a classification task. Given a set of relationships $R = \{r_1, r_2,..., r_n\}$, for each relationship r_i, we predicted its label to be either 1 or −1, where label 1 indicated that r_i was an influence relationship and label –1 indicated that r_i was not an influence relationship. The goal was to learn a model from the labels of known relationships and predict the labels for unlabeled relationships.

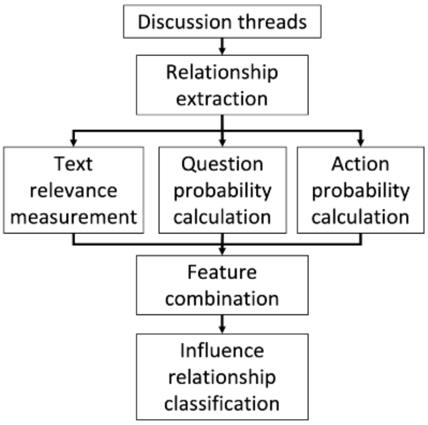
Model Design

Overview

In this section, we present the method to identify the influence relationships in decision-making threads in OHCs. Figure 3 presents the framework of the proposed method.

Given a set of discussion threads as the input, we first extracted the triple relationships using the relationship extraction module. Text relevance features, question probability features, and action probability features were then calculated using the text relevance measurement module, the question probability calculation module, and the action probability calculation module, respectively. Finally, all the features were combined using a deep learning model in the feature combination module to generate the probability of a relationship being an influence relationship.

Figure 3. Workflow of influence relationship identification.



Relationship Extraction Module

In this section, we introduce the relationship extraction module, which extracted all relationships defined in definition 1.

In the first step of relationship extraction, we built the reply tree structure based on the indented format in html files. For each adjacent post pair, the post that was posted earlier was treated as the parent of the latter post. The ancestor-descent distance between a post and the initial post was represented by the number of tab characters. The reply structure of a thread is illustrated in Figure 2. Each post is a node in the thread tree, and each edge represents a reply relationship. The root of the thread tree is the initial post (ie, p_A) in definition 1.

Existing work observes that, in some forums, the reply structure in a discussion thread may not be fully available and proposes techniques to construct full reply structures [28]. The OHCs used in our experiments had a full reply structure. Existing techniques can be leveraged if needed for other forums.

We then navigated the thread tree to extract all relationship triples, as defined in definition 1. Each triple started with the initial post followed by a reply to the initial post written by another author and then a subsequent reply by the initial author, all of which were on the same path in the thread tree. For example, $r_1 = (p_A, p_B, p_C)$ and $r_2 = (p_A, p_{B'}, p_{C'})$ are 2 relationships in the thread tree in Figure 2.

Text Relevance Measurement Module

The text relevance measurement module measures the content relevance, or text semantic similarity, of 2 posts using a relevance score between 0 and 1.

There are mainly 2 types of deep learning–based methods in the literature that measure text relevance. The first type of method extracts content feature vectors of 2 input texts and then combines them to make a prediction, such as the Deep Structured Semantic Models (DSSM) [29], the Convolutional DSSM [30], and Architecture-I (ARC-I) [31]. The intuition of this method is to highlight the important information of the original texts so that irrelevant content can be removed before the feature combination phase. However, the drawback of this type of method is that it runs the risk of losing detail [32].

The second type generates the word-level relevance first and then uses neural networks to learn the hierarchical interaction patterns for content-level relevance, such as DeepMatch [33], Architecture-II (ARC-II) [31], and MatchPyramid [34]. The motivation is that making a good relevance judgment requires considering the interactions in the text relevance measurement process, starting from the interactions between words to patterns in phrases and those in whole sentences [34]. However, the training process for the second type is much more expensive than for the first one.

We evaluated both approaches to measure text relevance in experiments. We chose 2 state-of-the-art representative methods for the text relevance measurement module in the evaluation. For the first type, we chose ARC-I [31], which uses a multilayer perceptron to combine relevance feature vectors. It shows better performance than the DSSM [29] and Convolutional DSSM [30], both of which use cosine similarity [34]. We chose MatchPyramid [34] to represent the second type of method as it exhibits better performance than the other 2 methods (DeepMatch [33] and ARC-II [31]) in experiments on multiple data sets [34].

We further proposed the adaptation of Bidirectional Encoder Representations from Transformers (BERT) [35] as the embedding layer in the ARC-I and MatchPyramid models. BERT is a state-of-the-art embedding method for word representation in many natural language understanding tasks, trained on BookCorpus and English Wikipedia. We considered both BERT (trained on Wikipedia) and word2vec (trained on the training data set) as the embedding methods for both ARC-I and MatchPyramid. Different variations of the text relevance measurement module are evaluated in the *Text Relevance Evaluation* section.

Question Probability Calculation Module

We now discuss how to calculate the probability of a post containing a question using the question probability calculation module.

There are 2 types of methods to identify question sentences in forums: a rule-based approach and a learning-based approach. In a rule-based approach, question marks and 5W1H words (what, who, when, where, why, and how) are used to identify question sentences [36]. A learning-based approach uses sequential question patterns to train a binary classifier on labeled data [37-40]. Liu and Jansen [37] used the question mark to extract question posts from Sina Weibo. In the studies by Ranganath et al [38,39], frameworks were proposed to identify rhetorical questions by modeling the motivation of the user for posting them. In the study by Ojokoh et al [40], questions from ResearchGate were identified based on the maximum probability value of a naïve Bayes classification with part-of-speech tag features.

Both rule-based and learning-based approaches can achieve excellent performances. A study shows that a rule-based approach can outperform complicated learning-based approaches [36]. Thus, we followed a rule-based method [36] to identify question presence in the posts. In total, 2 types of rules were considered: question marks and 5W1H words. We made adaptations of this approach for OHCs. As a question mark is the most significant sign of a question, we gave a higher confidence score to a sentence with a question mark. We also set some constraints on 5W1H words to simulate the pattern of question sentences. First, 5W1H must appear at the beginning of a sentence. Second, auxiliary words were added to the original words for more specific patterns. For example, we considered *what is*, *what are*, *what does*, and *what do* instead of *what*.

After the question probability of each sentence in a post p_i was calculated, the maximum probability was used as the likelihood of post p_i containing at least one question, denoted as $Q(p_i)$.

Action Probability Calculation Module

This section presents the action probability calculation module, which generated the probability of action presence in a post.

The indication of a future action can be captured by the presence of verbs and appropriate sentence tense. The Natural Language Toolkit (NLTK) [41] tagger module defines a standard interface for augmenting each token of a text with supplementary information, such as its part of speech or its WordNet synset tag, and provides several different implementations for this interface. We leveraged the NLTK tagger module to assess the likelihood of a post containing future actions by checking the existence of words with a future tense verb tag (eg, *will consider* in Figure 1B) or a modal auxiliaries tag (eg, *can*, *could*, *may*, and *must*). To count on the cases where future tenses may not be identified because of forum users' typos or informal writing, we set the probability of future action to be 0.5 when the rules failed to identify future actions. Equation 1 shows the calculation formula to generate the action probability of a post p_i.

Note that we did not consider negation in the action probability calculation module. For example, in post p_C, the initial author disagrees with the suggestions proposed in p_B and decides to do something different. For those cases, the overall meaning of p_B and p_C would be the opposite and, therefore, would be captured by the relevance vectors generated in the text relevance measurement module. Thus, we did not consider negations in this phase to avoid double counting.

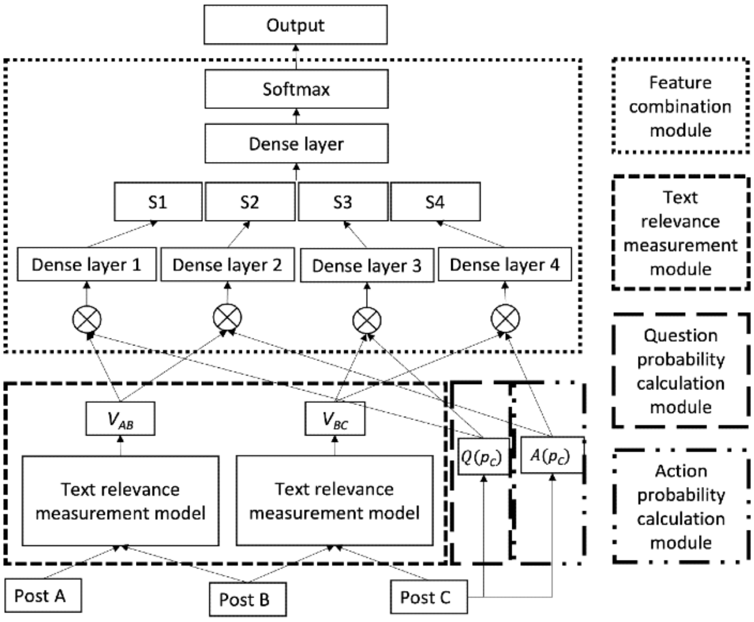
Feature Combination Module

Overview

Referring to Figure 4, the text relevance measurement module calculated P_{AB}—the relevance score between p_A and p_B—and P_{BC}—the relevance score between p_B and p_C. The question probability calculation module and action probability calculation module calculated the question probability $Q(p_C)$—or Q in short—and action probability $A(p_C)$—or A in short—based on the text of p_C.

We now discuss the feature combination module that measures the influence score based on these features. We discuss 2 alternative methods: *a baseline approach* and *a deep learning model*.

Figure 4. Architecture of the feature combination module.



Baseline Approach

Recall that, according to definition 2, the presence of an influence relationship requires the relevance between post p_A and post p_B, the relevance between post p_B and post p_C, and the presence of a question or action in post p_C. We started with an intuitive method to detect influence relationships based on the definition using Equation 2.

$$P_{baseline} = P_{AB} \times P_{BC} \times \max\,[Q(p_C), A(p_C)]\ \mathbf{(2)}$$

We set the thresholds to 0.5, 0.5, and 0.9 for each component.

Deep Learning Approach

We further proposed a deep learning model that combines the text relevance, the likelihood of question presence, and the likelihood of future action presence to identify influence relationships. The architecture of this model is shown in Figure 4.

Compared with the *baseline* approach, there are 3 major benefits of using a deep learning model. First, it is labor-intensive, time-consuming, and difficult to determine appropriate thresholds for cutting off the probabilities using a rule-based approach such as the *baseline* approach. A threshold that works well for one data set may not be optimal for another. Both a rule-based approach and a deep learning model require different thresholds for different data sets. A rule-based approach requires manual parameter tuning for each data set. In contrast, a deep learning approach learns thresholds from the ground truth and, thus, can easily adapt to a new data set with minimal human intervention [42]. Second, the question and action features may have different interactions with the relevance features. We observed that questions are often relevant, but actions are not necessarily. People typically express appreciation in post p_C or sometimes even mention actions totally irrelevant to post p_B, such as the plan to travel or shop. Being relevant is more important to consider in the presence of actions compared with in the presence of questions. However, in the *baseline* approach, the question and action features are merged before being combined with the relevance features, resulting in the loss of important information. Furthermore, we used relevance vectors as inputs to the deep learning model to calculate the influence score. Compared with the *baseline* approach, which uses the relevance scores as input to measure the influence score, relevance vectors provide much richer information. This can be especially helpful when there are several topics involved in the discussion. The relevance information is also leveraged during the phase of combining the relevance features with the question or action features.

Let V_{AB} denote the relevance vector between p_A and p_B and V_{BC} denote the relevance vector between p_B and p_C. We generated V_{AB},V_{BC} from p_A, p_B, and p_C and calculated Q and A from p_C.

These features were then connected. The question or future action in p_C must be related to the content of p_A and p_B. Thus, we combined V_{AB} and V_{BC} with Q and A using one of the following two operators: (1) *cat* (concatenating each relevance vector with question or action probability) and (2) *dot* (multiplying each relevance vector with question or action probability).

There are 2 major differences between these 2 operators for connecting the features: *cat* and *dot*. First, *dot* makes sure that Q and A affect each dimension in the relevance vectors, whereas *cat* cannot guarantee this as some neurons or nodes are dropped out. Some interactions between questions or actions and text relevance may be ignored by the *cat* operator. Second, the training process of the *cat* is more expensive than that of the *dot* because, for each dense layer 1 to 4, there is an additional dimension for the *cat* compared with for the *dot*.

In Figure 4, we use ⊗ to present the combination operator, which can be either *cat* or *dot*. The combination step produces 4 feature vectors: $V_{AB} \otimes Q$, $V_{AB} \otimes A$, $V_{BC} \otimes Q$, and $V_{BC} \otimes A$. To extract the key information from these combined feature vectors, 4 dense (fully connected) layers were used to populate the summarized feature vectors (S_1, S_2, S_3, S_4). The concatenation of these 4 summarized feature vectors was passed through 2 dense layers. The first one was used to further combine the summarized feature vectors. The second one aimed to generate the probability distribution over the labels. To avoid gradient vanishing and exploding [43], we chose the *Relu* function as the activation function for all the dense layers except the output layer, which uses the *softmax* function to populate the probabilities.

We trained the model using the binary cross-entropy loss function defined in Equation 3, which minimizes the distance between the probability distributions of the ground truth and those of the predicted score.

Where y_i is the ground truth label of the ith training sample and s_i is the score predicted by the model. The Adam optimizer [43] was leveraged for optimization because of its advantage of processing sparse features and obtaining faster convergence compared with the normal stochastic gradient descent with momentum.

Ethics Approval

All materials were obtained from anonymous open-source data. Thus, ethics approval was not required.

Results

Experiment Setting and Evaluation Metrics

We implemented a prototype system for influence relationship identification on discussion threads. The prototype system and data sets used in the evaluation are publicly available at GitHub [44].

For empirical evaluation, we collected 25,208 threads that were publicly available in the CSN breast cancer forum [9]. The webpages were collected and processed by a web crawler we developed leveraging the Spider Crawler library [45]. There were 321,000 posts with 1.9 million sentences in total. We applied the classifier proposed by Li et al [1] on all 25,208 threads to identify the ones that were related to patient decision-making and obtained 11,815 (46.87%) such threads. Note that other models for classifying decision-making threads can also be plugged in.

We then extracted relationships from the decision-making threads using the relationship extraction module and obtained 9053 relationships. We randomly picked 853 (9.42%) of them to label. A total of 4 PhD students worked on the manual labeling. All the relationship triples and post pairs were first independently labeled. In case of disagreement, a consensus was reached after discussion. A total of 261 relationships were labeled as influence relationships. Recall that, per definition 1, each relationship is presented as a triple (p_A, p_B, p_C). We also labeled whether posts p_A and p_B were relevant (ie, P_{AB}) and whether posts p_B and p_C were relevant (ie, P_{BC}). We observed some reply posts with content expressing only comfort or wishes. Although they express care about the initial author's conditions and seem relevant, they are generic. After discussion, we reached an agreement that, when the initial post and reply post shared similar medical terms (such as *chemotherapy* and *chemo*), we would label them as relevant. All 1706 post pairs (p_A, p_B) and (p_B, p_C) of the 853 relationships were labeled. Of the 1706 pairs, 1210 (70.93%) were relevant pairs, and the remaining 496 (29.07%) were irrelevant. We split the set of relationships into a training set (90%) and a testing set (10%). The post pairs in the aforementioned training and test sets were used for text relevance training and testing, respectively.

The metrics used for evaluation included precision, recall, F_1 score, accuracy, area under the receiver operating characteristic curve (ROC AUC), and area under the precision-recall curve (PR AUC). They evaluated the effectiveness of a system using different aspects: (1) *precision*, also known as positive predictive value, is the fraction of relevant instances among the retrieved instances; (2) *recall*, also known as sensitivity, is the fraction of relevant instances that are retrieved among all relevant instances; (3) *F_1 score* measures a model's performance by calculating the harmonic mean of the precision and recall, as shown in the following equation: **(4)**; (4) *accuracy* is a common evaluation metric for binary classification problems and is defined as the fraction of corrected predictions among the total number of predictions; (5) *ROC AUC* is a common evaluation metric for binary classification problems and is created by plotting the true positive rate against the false positive rate at various threshold settings; and (6) *PR AUC* is commonly used to evaluate the performance of a model on data sets with imbalanced labels.

Text Relevance Evaluation

Table 1 presents the classification results of the text relevance measurement module. In total, 2 observations were made. The first observation was that the models using BERT achieved high recall but low precision, whereas the models with word-embedding vectors trained on OHC data obtained balanced precision and recall values. There are 2 reasons for these results. First, OHC data are domain-sensitive and can benefit from domain-specific word representation. Second, the BERT transformer tends to link words in adjacent sentences by mistake. In the text relevance measurement module, precision was more important than recall as the accuracy of influence relationship identification depended on the precision of relevance classification. Thus, we used the word vectors trained on OHC data instead of BERT in the following experiments.

Table 1. Text relevance measurement module results.

	Precision	Recall	F_1	Accuracy	ROC AUC[a]	PR AUC[b]
MatchPyramid with BERT[c] (trained on Wikipedia)	0.578	*0.992*[d]	0.730	0.512	0.502	0.583
MatchPyramid with word2vec (trained on the training data set)	0.781	0.820[d]	*0.806*	0.692	0.763	0.854
ARC-I[e] with BERT (trained on Wikipedia)	0.523	0.890[d]	0.659	0.503	0.493	0.554
ARC-I with word2vec (trained on the training data set)	*0.832*	0.747[d]	0.785	*0.784*	*0.848*	*0.903*

[a] ROC AUC: area under the receiver operating characteristic curve.

[b] PR AUC: area under the precision-recall curve.

[c] BERT: Bidirectional Encoder Representations from Transformers.

[d] The P value is statistically significant at P=.05.

[e] ARC-I: Architecture-I.

The second observation was that, with word vector embedding, ARC-I achieved a better performance than MatchPyramid in most of the evaluation metrics. In the ARC-I model, each input text goes through an embedding layer, a convolution layer, and a max pooling layer, and the extracted feature vectors are then concatenated together as the input to a fully connected layer that calculates the predicted relevance scores. MatchPyramid populates the local word relevance matrix first. Each cell of the matrix presents the dot product of the word-embedding vectors of the words in the text input. The patterns of these interactions are then extracted using a convolutional neural network [46]. Thus, ARC-I focuses on checking relevance based on the meaning of the whole text, whereas MatchPyramid focuses on summarizing the important relevance features based on local word similarity. For OHC data sets, posts were relatively long and often contained noisy information; thus, considering the meaning of the entire post text was more important than focusing on adjacent words. This is why the performance of ARC-I was better than that of MatchPyramid in our evaluation. We also observed that *ARC-I* with word2vec outperformed *MatchPyramidwith word2vec* in both *ROC AUC* and *PR AUC* but had an inferior *F_1 score*. Note that F_1 averages the performance of all the samples by combining the precision and recall, whereas the ROC AUC and PR AUC cumulate the precisions among all samples with different recall thresholds. This indicates that the average performance of *MatchPyramidwith word2vec* was better, but the overall performance of *ARC-Iwith word2vec* was better.

Question and Action Probability Evaluation

Now, we present the evaluation of the question probability calculation module and the action probability calculation module. The performance is shown in Table 2. Good performance was achieved for question identification. For future action identification, a high score was achieved on recall but not on precision. The following are a few examples of posts that are classified as containing future actions but actually do not have action intent: *I will tell you though I hated my silicone* or *I would worry about it*. These sentences have verbs in the future tense, but those verbs only convey opinions or feelings rather than taking action on health care. We plan to improve action detection by training action sentence models as future work.

Recall that in the *deep learning* approach, question and action probabilities are considered as input features instead of imposing a strict requirement on their presence. We conducted an analysis on the test data in terms of their presence. All positive cases either had a probability of action presence of 1.0 or had a high probability of question presence, with an average probability of 0.986 (SD 0.033). This indicates that the *deep learning* approach captures definition 2 well, ensuring the high likelihood that either a question or a future action is present.

Table 2. Question and action calculation module results.

	Precision	Recall	F_1	Accuracy	ROC AUC[a]	PR AUC[b]
Question probability calculation module	*1.000*	*1.000* [c]	*1.000*	*1.000*	*1.000*	*1.000*
Action probability calculation module	0.771	*1.000* [c]	0.871	0.810	0.733	0.771

[a] ROC AUC: area under the receiver operating characteristic curve.

[b] PR AUC: area under the precision-recall curve.

[c] The *P* value is statistically significant at *P*=.05.

Influence Relationship Classification Evaluation

Table 3 shows the performance of the *baseline* and *deep learning* approaches with alternative ways to combine text relevance vectors, question features, and action features. Recall that, for the feature combination module, *baseline* combines the text relevance score, the likelihood of question presence, and the likelihood of future action presence to identify influence relationships. *MatchPyramid+cat Q/A* represents the model using *MatchPyramid* to calculate the text relevance score and *cat* as the combination operator ⊗, whereas *MatchPyramid+dot Q/A* uses *dot* as the combination operator ⊗. *ARC-I+cat Q/A* represents the model using *ARC-I* to calculate the relevance score and *cat* as the combination operator ⊗, whereas *ARC-I+dot Q/A* uses *dot* as the combination operator ⊗.

Table 3. Influence relationship classification results.

	Precision	Recall	F_1	Accuracy	ROC AUC[a]	PR AUC[b]
Baseline	0.300	0.231[c]	0.261	0.595	0.495	0.307
MatchPyramid+cat Q/A[d]	0.667	0.154[c]	0.25	0.714	0.560	0.442
MatchPyramid+dot Q/A[e]	0.633	*0.577* [c]	*0.603*	0.667	0.634	0.481
ARC-I+cat Q/A[f]	0.667	0.154[c]	0.25	0.714	0.637	0.515
ARC-I+dot Q/A[g]	*0.750*	0.462[c]	0.571	*0.786*	*0.724*	*0.631*

[a] ROC AUC: area under the receiver operating characteristic curve.

[b] PR AUC: area under the precision-recall curve.

[c] The *P* value is statistically significant at *P*=.05.

[d] MatchPyramid+cat Q/A: model using *MatchPyramid* to calculate the text relevance score and *cat* as the combination operator ⊗.

[e] MatchPyramid+dot Q/A: model using *MatchPyramid* to calculate the text relevance score and *dot* as the combination operator ⊗.

[f] ARC-I+cat Q/A: model using *Architecture-I* to calculate the relevance score and *cat* as the combination operator ⊗.

[g] ARC-I+dot Q/A: model using *Architecture-I* to calculate the relevance score and *dot* as the combination operator ⊗.

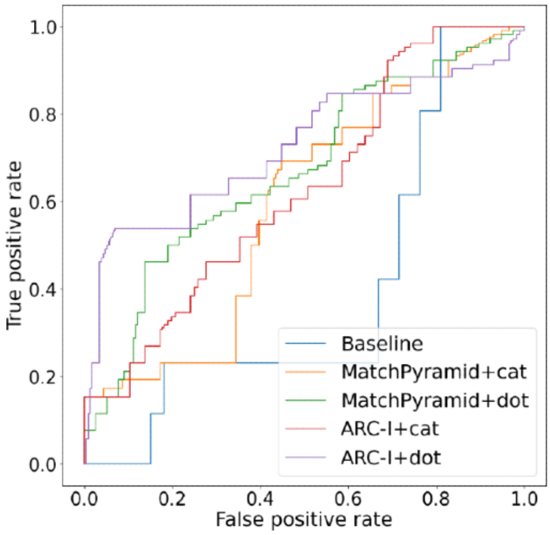
We also visualized the operating characteristic curves of all methods, as shown in Figure 5. From Table 3 and Figure 5, we have the following observations.

First, all proposed *deep learning* methods, which use relevance features and consider the interaction between relevance and the presence of questions or actions, significantly outperformed the *baseline* approach. This indicates that the relevance feature vectors generated by the text relevance measurement module were effective in capturing relevant content. Combining these feature vectors with the features of question presence and action presence helped capture their interactions and achieved good performance in influence relationship classification. In contrast, the *baseline* approach, which directly follows definition 2, did not perform well. This was due to the inability to capture the interactions between text relevance and question or action presence and the challenge of manually setting an appropriate cutoff threshold for each module.

Second, the models using the *dot* operator performed better than those using the *cat* operator. There are mainly 2 reasons for this. First, question probability and action probability may interact with V_{AB} and V_{BC} relevance vectors, which can be captured well by the *dot* operator. Figure 1B shows an example in which the action in p_C is related to the discussion in p_A and p_B. The action in p_{C2} is related to *chemo*, which is the common content of p_A and p_{B2}. In this case, the action probability needs to be combined with V_{AB}. Although, in another case, the action refers to an option mentioned in p_B, the interaction between p_B and p_C is more likely to be the context of the action and, thus, the action probability needs to be combined with V_{BC}. In contrast, the *cat* operator ignores some interactions between questions (actions) and the context because of the dropout of some neutrals. Therefore, the *cat*-based methods had a much lower recall than the *dot*-based methods. The results show that interactions between action and context are important for influence identification.

Furthermore, the *ARC-I+dot Q/A* had a much better precision, accuracy, ROC AUC, and PR AUC than *MatchPyramid+dot Q/A* but had lower recall and slightly lower F_1. This is because ARC-I achieved a better performance than MatchPyramid in the text relevance measurement module. *ARC-I+dot Q/A* was stricter than *MatchPyramid+dot Q/A* when fitting the model to the relevance factor. For applications that want to analyze the writing style and patterns of posts that have influence, precision is critical. *ARC-I+dot Q/A* is effective for locating such discussions. In contrast, for applications that want to check the information quality of the posts that have influence to prevent and mitigate the spread of misleading information, *MatchPyramid+dot Q/A* is more suitable because of its higher recall.

Figure 5. Influence relationship classification.



A Case Study

Figure 1 shows an example of 3 relationships, (p_A, p_{B1}, p_{C1}), (p_A, p_{B2}, p_{C2}), and (p_A, p_{B3}, p_{C3}), where (p_A is the initial post of the thread. The scores of these 3 relationships calculated using our system were 0.282, 0.793, and 0.622, respectively. Our system identified (p_A, p_{B2}, p_{C2}) and (p_A, p_{B3}, p_{C3}) as each containing an influence relationship, and (p_A, p_{B1}, p_{C1}) does not. As we can see from the post content, p_{B2} provides suggestions to the initial author regarding the treatment decision. In post p_{C2}, the initial author expresses actions to take based on the suggestions in p_{B2}. In post p_{B3}, the replier recommends that the author use wigs. The initial author then asks further questions about the wig information. Both relationships indicate that the initial author was influenced. In contrast, p_{B1} discusses general information and comforts the initial author, and the initial author expresses thanks in p_{C1}, but there is no indication of being influenced.

Discussion

Principal Findings

To the best of our knowledge, this is the first study that defines the influence relationships of discussion posts related to decision-making in OHCs. We proposed a deep learning–based natural language processing prototype to identify influence relationships. We then applied the developed techniques to identify the influence relationships in an OHC, the CSN breast cancer forum. There were 2 major observations.

First, we found that there is a significant amount of influence relationships in the OHC. Of the 9052 relationships in decision-making threads identified by Li et al [1], 3069 (33.9%) were identified as influence relationships. That is, approximately one-third of the communications influence the initial authors on their decision-making. Furthermore, of the 5143 decision-making threads, which have at least one relationship, 2417 (47%) contain at least one influence relationship. Owing to the prevalence, it is important to study posts that have influence.

Second, we also observed that posts that have influence may contribute to engaging users in discussions. The average number of posts in threads containing at least one influence relationship was 15.5, whereas the average number of posts in threads containing no influence relationship was 12.6. Our conjecture is that posts that have an influence likely provide helpful information or good reasoning, which are thought-provoking and help engage users in discussions.

On the basis of these observations, there are several applications that can benefit from the identification and analysis of influence relationships.

First, analyzing the quality of posts that have influence helps improve the quality of the influence. As discussed in the first observation, influence relationships are common. Quality checking of those posts is more critical than that of other posts in terms of improving the effect of influences and mitigating the spread of misleading information.

On the basis of the identification of influence relationships, we can further identify influential users in OHCs. We can use existing techniques that analyze the network features to identify influential users [16-20], where this work calculates the edge weights (ie, the influence of a post). Identifying and checking influential users contributes to high-quality information dissemination.

Second, based on the second observation, analyzing the writing style of posts that have influence provides insights to health care professionals about effective communication for patient engagement.

Furthermore, identifying influence relationships contributes to effective information recommendations for addressing the information overload problem. When a user searches for information in OHCs, it is important to rank discussion threads and posts and recommend to users the most relevant and helpful discussions. On the basis of the analysis of influence relationships and the second observation, discussions that contain influence relationships are more likely to provide helpful information and encourage patient engagement. Thus, the presence of influence relationships is a positive factor in ranking.

Limitations

Our results are not without limitations. First, our definition of relationship was based on 3 posts, including the initial post in the thread. Therefore, we only identified the posts that had an influence on the initial author. However, any 3 posts that have a sequential reply relationship with the first and third posts from the same author can represent a relationship. We conjecture that the proposed techniques can be used to identify influence relationships among the generalized relationships and plan to study that problem in the future. Second, in this study, we considered text relevance between the posts in the relationship. Sometimes, even though 2 posts, p_B and p_C, are relevant overall, the specific sentence that has a question or future action indication in p_C may not be relevant to the suggestions in p_B. In addition, the current technique for future action detection sometimes generates false positives. To address these issues, we will investigate how to leverage part-of-speech and reference resolution techniques [47] to improve natural language understanding.

Conclusions and Future Work

We studied the problem of identifying influence relationships of web-based discussions and developed techniques and a prototype system for identifying influence relationships in OHCs. The proposed deep learning model demonstrates the performance advantage of the compared methods. As future work, we will address the aforementioned limitations to improve the generality and accuracy of the proposed techniques.

Acknowledgments

This material is based on work partially supported by the Leir Foundation and a grant from the National Institutes of Health (UL1TR003017).

Conflicts of Interest

None declared.

References

1. Li M, Shi J, Chen Y. Analyzing patient decision making in online health communities. In: Proceedings of the 2019 IEEE International Conference on Healthcare Informatics. 2019 Presented at: ICHI '19; June 10-13, 2019; Xi'an, China p. 1-8. [doi: 10.1109/ICHI.2019.8904879]
2. Fan H, Smith SP, Lederman R, Chang S. Why people trust in online health communities: an integrated approach. In: Proceedings of the 21st Australasian Conference on Information Systems. 2010 Presented at: ACIS '10; August 18-20, 2010; Yamagata, Japan.
3. Kim HS, Mrotek A. A functional and structural diagnosis of online health communities sustainability: a focus on resource richness and site design features. Comput Human Behav 2016 Oct;63:362-372. [doi: 10.1016/j.chb.2016.05.004]
4. Demiris G. The diffusion of virtual communities in health care: concepts and challenges. Patient Educ Couns 2006 Aug;62(2):178-188. [doi: 10.1016/j.pec.2005.10.003] [Medline: 16406472]

5. Nasralah T, Noteboom C, Wahbeh A, Al-Ramahi MA. Online health recommendation system: a social support perspective. Dakota State University. 2017. URL: https://scholar.dsu.edu/cgi/viewcontent.cgi?article=1008&context=bispapers [accessed 2022-06-10]
6. Ji X, Geller J, Chun S. Social InfoButtons for patient-oriented healthcare knowledge support. In: Proceedings of the 2014 AMIA Annual Symposium. 2014 Presented at: AMIA '14; November 15-19, 2014; Washington, DC, USA.
7. Waterworth S, Luker KA. Reluctant collaborators: do patients want to be involved in decisions concerning care? J Adv Nurs 1990 Aug;15(8):971-976. [doi: 10.1111/j.1365-2648.1990.tb01953.x] [Medline: 2229694]
8. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
9. Breast Cancer. Cancer Survivors Network. URL: https://csn.cancer.org/forum/127 [accessed 2022-06-10]
10. Yaraghi N, Wang W, Gao GG, Agarwal R. How online quality ratings influence patients' choice of medical providers: controlled experimental survey study. J Med Internet Res 2018 Mar 26;20(3):e99 [FREE Full text] [doi: 10.2196/jmir.8986] [Medline: 29581091]
11. Zhang D, Yue WT. Social media use in decision making: special issue of decision support systems for the 10th workshop on e-business. Decis Support Syst 2014 Jul;63:65-66. [doi: 10.1016/j.dss.2013.08.007]
12. Wu B, Jiang S, Chen H. Effects of individuals' motivations on communications in online health forums. Soc Behav Pers 2016 Mar 23;44(2):299-312. [doi: 10.2224/sbp.2016.44.2.299]
13. Pan W, Shen C, Feng B. You get what you give: understanding reply reciprocity and social capital in online health support forums. J Health Commun 2017 Jan;22(1):45-52. [doi: 10.1080/10810730.2016.1250845] [Medline: 28027009]
14. Li S, Feng B, Wingate VS. Give thanks for a little and you will find a lot: the role of a support seeker's reply in online support provision. Commun Monogr 2019;86(2):251-270. [doi: 10.1080/03637751.2018.1539237]
15. Yu J, Hu Y, Yu M, Di Z. Analyzing netizens' view and reply behaviors on the forum. Phys A Stat Mech Appl 2010 Aug;389(16):3267-3273. [doi: 10.1016/j.physa.2010.03.043]
16. Tang X, Yang CC. Identifying influential users in an online healthcare social network. In: Proceedings of the 2010 IEEE International Conference on Intelligence and Security Informatics. 2010 Presented at: ISI '10; May 23-26, 2010; Vancouver, Canada p. 43-48. [doi: 10.1109/ISI.2010.5484779]
17. Cha M, Haddadi H, Benevenuto F, Gummadi K. Measuring user influence in Twitter: the million follower fallacy. Proc Conf AAAI Artif Intell 2010;4(1):10-17.
18. Ishfaq U, Khan HU, Iqbal K. Identifying the influential bloggers: a modular approach based on sentiment analysis. J Web Eng 2017;16(5-6):505-523.
19. Zhao K, Qiu B, Caragea C, Wu D, Mitra P, Yen J, et al. Identifying leaders in an online cancer survivor community. In: Proceedings of the 21st Annual Workshop on Information Technologies and Systems. 2011 Presented at: WITS '11; December 3-4, 2011; Shanghai, China p. 115-120.
20. Hou X, Lei CU, Kwok YK. OP-DCI: a riskless K-means clustering for influential user identification in MOOC forum. In: Proceedings of the 16th IEEE International Conference on Machine Learning and Applications. 2017 Presented at: ICMLA '17; December 18-21, 2017; Cancun, Mexico p. 936-939. [doi: 10.1109/icmla.2017.00-34]
21. Zhao K, Yen J, Greer G, Qiu B, Mitra P, Portier K. Finding influential users of online health communities: a new metric based on sentiment influence. J Am Med Inform Assoc 2014 Oct;21(e2):e212-e218 [FREE Full text] [doi: 10.1136/amiajnl-2013-002282] [Medline: 24449805]
22. Cercel DC, Trausan-Matu S. Opinion influence analysis in online forum threads. In: Proceedings of the 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. 2014 Presented at: SYNASC '14; September 22-25, 2014; Timisoara, Romania p. 228-235. [doi: 10.1109/synasc.2014.38]
23. Qiu B, Zhao K, Mitra P, Wu D, Caragea C, Yen J, et al. Get online support, feel better--sentiment analysis and dynamics in an online cancer survivor community. In: Proceedings of the IEEE 3rd International Conference on Privacy, Security, Risk and Trust and IEEE 3rd International Conference on Social Computing. 2011 Presented at: SocialCom '11; October 9-11, 2011; Boston, MA, USA p. 274-281. [doi: 10.1109/PASSAT/SocialCom.2011.127]
24. Definition of influence. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/influence [accessed 2022-06-10]
25. Kidd C, Hayden BY. The psychology and neuroscience of curiosity. Neuron 2015 Nov 04;88(3):449-460 [FREE Full text] [doi: 10.1016/j.neuron.2015.09.010] [Medline: 26539887]
26. Palloff RM, Pratt K. Lessons from the Virtual Classroom: The Realities of Online Teaching. 2nd edition. Hoboken, NJ, USA: John Wiley & Sons; 2013.
27. Adjei MT, Noble SM, Noble CH. The influence of C2C communications in online brand communities on customer purchase behavior. J Acad Mark Sci 2009 Nov 13;38(5):634-653. [doi: 10.1007/s11747-009-0178-5]
28. Liu Y, Chen F, Chen Y. Learning thread reply structure on patient forums. In: Proceedings of the 2013 International Workshop on Data Management & Analytics for Healthcare. 2013 Presented at: DARE '13; November 1, 2013; San Francisco, CA, USA p. 1-4. [doi: 10.1145/2512410.2512426]
29. Huang PS, He X, Gao J, Deng L, Acero A, Heck L. Learning deep structured semantic models for web search using clickthrough data. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management.

2013 Presented at: CIKM '13; October 27-November 1, 2013; San Francisco, CA, USA p. 2333-2338. [doi: 10.1145/2505515.2505665]

30. Shen Y, He X, Gao J, Deng L, Mesnil G. Learning semantic representations using convolutional neural networks for Web search. In: Proceedings of the 23rd International Conference on World Wide Web. 2014 Presented at: WWW '14 Companion; April 7-11, 2014; Seoul, South Korea p. 373-374. [doi: 10.1145/2567948.2577348]
31. Hu B, Lu Z, Li H, Chen Q. Convolutional neural network architectures for matching natural language sentences. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. 2014 Presented at: NIPS '14; December 8-13, 2014; Montreal, Canada p. 2042-2050.
32. Wang YC, Kraut R, Levine JM. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12; February 11-15, 2012; Seattle, WA, USA p. 833-842. [doi: 10.1145/2145204.2145329]
33. Lu Z, Li H. A deep architecture for matching short texts. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1. 2013 Presented at: NIPS '13; December 5-10, 2013; Lake Tahoe, NV, USA p. 1367-1375. [doi: 10.7551/mitpress/11474.003.0014]
34. Pang L, Lan Y, Guo J, Xu J, Wan S, Cheng X. Text matching as image recognition. Proc Conf AAAI Artif Intell 2016 Mar 5;30(1):10341. [doi: 10.1609/aaai.v30i1.10341]
35. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv 2018 Oct 11. [doi: 10.48550/arXiv.1810.04805]
36. Efron M, Winget M. Questions are content: a taxonomy of questions in a microblogging environment. Proc Am Soc Info Sci Tech 2011 Feb 03;47(1):1-10. [doi: 10.1002/meet.14504701208]
37. Liu Z, Jansen BJ. Questioner or question: predicting the response rate in social question and answering on Sina Weibo. Inf Process Manag 2018 Mar;54(2):159-174. [doi: 10.1016/j.ipm.2017.10.004]
38. Ranganath S, Hu X, Tang J, Wang S, Liu H. Identifying rhetorical questions in social media. Proc Conf AAAI Artif Intell 2016;10(1):667-670.
39. Ranganath S, Hu X, Tang J, Wang S, Liu H. Understanding and identifying rhetorical questions in social media. ACM Trans Intell Syst Technol 2018 Mar 31;9(2):1-22. [doi: 10.1145/3108364]
40. Ojokoh B, Igbe T, Araoye A, Ameh F. Question identification and classification on an academic question answering site. In: Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries. 2016 Presented at: JCDL '16; June 19-23, 2016; Newark, NJ, USA p. 223-224. [doi: 10.1145/2910896.2925442]
41. Bird S, Loper E. NLTK: the natural language toolkit. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions. 2004 Presented at: ACL '04; July 21-26, 2004; Barcelona, Spain p. 214-217.
42. Goodfellow I, Yoshua B, Courville A. Deep Learning. Cambridge, MA, USA: MIT Press; Nov 2016:4.
43. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. 2015 Presented at: ICLR '15; May 7-9, 2015; San Diego, CA, USA.
44. Mingda Li, Jinhe Shi, Yi Chen. NJIT-AI-in-Healthcare / Influence-Relationship-Classification. GitHub. 2022 Feb 27. URL: https://github.com/NJIT-AI-in-Healthcare/Influence-Relationship-Classification [accessed 2022-06-10]
45. Scrapy. URL: https://scrapy.org/ [accessed 2022-06-10]
46. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 Jun;60(6):84-90. [doi: 10.1145/3065386]
47. Liu Y, Shi J, Chen Y. Patient-centered and experience-aware mining for effective adverse drug reaction discovery in online health forums. J Assoc Inf Sci Technol 2017 Sep 22;69(2):215-228. [doi: 10.1002/asi.23929]

Abbreviations

ARC-I: Architecture-I
ARC-II: Architecture-II
BERT: Bidirectional Encoder Representations from Transformers
CSN: Cancer Survivors Network
DSSM: Deep Structured Semantic Models
OHC: online health community
PR AUC: area under the precision-recall curve
ROC AUC: area under the receiver operating characteristic curve

Edited by R Kukafka; submitted 03.06.21; peer-reviewed by S Liu, G Lim; comments to author 02.08.21; revised version received 05.10.21; accepted 21.03.22; published 31.08.22.

Please cite as:
Li M, Shi J, Chen Y
Identifying Influences in Patient Decision-making Processes in Online Health Communities: Data Science Approach
J Med Internet Res 2022;24(8):e30634
URL: https://www.jmir.org/2022/8/e30634
doi:10.2196/30634
PMID:36044266

Original Paper

Improving the Performance of Outcome Prediction for Inpatients With Acute Myocardial Infarction Based on Embedding Representation Learned From Electronic Medical Records: Development and Validation Study

Yanqun Huang[1,2], BS; Zhimin Zheng[1,2], BS; Moxuan Ma[1,2], MS; Xin Xin[1,2], MS; Honglei Liu[1,2], PhD; Xiaolu Fei[3], PhD; Lan Wei[3], PhD; Hui Chen[1,2], PhD

[1]School of Biomedical Engineering, Capital Medical University, Beijing, China
[2]Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, China
[3]Information Center, Xuanwu Hospital, Capital Medical University, Beijing, China

Corresponding Author:
Hui Chen, PhD
School of Biomedical Engineering
Capital Medical University
No. 10, Xitoutiao, You An Men, Fengtai District
Beijing, 100069
China
Phone: 86 01083911545
Email: chenhui@ccmu.edu.cn

Abstract

Background: The widespread secondary use of electronic medical records (EMRs) promotes health care quality improvement. Representation learning that can automatically extract hidden information from EMR data has gained increasing attention.

Objective: We aimed to propose a patient representation with more feature associations and task-specific feature importance to improve the outcome prediction performance for inpatients with acute myocardial infarction (AMI).

Methods: Medical concepts, including patients' age, gender, disease diagnoses, laboratory tests, structured radiological features, procedures, and medications, were first embedded into real-value vectors using the improved skip-gram algorithm, where concepts in the context windows were selected by feature association strengths measured by association rule confidence. Then, each patient was represented as the sum of the feature embeddings weighted by the task-specific feature importance, which was applied to facilitate predictive model prediction from global and local perspectives. We finally applied the proposed patient representation into mortality risk prediction for 3010 and 1671 AMI inpatients from a public data set and a private data set, respectively, and compared it with several reference representation methods in terms of the area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), and F1-score.

Results: Compared with the reference methods, the proposed embedding-based representation showed consistently superior predictive performance on the 2 data sets, achieving mean AUROCs of 0.878 and 0.973, AUPRCs of 0.220 and 0.505, and F1-scores of 0.376 and 0.674 for the public and private data sets, respectively, while the greatest AUROCs, AUPRCs, and F1-scores among the reference methods were 0.847 and 0.939, 0.196 and 0.283, and 0.344 and 0.361 for the public and private data sets, respectively. Feature importance integrated in patient representation reflected features that were also critical in prediction tasks and clinical practice.

Conclusions: The introduction of feature associations and feature importance facilitated an effective patient representation and contributed to prediction performance improvement and model interpretation.

(J Med Internet Res 2022;24(8):e37486) doi:10.2196/37486

KEYWORDS

representation learning; skip-gram; feature association strengths; feature importance; mortality risk prediction; acute myocardial infarction

Introduction

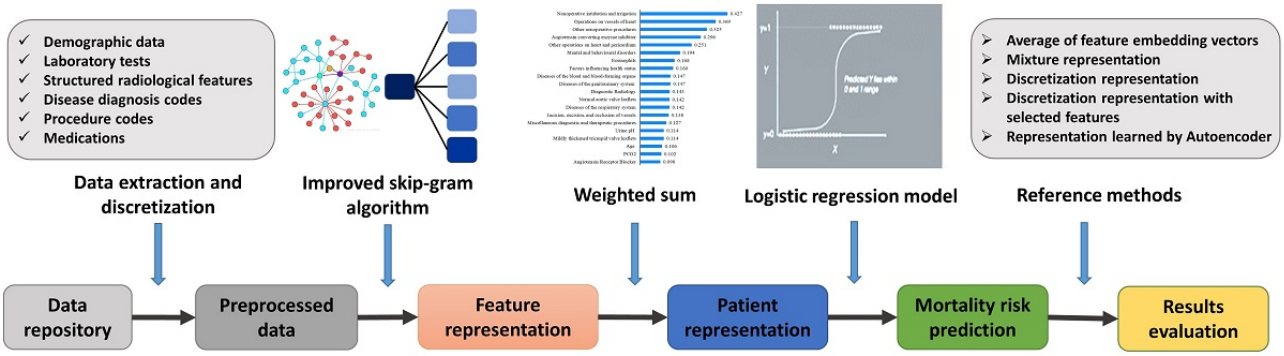
Electronic medical records (EMRs) contain diverse and heterogeneous information, such as demographic data, disease diagnoses, laboratory tests, radiological findings, examinations and procedures, and medications. EMR data can be used to not only reflect the health status of patients and record the treatment trajectory, but also help doctors in making clinical decisions [1-6] and improving the efficiency of diagnosis and treatment [1,7,8]. One of the most prevalent and practical tasks of the secondary use of EMR data is building models to predict the disease status [8-10] and treatment outcomes [11-17] for a patient, using machine learning algorithms.

However, the high dimensionality, sparsity, and heterogeneity of EMR data [12,18] pose many obstacles for directly inputting the raw data into machine learning–based predictive models. Some manual and data-driven feature engineering methods [15,19], though time-consuming and laborious, were used to select important features or extract useful information for predictive tasks. Moreover, the performance of predictive models relies heavily on the representation of data. It was reported that effective representation methods could make the downstream modeling simpler and more flexible, and greatly improve the predictive performance [18,20]. By transforming raw features into compact vectors, representation learning can make it easier to automatically extract useful information when building predictive models [16,21,22]. One widely used representation method for EMR data is the skip-gram algorithm [23], a distributed embedding method that treats patient records as sentences and medical concepts as words. An inevitable problem in the skip-gram algorithm is that contrary to words within a sentence, medical concepts in a patient's record do not have a natural order, making it difficult to learn meaningful representations of concepts that have potential associations. One solution for this problem was randomly shuffling the concepts within a record to learn concept embeddings [12,24-26]. It could reduce the impact of the disorder attribute of medical concepts on the algorithm to some degree, while associations among these concepts were still not taken into consideration.

Acute myocardial infarction (AMI) is an acute ischemic heart disease and is the second leading cause of death. One in every 6 deaths is caused by ischemic heart disease, where AMI accounts for the majority of deaths [27,28]. Mortality risk prediction for AMI patients plays a crucial role in clinical work, helping doctors identify potential clinical factors, take early intervention measures based on timely alerts of patients' adverse health statuses, and reduce the burdensome expenditure of related health care expenses. Therefore, researchers [19,29-31] have focused on building machine learning models for the outcome prediction of AMI patients, and most of them used specific clinical features, such as laboratory test results (eg, albumin), comorbidities (eg, diabetes), and demographic data (eg, gender).

In this study, we aimed to represent various structured features extracted from EMR data as fixed-length embedding vectors, which were then used to improve the performance of predictive models for the death risk of AMI patients. Specifically, we introduced the association strengths into the skip-gram algorithm to learn more informative representations of features. We also introduced the Shapley additive explanations (SHAP) [32] technique to facilitate representation at the patient level and enhance the interpretability of the predictive model. An overview of our proposed representation learning framework and its application is shown in Figure 1.

Figure 1. Overview of the proposed representation learning method for patients' mortality risk prediction. First, feature representations were learned by the skip-gram algorithm using an adaptive context window. Then, patient representations were constructed based on feature representations weighted by the feature importance. Finally, the proposed patient representation was applied in the mortality risk prediction for acute myocardial infarction in-patients from a public data set and a private data set, and compared with reference methods.



Methods

Skip-Gram–Based Patient Representation

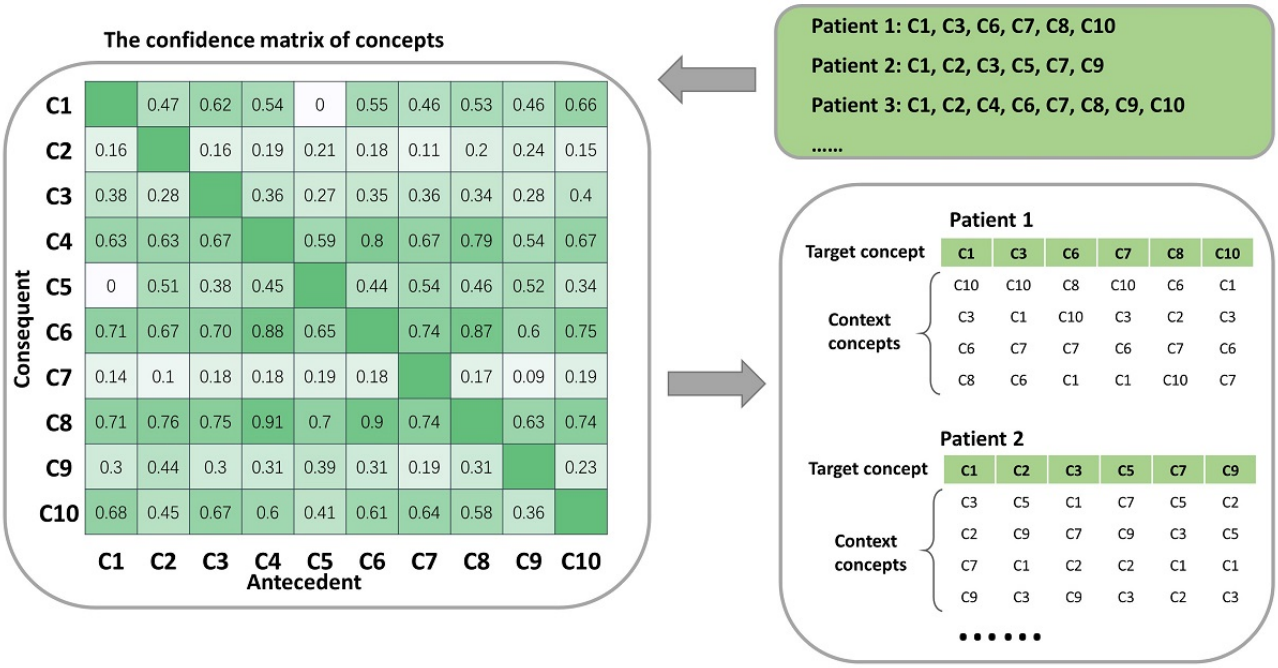
The representation was learned hierarchically at the following 3 levels: the concept, feature, and patient levels. At the concept level, we employed the improved skip-gram algorithm [23] to represent a concept as an embedding vector. In the natural language processing domain, the basic idea of skip-gram was to maximize the occurrence probabilities of the target words and the context words in the predefined context window, making the words that appear in the same context window closer in the embedding space. Unlike words with natural orders in a sentence, medical concepts appeared out of order in a patient record for a certain hospital stay. This made it difficult to determine the context window that contained relevant concepts for the target concept, especially when the number of concepts in a record was far larger than the size of the context window.

Thus, for a concept within a record, we identified relevant concepts using its association strength with a candidate concept in the same record. The association strength was defined as the confidence (equation 1) of an association rule with one candidate concept as the unique antecedent (or consequent).

$$\text{Confidence}\ (C1, C2) = |C1 \cap C2| \,/\, |C1| \quad \textbf{(1)}$$

where C1 and C2 are the antecedent and consequent concepts, respectively, of an association rule C1→C2, and |C1| and |C1∩C2| are the numbers of patient records containing C1 and both C1 and C2, respectively. The greater the confidence, the stronger the association between the 2 concepts. Antecedent (or consequent) concepts in association rules with the top N highest confidences were included in the context window of the target concept. We called these selection schemes of context concepts *antecedent-based* (or *consequent-based*) embeddings. Figure 2 provides an example of the consequent-based selection scheme of context concepts.

Figure 2. An illustration of context concept selection for the skip-gram algorithm using association strengths. All records are composed of 10 concepts (C1, C2, ……, and C10). In the confidence matrix, element Cij was the confidence of the association rule with Cj as antecedent and Ci as consequent. For patient 1 with 6 concepts (C1, C3, C6, C7, C8, and C10), the included concepts in C1's 4-concept context window were selected from the remining 5 candidate concepts, whose confidences were 0.66 (antecedent, C10), 0.62 (C3), 0.55 (C6), 0.53 (C8), and 0.46 (C7). Therefore, C10, C3, C6, and C8 were selected to construct the context window for C1.



The confidence matrix of concepts

Consequent	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1		0.47	0.62	0.54	0	0.55	0.46	0.53	0.46	0.66
C2	0.16		0.16	0.19	0.21	0.18	0.11	0.2	0.24	0.15
C3	0.38	0.28		0.36	0.27	0.35	0.36	0.34	0.28	0.4
C4	0.63	0.63	0.67		0.59	0.8	0.67	0.79	0.54	0.67
C5	0	0.51	0.38	0.45		0.44	0.54	0.46	0.52	0.34
C6	0.71	0.67	0.70	0.88	0.65		0.74	0.87	0.6	0.75
C7	0.14	0.1	0.18	0.18	0.19	0.18		0.17	0.09	0.19
C8	0.71	0.76	0.75	0.91	0.7	0.9	0.74		0.63	0.74
C9	0.3	0.44	0.3	0.31	0.39	0.31	0.19	0.31		0.23
C10	0.68	0.45	0.67	0.6	0.41	0.61	0.64	0.58	0.36	

Antecedent

Patient 1: C1, C3, C6, C7, C8, C10

Patient 2: C1, C2, C3, C5, C7, C9

Patient 3: C1, C2, C4, C6, C7, C8, C9, C10

……

Patient 1

Target concept	C1	C3	C6	C7	C8	C10
Context concepts	C10	C10	C8	C10	C6	C1
	C3	C1	C10	C3	C2	C3
	C6	C7	C7	C6	C7	C6
	C8	C6	C1	C1	C10	C7

Patient 2

Target concept	C1	C2	C3	C5	C7	C9
Context concepts	C3	C5	C1	C7	C5	C2
	C2	C9	C7	C9	C3	C5
	C7	C1	C2	C2	C1	C1
	C9	C3	C9	C3	C2	C3

••••••

Moreover, to reduce the high dimensionality and sparsity of a large number of concepts, while preserving the clinical information as much as possible, we aggregated the concepts of disease diagnoses and procedures into several clinically meaningful feature groups according to International Classification of Diseases, 10th revision (ICD-10) codes and International Classification of Diseases, 9th revision (ICD-9) codes, and with the help of clinical experts. For example, disease diagnoses of type 1 diabetes mellitus and type 2 diabetes mellitus were grouped into the feature group of diabetes mellitus. The average of all embedding vectors of concepts from the same feature group in a patient record was treated as the representation at the feature level for the patient.

The representation at the patient level was the weighted sum of feature-level representations. The feature weights were obtained under the guidance of the predictive task, indicating the importance of each feature involved in the patient representation. In this study, we used SHAP values as the feature weights. The SHAP framework is a machine learning interpretation technique based on the idea of game theory. It approximated a trained prediction model with a different but simple model that could easily calculate the contribution in the form of a SHAP value for each feature in the prediction model and performed additive feature attribution to explain the combination of features [32]. A positive or negative SHAP value reflected a positive or negative influence on the prediction. A feature's importance was then computed as the average of its absolute SHAP values from all samples.

Experiments and Evaluations

Data Sets and Data Preprocessing

In this study, we used a public data set, the freely accessible critical care database Medical Information Mart for Intensive Care III (MIMIC-III data set [33]), and a private data set for the experiments.

The MIMIC-III data set was collected between June 2001 and October 2012, and involved 46,520 patients admitted to intensive care units at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. It includes patient health information, such as demographics, vital signs, laboratory test results, medications, procedures, diagnosis codes, and clinical notes. The informative MIMIC-III data set was widely used in some medical machine learning modeling and algorithm evaluations, providing strong data support for researchers to establish models and evaluate algorithms [14,18].

The private data set was derived from the EMR system of a tertiary hospital, Xuanwu Hospital, Capital Medical University, Beijing, China, between January 2014 and December 2016. Patient features included hospital admission and discharge information, demographic data, disease diagnoses, laboratory tests, examinations and procedures, medications, and radiology reports of chest X-ray or color sonography examination.

We extracted the records of all 3010 and 1671 AMI patients from the public and private data sets, respectively. The diagnosis of AMI was confirmed with the ICD-9 codes 410.01 to 410.91 or ICD-10 codes I21 and I22. There were 254 (8.1%) and 103 (6.2%) patients who died in the hospital from the public and private data sets, respectively.

We maintained patients' first hospitalization data to evaluate the proposed method. Demographic data (age and gender) and the following AMI-related features were maintained in both data sets: AMI-relevant items of laboratory tests that at least 95% of patients carried out, AMI-relevant radiological features extracted from radiology reports [34], 7 commonly prescribed medications, and all recorded disease diagnoses and procedures. For laboratory tests performed more than once, only the results obtained in the first test (usually at admission) were retained, which could reflect a patient's health status and the severity of illness.

Since initially proposed in the field of natural language processing, the skip-gram algorithm was used to train embeddings for discrete words or symbols. Therefore, to use the skip-gram algorithm for the embedding representation of the structured data, all patient features should be categorical, where each discrete value is treated as a concept. For example, male and female were 2 concepts for gender. Different from raw categorical features, such as gender, disease diagnoses, procedures, and medications, that might remain unchanged, the continuous variables age and laboratory test results had to be discretized into two or more concepts. Age was discretized into 2 concepts (>60 years and ≤60 years). Each laboratory test result was also discretized into 2 concepts (normal and abnormal with reference to clinical standards). In total, 3326 and 1073 medical concepts were identified and further aggregated to 104 and 108 feature groups in the public and private data sets, respectively (Table 1). All feature groups of the private and public data sets are listed in Multimedia Appendix 1 and Multimedia Appendix 2, respectively.

Table 1. Concepts and feature groups of both the public and private data sets.

Feature category	Public data set		Private data set		Concept examples
	Feature groups (n=104), n	Concepts (n=3326), n	Feature groups (n=108), n	Concepts (n=1073), n	
Age	1	2	1	2	>60 years and ≤60 years
Gender	1	2	1	2	Male and female
Laboratory tests	19	38	40	80	Abnormal serum triglyceride and normal serum creatinine
Radiological features	34	34	36	36	Cardiac image enlargement and sharp costophrenic angle
Disease diagnoses	24	2600	15	739	Hypertension and brainstem infarction
Procedures	18	643	8	207	Coronary stenting and pericardiocentesis
Medications	7	7	7	7	Angiotensin-converting enzyme inhibitor and heparin

Representation Evaluation

To evaluate the effectiveness and advantages of the proposed representation, we used 2 additional kinds of simple reference representation methods, namely, the 3-layer autoencoder with learning and the feature selection method without learning. Table 2 describes the details of the proposed and reference representation methods.

The proposed representation method was first evaluated at the concept level. Cluster analyses were used to cluster laboratory test concepts into 2 clusters for the quantitative evaluation. The adjusted Rand index (ARI) [35] (ranging from −1 to 1) was used to evaluate the cluster solutions. Greater ARI values indicated higher ability of discriminating from categories with different real labels (normal and abnormal). We also applied the t-distributed stochastic neighbor algorithm to project the embedding vectors of laboratory test concepts into a 2-dimensional space to visually observe the distribution of embeddings.

The proposed representation method was then evaluated at the patient level with a downstream prediction task using the logistic regression model. The predicted outcome was the in-hospital death of AMI patients during hospital stay. The input for prediction was the patient representation derived from the entire feature set listed in Table 1. We also extracted a treatment-free feature subset that excluded medications and procedures from the entire feature set, trying to clarify that the performance of the proposed patient representation was related to the features that were involved in the representation and that the treatment-related features played a crucial role in predicting patient outcome even if they had been represented as embedding vectors.

Table 2. Descriptions of the proposed and reference representation methods.

Representation method	Descriptions	Representation examples
Mixture	The mixture of discretization codes for original discrete features and original values for continuous features. The missing values in the laboratory tests were interpolated using the mean of the corresponding laboratory tests.	(0,1,1,0,0,0,1,12,8.5,3,8) for a patient with 11 features
Discretization	The 0-1 vector where the digit 1 represented the patient having the specific disease, procedure, radiological feature, and medication, and 0 otherwise. Age of 1 meant >60 years and 0 meant ≤60 years, gender of 1 meant male and 0 meant female, and a laboratory test item of 1 meant abnormal and 0 meant normal. Missing values for laboratory tests were interpolated by the corresponding mode.	(0,1,1,0,0,0,1,1,0,1,1) for a patient with 11 discretization features
DIS_FS[a]	The selected features with discretization representations were statistically different between patients with and without the label "death."	(0,0,1,0,0,1,0,1) for a patient with 8 selected features
DIS_AE[b]	The hidden-layer vector of a 3-layer autoencoder with discretization vectors as inputs and outputs. The dimension of the hidden layer was set to 64.	(0.7,1.9,0.5,−1,−3.1,2.4) for a patient with a 6-dimensional vector
RAN_EM_AVE[c]	The average of feature embedding vectors learned from the skip-gram algorithm using the random selection method to determine the context window.	(1.6,−0.5,1.1,0.1,−1.3,0.6) for a patient with a 6-dimensional embedding vector
RAN_EM_WGT[d]	The weighted sum of the feature embedding vectors learned from the skip-gram algorithm using the random selection method to determine the context window.	(1.2,−0.9,1.3,0.4,−1.9,1.0) for a patient with a 6-dimensional embedding vector
ANT_EM_AVE[e]	The average of the feature embedding vectors learned from the skip-gram algorithm using the confidence with the target concept as the antecedent.	(0.9,−0.6,1.2,1.4,−1.9,0.6) for a patient with a 6-dimensional embedding vector
ANT_EM_WGT[f]	The weighted sum of the feature embedding vectors learned from the skip-gram algorithm using the confidence with the target concept as the antecedent.	(1.2,−1.5,1.1,0.1,−0.6,0.6) for a patient with a 6-dimensional embedding vector
CON_EM_AVE[g]	The average of the feature embedding vectors learned from the skip-gram algorithm using the confidence with the target concept as the consequent.	(1.6,−0.8,2.1,1.6,−1.4,1.5) for a patient with a 6-dimensional embedding vector
CON_EM_WGT[h]	The weighted sum of the feature embedding vectors learned from the skip-gram algorithm using the confidence with the target concept as the consequent.	(1.1,−0.4,−0.7,1.6,−0.3,0.9) for a patient with a 6-dimensional embedding vector

[a] DIS_FS: discretization representations with feature selection.

[b] DIS_AE: hidden vector of an autoencoder-based representation.

[c] RAN_EM_AVE: average of the random selection–based embedding representation.

[d] RAN_EM_WGT: weighted sum of the random selection–based embedding representation.

[e] ANT_EM_AVE: average of the antecedent-based embedding representation.

[f] ANT_EM_WGT: weighted sum of the antecedent-based embedding representation.

[g] CON_EM_AVE: average of the consequent-based embedding representation.

[h] CON_EM_WGT: weighted sum of the consequent-based embedding representation.

We randomly split samples into training and test data sets by the ratio of 7:3. The training samples were first represented in the discretization vectors and used to build a predictive model for calculating all features' SHAP values for the further patient embedding representations of all study samples. After being represented as embedding vectors, the training and test samples were used to build and validate a logistic regression-based predictive model, respectively. The area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), and F1-score were the main evaluation metrics. Other relevant performance metrics from the confusion matrix included precision, recall, and accuracy. To eliminate the performance bias introduced by the initialization of a skip-gram model and the training/test data set split, we performed the comparative experiment 100 times. In each experiment round, the above processes were repeated. The mean with its 95% CI of each performance evaluation metric was reported.

In the skip-gram algorithm, the size of the context window and the dimension of the embedding vector were determined by trial and error. We conducted a group of predictive experiments on the public data set, using possible combinations of window sizes of 5, 10, 15, and 20, and vector dimensions of 50, 100, 200, and 300. Experimental results (listed in Multimedia Appendix 3) showed that the skip-gram algorithm with the combination of a window size of 10 and a vector dimension of 300 had the highest representation performance. Therefore, the size of the context window and the dimension of the embedding vector were set to 10 and 300, respectively. We applied the negative sampling mechanism (20 negative samples in this study) to accelerate the concept embedding training process. Other parameters were as follows: learning rate, 0.001; number of iterations, 50; batch size, 64. The gradient calculation method was Adam. We implemented representation learning, SHAP value computation, and prediction modeling in Python 3.7 and TensorFlow 2.0. In the step of patient representation, we used the L2 regularization penalty with "liblinear" solver for the logistic regression model, and the inverse of regularization strength was set to 0.1.

Ethics Approval

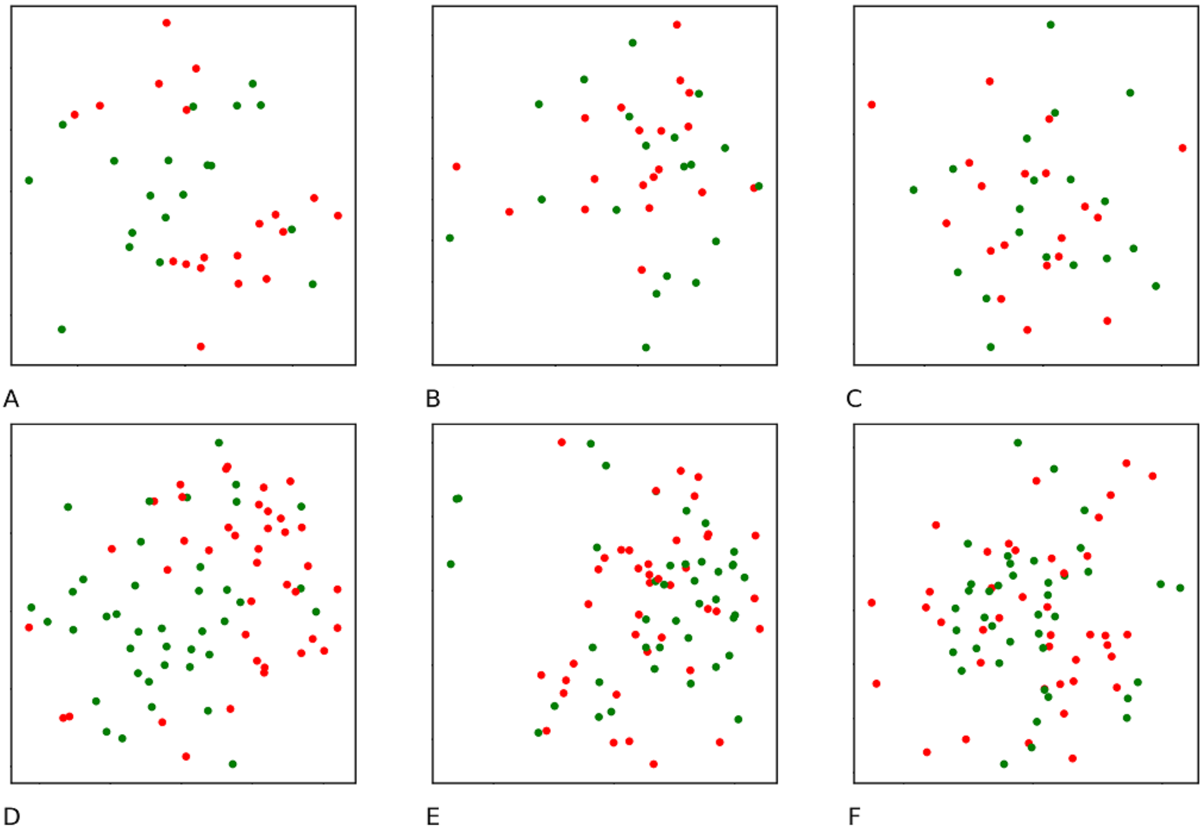
The study was approved by the Human Research Ethics Committees of Xuanwu Hospital, Capital Medical University (approval number: Clinical Scientific Research 2020-070).

Results

Concept Representation Evaluation

Embedding vectors for laboratory test concepts were visualized in a plane space (Figure 3). Concepts of normal and abnormal laboratory tests (Figure 3) were farther away when they were represented by the consequent-based embeddings (Figures 3A and 3D) than by the antecedent-based embeddings (Figures 3B and 3E) and the random selection–based embeddings (Figures 3C and 3F). In cluster analyses for laboratory tests, the consequent-based embeddings achieved higher ARIs (0.317 and 0.520 on the public and private data sets, respectively) than the antecedent-based embeddings (0.112 and 0.149, respectively) and the random selection–based embeddings (0.043 and 0.028, respectively). The best cluster performance of the consequent-based embeddings among the 3 embeddings indicated that the consequent-based embeddings might contain more feature association information.

Figure 3. Visualization of the embedding laboratory tests using different selection schemes for contextual concepts in the skip-gram algorithm (the t-distributed stochastic neighbor embedding algorithm was used). Dots in red and green represent abnormal and normal laboratory test results, respectively. A to C for the public data set: the contextual concepts of a target concept consist of its consequent concepts (A) or antecedent concepts (B) in association rules, or randomly selected concepts (C). D to F are the counterparts of A to C on the private data set.



Predictive Performance

Table 3 and Multimedia Appendix 4 list the predictive performances using various representation methods on the private and public data sets, respectively. The proposed representation method, the weighted sum of the consequent-based embedding representation (CON_EM_WGT), showed the highest predictive performances, with maximum AUROCs of 0.878, 0.973, and 0.926 using all features of the public data set and the entire and treatment-free feature sets of the private data set, respectively. When the performance was measured by AUPRC and F1-score, the proposed representation method outperformed all the other methods regardless of the data sets and feature sets.

Compared with the reference representations, most of the embedding-based representations on both data sets showed a performance improvement. The average AUROC, AUPRC, and F1-score of the 6 representation methods with embeddings were greater than those of the 4 reference methods without embeddings (0.855 vs 0.831, 0.203 vs 0.185, and 0.354 vs 0.328, respectively) on the public data set with the entire feature set. Further, among the 6 representations based on the skip-gram algorithm, representations with algorithm improvement based on the association strength achieved superior performance than those without.

When assembling feature representations into a patient representation, the assembling method and the involved features did matter. Representations based on the idea of weighted sum outperformed those based on the idea of average, on either the public data set with the entire feature set (AUROC, 0.863 to 0.878 vs 0.834 to 0.850) or the private data set with the entire feature set (0.967 to 0.973 vs 0.948 to 0.957). On the other hand, consistently superior predictive performance was achieved on both data sets with the entire feature set compared with the treatment-free feature set. Multimedia Appendix 5 shows the average predictive performance of patient representation methods on the public and private data sets with and without treatment feature sets.

Table 3. Predictive performance of patient representation methods on the private data set.

Feature set and representation methods	AUROC[a], mean (95% CI)	AUPRC[b], mean (95% CI)	F1-score, mean (95% CI)
Entire feature set			
Embedding-based representation methods			
CON_EM_WGT[c]	0.973 (0.951-0.995)	0.505 (0.278-0.732)	0.674 (0.468-0.880)
CON_EM_AVE[d]	0.957 (0.933-0.981)	0.312 (0.159-0.465)	0.479 (0.301-0.657)
ANT_EM_WGT[e]	0.972 (0.948-0.996)	0.489 (0.258-0.720)	0.658 (0.442-0.874)
ANT_EM_AVE[f]	0.953 (0.929-0.977)	0.310 (0.185-0.435)	0.478 (0.329-0.627)
RAN_EM_WGT[g]	0.967 (0.942-0.992)	0.486 (0.263-0.709)	0.660 (0.460-0.860)
RAN_EM_AVE[h]	0.948 (0.923-0.973)	0.287 (0.167-0.407)	0.451 (0.306-0.596)
Reference representation methods			
DIS_AE[i]	0.884 (0.845-0.923)	0.207 (0.144-0.270)	0.361 (0.279-0.443)
DIS_FS[j]	0.938 (0.907-0.969)	0.283 (0.167-0.399)	0.452 (0.309-0.595)
Discretization	0.939 (0.908-0.970)	0.283 (0.165-0.401)	0.454 (0.307-0.601)
Mixture	0.904 (0.849-0.959)	0.251 (0.135-0.367)	0.417 (0.264-0.570)
Treatment-free feature set			
Embedding-based representation methods			
CON_EM_WGT	0.926 (0.883-0.969)	0.282 (0.139-0.425)	0.456 (0.282-0.630)
CON_EM_AVE	0.915 (0.876-0.954)	0.248 (0.156-0.340)	0.413 (0.297-0.529)
ANT_EM_WGT	0.919 (0.874-0.964)	0.278 (0.133-0.423)	0.455 (0.275-0.635)
ANT_EM_AVE	0.912 (0.869-0.955)	0.256 (0.162-0.350)	0.423 (0.307-0.539)
RAN_EM_WGT	0.915 (0.868-0.962)	0.248 (0.119-0.377)	0.416 (0.238-0.594)
RAN_EM_AVE	0.897 (0.850-0.944)	0.225 (0.133-0.317)	0.385 (0.265-0.505)
Reference representation methods			
DIS_AE	0.884 (0.845-0.923)	0.207 (0.144-0.270)	0.361 (0.279-0.443)
DIS_FS	0.903 (0.862-0.944)	0.214 (0.124-0.304)	0.367 (0.236-0.498)
Discretization	0.905 (0.862-0.948)	0.224 (0.122-0.326)	0.381 (0.238-0.524)
Mixture	0.867 (0.806-0.928)	0.202 (0.116-0.288)	0.356 (0.227-0.485)

[a] AUROC: area under the receiver operating characteristic curve.

[b] AUPRC: area under the precision-recall curve.

[c] CON_EM_WGT: weighted sum of the consequent-based embedding representation.

[d] CON_EM_AVE: average of the consequent-based embedding representation.

[e] ANT_EM_WGT: weighted sum of the antecedent-based embedding representation.

[f] ANT_EM_AVE: average of the antecedent-based embedding representation.

[g] RAN_EM_WGT: weighted sum of the random selection–based embedding representation.

[h] RAN_EM_AVE: average of the random selection–based embedding representation.

[i] DIS_AE: discretization representations with features selection.

[j] DIS_FS: hidden vector of an autoencoder-based representation.

Predictive Model Interpretation

Figure 4 illustrates the global feature attributions for the top 20 most important features from the private data set when predicting in-hospital death risk. The treatment-related features played an important role in the mortality prediction. These features included other surgery (mean absolute SHAP value: 0.413), diagnostic ultrasound (0.279), contrast agent cardiovascular angiography (0.197), etc (Figure 4A). Moreover, comorbidity diseases like hypertension (mean absolute SHAP value: 0.252) and heart disease complications (0.236), and laboratory tests like serum glucose (0.188) and serum lactate dehydrogenase (0.139) had strong associations with in-hospital death (Figure 4B). SHAP values of features in the public data set are shown in Multimedia Appendix 6.

In addition to the feature's global importance in the specific predictive task, SHAP values were helpful in distinguishing the feature's local importance, that is, the importance for an individual sample. Figure 5 illustrates how the mortality risk was predicted with SHAP values for a patient who died during hospital stay and another patient who did not die. The positive SHAP values of most features of the patient who died during hospital stay increased the total SHAP value from an average value of −3.739 to a final value of −0.499 (Figures 5A and 5C), meaning that the patient had a higher risk of in-hospital death than the average. In this incremental process, gender as female, for example, contributed a SHAP value of +0.21 (Figures 5C). On the contrary, the negative SHAP values of most features of the patient who was discharged alive decreased the total SHAP value from −3.739 to −6.169 (Figures 5B and 5D), indicating a lower death risk. In this decremental process, male gender contributed a SHAP value of −0.09 (Figures 5D). We have shown 2 examples of patients from the public data set in Multimedia Appendix 7.

Figure 4. The mean absolute Shapley additive explanations (SHAP) values of the top 20 features of the private data set within the entire feature set (A) and the treatment-free feature set (B).

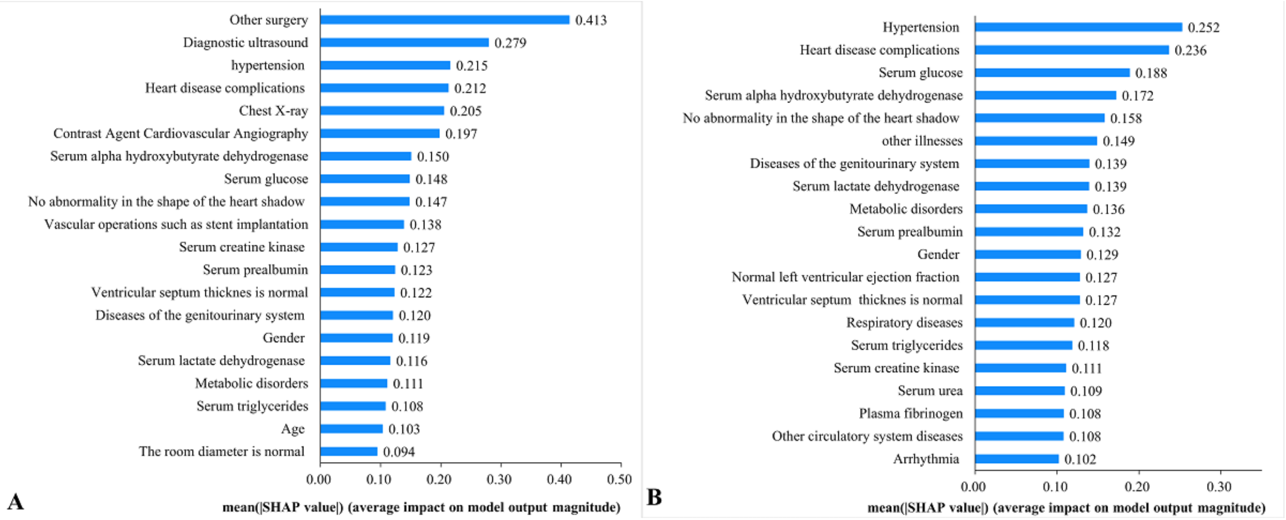
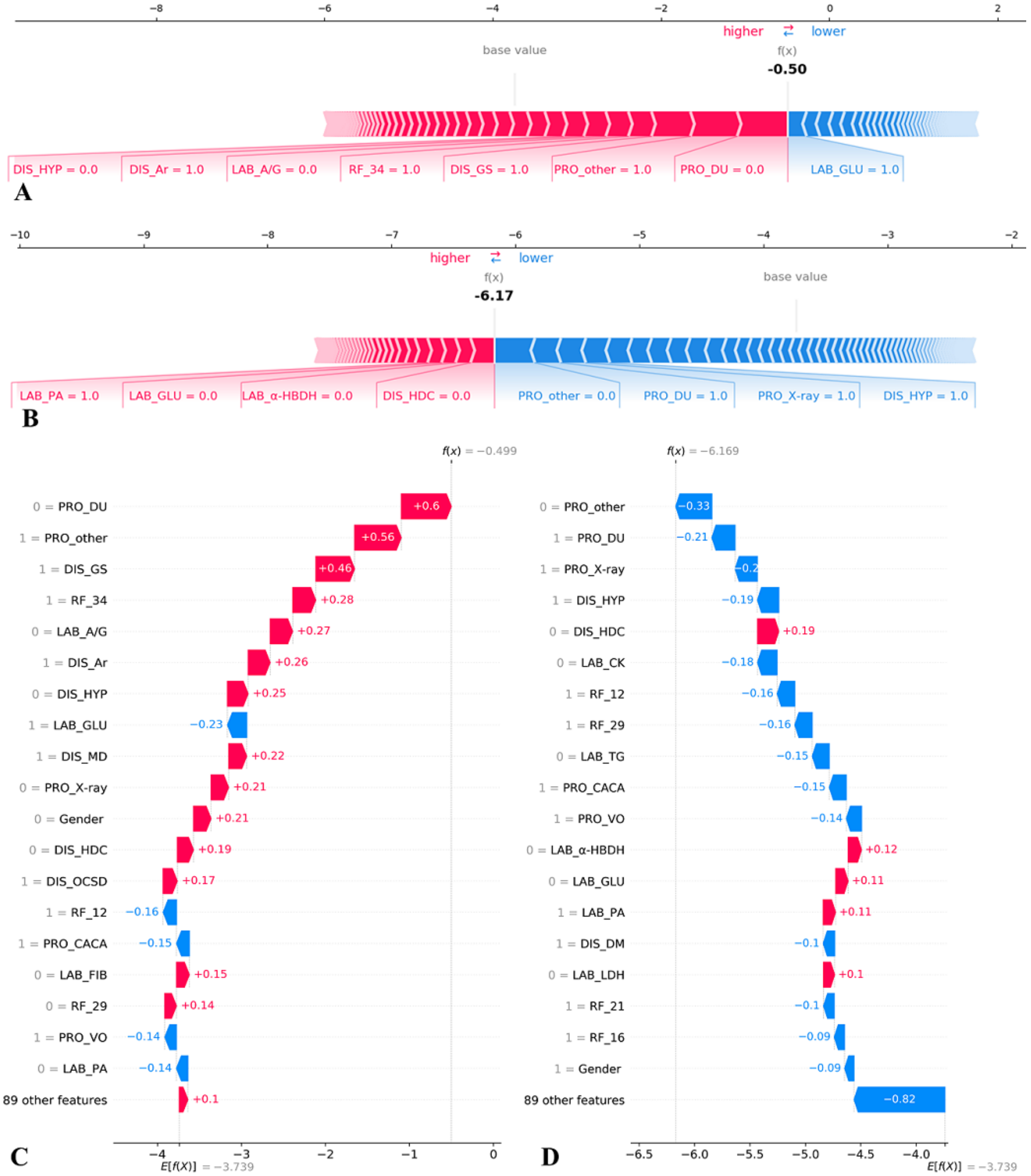


Figure 5. Shapley additive explanations (SHAP) values for a patient who died during hospital stay (A and C) and another patient who did not die (B and D). Both patients were selected from the private data set with the entire feature set. A and B, all features with their SHAP values. C and D, 20 features with the greatest absolute SHAP values. Features in blue tend to reduce the possibility of a patient being classified as positive (death in this study), while features in red do the contrary. The meaning of each abbreviated feature name can be found in Multimedia Appendix 1.



Discussion

Principal Findings

With the widespread adoption of EMR data in building machine learning–based predictive models, one of the most fundamental research challenges was learning proper patient representations that might capture hidden semantic associations among medical concepts [18]. In this study, we proposed an improved skip-gram–based patient representation method where the association strength among medical concepts and the task-specific feature importance were integrated. Compared with other representation methods, the proposed patient representation improved the performance of the mortality risk prediction for AMI patients.

In previous studies, deep learning models [9,10,12,25,36] were used in training embedding representations of medical concepts for the subsequent patient representation. When using the skip-gram algorithm, the order of medical concepts that was independent of feature relevance hindered the algorithm from learning high-quality representation. Prior work recommended the shuffling mechanism for medical concepts in a patient record to reduce the impact of the out-of-order characteristics on the algorithm [12,24-26]. In this study, we introduced the association strength between 2 concepts that was defined as the confidence of an association rule involving just the 2 concepts. Experiments from various aspects indicated that this ingenious improvement was effective in revealing potential associations among medical concepts and further enhancing the performance of downstream prediction tasks.

In addition to the representation algorithm, features used to represent a patient were also critical. Many previous studies focused on some features in the original form of medical codes, such as disease diagnoses, procedures, and medications [1,11,14,37]. For laboratory tests that contained much diagnosis and prognosis-relevant information about patients, we included the normal status of the laboratory tests into the feature sets, rather than simply using the number of laboratory tests and test co-occurrences [12,38]. We further extracted radiological features from free-text radiological reports. Admittedly, richer features may lead to a feature representation with more information, even if the dimension of patient representation remains unchanged. In this study, predictive models using more features to represent a patient did reflect more information about the patient and showed higher performance than those using fewer features. Our finding is similar to the results of other studies [39,40].

Prior studies employed neural networks to train predictive models for clinical outcomes using EMR data [2,16,22]. They focused on end-to-end prediction models built on large data sets, where the last hidden layer of the neural network was regarded as the patient representation. Although the deep end-to-end neural network–based patient representation improved the predictive accuracy, the lack of interpretability could not be ignored. Some studies [26,41] constructed patient representations using the average of concept representations learned by word embedding methods, which did not make full use of the importance of different clinical features for patients. As an advanced interpretability method, the SHAP value [32] was successfully used to analyze and explain the predictive models in some previous studies [40,42-44]. We introduced SHAP values as feature importance into the patient representation, and further explained the predictive model with SHAP values. SHAP values can be used to not only rank the overall importance and identify the important factors for the prediction task, but also explore the key factors for predicting the mortality risk for a specific patient. In our predictive task for AMI patients, the most important features identified by SHAP values were really closely related to AMI [45-47], such as serum glucose and serum creatine kinase, which are 2 critical laboratory tests for AMI diagnosis and prognosis in clinical practice.

In our predictive task, the model that took all available patient characteristics represented by the proposed patient representation method as inputs showed a higher performance than other models on the same task in previous studies (AUROC, 0.973 vs 0.905 to 0.935 [19,29-31,48]). This may be because the embedding representation contained a large number of diverse features extracted from a general EMR system, while many researchers selected AMI-related features with the assistance of clinical experts. For example, basic demographic data and few laboratory tests, as well as several specific features of AMI like Killip classification and left ventricular ejection fraction [19,30] were directly added into the machine learning model to predict mortality risk. Further, compared with other simple feature extraction methods like Principal Component Analysis [29] and the 3-layer autoencoder model, the proposed method took the association strength and feature importance into consideration, achieving higher predictive performance.

Limitations

This study had some limitations. First, only patients' laboratory tests for the first time during hospital stay were included in this study, while many patients took two or more laboratory tests. Since temporal data, especially multiple laboratory tests, may reflect the dynamic health status and the treatment effect of a patient over time, the lack of temporal characteristics of laboratory tests in the patient representation may lead to performance loss in downstream tasks. A future study will focus on integrating this uneven and irregular temporal data into the current patient representation. Second, the skip-gram algorithm was used in training concept embeddings. The algorithm is popular in the natural language processing domain, possibly having a limited ability to represent structured and disordered EMR data. A transformer-based pretrain model, Med-Bert, has been trained to represent disease diagnoses originally expressed in ICD-10 and ICD-9 codes, showing higher performance with AUROCs of 85.39% and 82.23% in heart failure and pancreatic cancer prediction tasks, respectively [49]. Therefore, more complicated deep learning methods will be adopted for a more informative patient representation in the future. Lastly, we carried out only internal validation of the predictive model built on the proposed patient representation. External validation of high quality will be more convincing and will help in continuous algorithm improvement. Moreover, the chosen reference methods for the performance comparison were simple feature selection methods and a 3-layer autoencoder. Comparison with state-of-the-art methods is needed to evaluate the performance and potential use of our proposed method.

Conclusions

In this study, we improved the embedding-based patient representation with the association strength of medical concepts and importance of patient features. After further training and fine-tuning, the model based on the proposed patient representation will hopefully be used to assist in prognostic prediction for AMI inpatients. This study puts forward a meaningful direction for the development of more effective and efficient clinical prediction models using EMR data. It is desirable for patient representation learning to serve as an essential part of building a predictive model for clinical outcomes.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 81971707).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient features of samples in the private data set.
[DOCX File , 35 KB - jmir_v24i8e37486_app1.docx]

Multimedia Appendix 2
Patient features of samples in the public MIMIC-III data set.
[DOCX File , 34 KB - jmir_v24i8e37486_app2.docx]

Multimedia Appendix 3
Predictive performance of skip-gram–based embedding representations with different combinations of the size of the context window and the dimension of the embedding vector.
[DOCX File , 24 KB - jmir_v24i8e37486_app3.docx]

Multimedia Appendix 4
Predictive performance of patient representation methods on the public data set.
[DOCX File , 21 KB - jmir_v24i8e37486_app4.docx]

Multimedia Appendix 5
The average predictive performance of patient representation methods on the public and private data sets with and without treatment feature sets.
[DOCX File , 22 KB - jmir_v24i8e37486_app5.docx]

Multimedia Appendix 6
The mean absolute Shapley additive explanations (SHAP) values of the top 20 features of the public data set within the entire feature set (A) and the treatment-free feature set (B).
[DOCX File , 356 KB - jmir_v24i8e37486_app6.docx]

Multimedia Appendix 7
Shapley additive explanations (SHAP) values for a patient who died during hospital stay (A and C) and another patient who did not die (B and D) from the public data set with the entire feature set.
[DOCX File , 394 KB - jmir_v24i8e37486_app7.docx]

References

1. Xiao X, Wei G, Zhou L, Pan Y, Jing H, Zhao E, et al. Treatment initiation prediction by EHR mapped PPD tensor based convolutional neural networks boosting algorithm. J Biomed Inform 2021 Aug;120:103840. [doi: 10.1016/j.jbi.2021.103840] [Medline: 34139331]
2. Li L, Jiang Y, Huang B. Long-term prediction for temporal propagation of seasonal influenza using Transformer-based model. J Biomed Inform 2021 Oct;122:103894. [doi: 10.1016/j.jbi.2021.103894] [Medline: 34454080]
3. Ma H, Sheng W, Li J, Hou L, Yang J, Cai J, et al. A novel hierarchical machine learning model for hospital-acquired venous thromboembolism risk assessment among multiple-departments. J Biomed Inform 2021 Oct;122:103892 [FREE Full text] [doi: 10.1016/j.jbi.2021.103892] [Medline: 34454079]
4. Tang Z, Yu Y, Ng K, Sow D, Hu J, Mei J. Disease network delineates the disease progression profile of cardiovascular diseases. J Biomed Inform 2021 Mar;115:103686 [FREE Full text] [doi: 10.1016/j.jbi.2021.103686] [Medline: 33493631]
5. Chen P, Dong W, Lu X, Kaymak U, He K, Huang Z. Deep representation learning for individualized treatment effect estimation using electronic health records. J Biomed Inform 2019 Dec;100:103303 [FREE Full text] [doi: 10.1016/j.jbi.2019.103303] [Medline: 31610264]

6. Dligach D, Afshar M, Miller T. Toward a clinical text encoder: pretraining for clinical natural language processing with applications to substance misuse. J Am Med Inform Assoc 2019 Nov 01;26(11):1272-1278 [FREE Full text] [doi: 10.1093/jamia/ocz072] [Medline: 31233140]
7. Kamal SA, Yin C, Qian B, Zhang P. An interpretable risk prediction model for healthcare with pattern attention. BMC Med Inform Decis Mak 2020 Dec 30;20(Suppl 11):307 [FREE Full text] [doi: 10.1186/s12911-020-01331-7] [Medline: 33380322]
8. Zhang X, Chou J, Liang J, Xiao C, Zhao Y, Sarva H, et al. Data-driven subtyping of Parkinson's disease using longitudinal clinical records: A cohort study. Sci Rep 2019 Jan 28;9(1):797 [FREE Full text] [doi: 10.1038/s41598-018-37545-z] [Medline: 30692568]
9. Luo X, Gandhi P, Zhang Z, Shao W, Han Z, Chandrasekaran V, et al. Applying interpretable deep learning models to identify chronic cough patients using EHR data. Comput Methods Programs Biomed 2021 Oct;210:106395. [doi: 10.1016/j.cmpb.2021.106395] [Medline: 34525412]
10. Li Z, Roberts K, Jiang X, Long Q. Distributed learning from multiple EHR databases: Contextual embedding models for medical events. J Biomed Inform 2019 Apr;92:103138 [FREE Full text] [doi: 10.1016/j.jbi.2019.103138] [Medline: 30825539]
11. Barbieri S, Kemp J, Perez-Concha O, Kotwal S, Gallagher M, Ritchie A, et al. Benchmarking deep learning architectures for predicting readmission to the ICU and describing patients-at-risk. Sci Rep 2020 Jan 24;10(1):1111 [FREE Full text] [doi: 10.1038/s41598-020-58053-z] [Medline: 31980704]
12. Steinberg E, Jung K, Fries JA, Corbin CK, Pfohl SR, Shah NH. Language models are an effective representation learning technique for electronic health record data. J Biomed Inform 2021 Jan;113:103637 [FREE Full text] [doi: 10.1016/j.jbi.2020.103637] [Medline: 33290879]
13. Carrasco-Ribelles LA, Pardo-Mas JR, Tortajada S, Sáez C, Valdivieso B, García-Gómez JM. Predicting morbidity by local similarities in multi-scale patient trajectories. J Biomed Inform 2021 Aug;120:103837. [doi: 10.1016/j.jbi.2021.103837] [Medline: 34119690]
14. Yang S, Zheng X, Ji C, Chen X. Multi-layer representation learning and its application to electronic health records. Neural Process Lett 2021;53(2):1417-1433 [FREE Full text] [doi: 10.1007/s11063-021-10449-2] [Medline: 33623481]
15. Wang Z, Wang B, Zhou Y, Li D, Yin Y. Weight-based multiple empirical kernel learning with neighbor discriminant constraint for heart failure mortality prediction. J Biomed Inform 2020 Jan;101:103340 [FREE Full text] [doi: 10.1016/j.jbi.2019.103340] [Medline: 31756495]
16. Rongali S, Rose AJ, McManus DD, Bajracharya AS, Kapoor A, Granillo E, et al. Learning latent space representations to predict patient outcomes: Model development and validation. J Med Internet Res 2020 Mar 23;22(3):e16374 [FREE Full text] [doi: 10.2196/16374] [Medline: 32202503]
17. Tahayori B, Chini-Foroush N, Akhlaghi H. Advanced natural language processing technique to predict patient disposition based on emergency triage notes. Emerg Med Australas 2021;33(3):480-484. [doi: 10.1111/1742-6723.13656] [Medline: 33043570]
18. Si Y, Du J, Li Z, Jiang X, Miller T, Wang F, et al. Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review. J Biomed Inform 2021 Mar;115:103671 [FREE Full text] [doi: 10.1016/j.jbi.2020.103671] [Medline: 33387683]
19. Kwon J, Jeon K, Kim HM, Kim MJ, Lim S, Kim K, et al. Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. PLoS One 2019;14(10):e0224502 [FREE Full text] [doi: 10.1371/journal.pone.0224502] [Medline: 31671144]
20. Ruan T, Lei L, Zhou Y, Zhai J, Zhang L, He P, et al. Representation learning for clinical time series prediction tasks in electronic health records. BMC Med Inform Decis Mak 2019 Dec 17;19(Suppl 8):259 [FREE Full text] [doi: 10.1186/s12911-019-0985-7] [Medline: 31842854]
21. Zhao J, Papapetrou P, Asker L, Boström H. Learning from heterogeneous temporal data in electronic health records. J Biomed Inform 2017 Jan;65:105-119 [FREE Full text] [doi: 10.1016/j.jbi.2016.11.006] [Medline: 27919732]
22. Morid MA, Sheng ORL, Kawamoto K, Abdelrahman S. Learning hidden patterns from patient multivariate time series data using convolutional neural networks: A case study of healthcare cost prediction. J Biomed Inform 2020 Nov;111:103565 [FREE Full text] [doi: 10.1016/j.jbi.2020.103565] [Medline: 32980530]
23. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. 2013 Presented at: 26th International Conference on Neural Information Processing Systems; December 5-10, 2013; Lake Tahoe, Nevada p. 3111-3119. [doi: 10.5555/2999792.2999959]
24. Glicksberg BS, Miotto R, Johnson KW, Shameer K, Li L, Chen R, et al. Automated disease cohort selection using word embeddings from Electronic Health Records. Pac Symp Biocomput 2018;23:145-156 [FREE Full text] [Medline: 29218877]
25. Cui L, Xie X, Shen Z. Prediction task guided representation learning of medical codes in EHR. J Biomed Inform 2018 Aug;84:1-10 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.013] [Medline: 29928997]

26. Huang Y, Wang N, Zhang Z, Liu H, Fei X, Wei L, et al. Patient representation from structured electronic medical records based on embedding technique: Development and validation study. JMIR Med Inform 2021 Jul 23;9(7):e19905 [FREE Full text] [doi: 10.2196/19905] [Medline: 34297000]
27. Chen H, Shi L, Xue M, Wang N, Dong X, Cai Y, et al. Geographic variations in in‐hospital mortality and use of percutaneous coronary intervention following acute myocardial infarction in China: A nationwide cross‐sectional analysis. J Am Heart Assoc 2018 Apr 17;7(8):52-62. [doi: 10.1161/JAHA.117.008131]
28. GBD 2013 Mortality Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet 2015 Jan 10;385(9963):117-171 [FREE Full text] [doi: 10.1016/S0140-6736(14)61682-2] [Medline: 25530442]
29. Lee HC, Park JS, Choe JC, Ahn JH, Lee HW, Oh J, Korea Acute Myocardial Infarction Registry (KAMIR) Korea Working Group on Myocardial Infarction (KorMI) Investigators. Prediction of 1-year mortality from acute myocardial infarction using machine learning. Am J Cardiol 2020 Oct 15;133:23-31. [doi: 10.1016/j.amjcard.2020.07.048] [Medline: 32811651]
30. Aziz F, Malek S, Ibrahim KS, Raja Shariff RE, Wan Ahmad WA, Ali RM, et al. Short- and long-term mortality prediction after an acute ST-elevation myocardial infarction (STEMI) in Asians: A machine learning approach. PLoS One 2021;16(8):e0254894 [FREE Full text] [doi: 10.1371/journal.pone.0254894] [Medline: 34339432]
31. Wang Q, Qian W, Sun Z, Zhu W, Liu Y, Chen X, et al. Nomograms based on pre-operative parametric for prediction of short-term mortality in acute myocardial infarction patients treated invasively. Aging (Albany NY) 2020 Dec 11;13(2):2184-2197 [FREE Full text] [doi: 10.18632/aging.202230] [Medline: 33323557]
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, California p. 4768-4777. [doi: 10.5555/3295222]
33. Medical Information Mart for Intensive Care. URL: https://mimic.mit.edu/ [accessed 2022-07-23]
34. Wang N, Wang M, Zhou Y, Liu H, Wei L, Fei X, et al. Sequential data-based patient similarity framework for patient outcome prediction: Algorithm development. J Med Internet Res 2022 Jan 06;24(1):e30720 [FREE Full text] [doi: 10.2196/30720] [Medline: 34989682]
35. Xie J, Gao H, Xie W, Liu X, Grant PW. Robust clustering by detecting density peaks and assigning points based on fuzzy weighted K-nearest neighbors. Information Sciences 2016 Aug;354:19-40. [doi: 10.1016/j.ins.2016.03.011]
36. Wang L, Wang Q, Bai H, Liu C, Liu W, Zhang Y, et al. EHR2Vec: Representation learning of medical concepts from temporal patterns of clinical notes based on self-attention mechanism. Front Genet 2020;11:630 [FREE Full text] [doi: 10.3389/fgene.2020.00630] [Medline: 32714371]
37. Bai T, Chanda AK, Egleston BL, Vucetic S. EHR phenotyping via jointly embedding medical concepts and words into a unified vector space. BMC Med Inform Decis Mak 2018 Dec 12;18(Suppl 4):123 [FREE Full text] [doi: 10.1186/s12911-018-0672-0] [Medline: 30537974]
38. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 May 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
39. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. BMC Med Inform Decis Mak 2020 Oct 29;20(1):280 [FREE Full text] [doi: 10.1186/s12911-020-01297-6] [Medline: 33121479]
40. Xu Y, Liu X, Pan L, Mao X, Liang H, Wang G, et al. Explainable dynamic multimodal variational autoencoder for the prediction of patients with suspected central precocious puberty. IEEE J. Biomed. Health Inform 2022 Mar;26(3):1362-1373. [doi: 10.1109/jbhi.2021.3103271]
41. Choi E, Schuetz A, Stewart WF, Sun J. Medical concept representation learning from electronic health records and its application on heart failure prediction. arXiv. 2016. URL: https://arxiv.org/abs/1602.03686 [accessed 2022-07-19]
42. Müller M, Gromicho M, de Carvalho M, Madeira SC. Explainable models of disease progression in ALS: Learning from longitudinal clinical data with recurrent neural networks and deep model explanation. Computer Methods and Programs in Biomedicine Update 2021;1:100018. [doi: 10.1016/j.cmpbup.2021.100018]
43. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
44. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018 Oct;2(10):749-760. [doi: 10.1038/s41551-018-0304-0] [Medline: 31001455]
45. Pinto DS, Grandin EW. Risk prediction in AMI shock: Goldilocks and the search for "just right". J Am Coll Cardiol 2017 Apr 18;69(15):1921-1923 [FREE Full text] [doi: 10.1016/j.jacc.2017.02.024] [Medline: 28408021]
46. Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV, Go AS. Population trends in the incidence and outcomes of acute myocardial infarction. N Engl J Med 2010 Jun 10;362(23):2155-2165. [doi: 10.1056/nejmoa0908610]
47. Shroff GR, Frederick PD, Herzog CA. Renal failure and acute myocardial infarction: clinical characteristics in patients with advanced chronic kidney disease, on dialysis, and without chronic kidney disease. A collaborative project of the United

States Renal Data System/National Institutes of Health and the National Registry of Myocardial Infarction. Am Heart J 2012 Mar;163(3):399-406 [FREE Full text] [doi: 10.1016/j.ahj.2011.12.002] [Medline: 22424010]

48. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu M, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. The Lancet 2021 Jan;397(10270):199-207. [doi: 10.1016/S0140-6736(20)32519-8]
49. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit Med 2021 May 20;4(1):86 [FREE Full text] [doi: 10.1038/s41746-021-00455-y] [Medline: 34017034]

Abbreviations

AMI: acute myocardial infarction
ARI: adjusted Rand index
AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
EMR: electronic medical record
ICD: International Classification of Diseases
SHAP: Shapley additive explanations

Edited by G Eysenbach; submitted 22.02.22; peer-reviewed by M Nuutinen, L Jorm, B Qian; comments to author 20.04.22; revised version received 02.06.22; accepted 18.07.22; published 03.08.22.

Please cite as:
Huang Y, Zheng Z, Ma M, Xin X, Liu H, Fei X, Wei L, Chen H
Improving the Performance of Outcome Prediction for Inpatients With Acute Myocardial Infarction Based on Embedding Representation Learned From Electronic Medical Records: Development and Validation Study
J Med Internet Res 2022;24(8):e37486
URL: https://www.jmir.org/2022/8/e37486
doi:10.2196/37486
PMID:35921141

Original Paper

Factors Influencing How Providers Assess the Appropriateness of Video Visits: Interview Study With Primary and Specialty Health Care Providers

Caroline Gray[1], BA, MA, PhD; Charlie Wray[2], MS, DO; Rebecca Tisdale[1], MPH, MD; Camila Chaudary[1], MPH; Cindie Slightam[1], MPH; Donna Zulman[1], MS, MD

[1]VA Palo Alto Healthcare System, Menlo Park, CA, United States

[2]Division of Hospital Medicine, San Francisco Veterans Affairs Medical Center, San Francisco, CA, United States

Corresponding Author:
Caroline Gray, BA, MA, PhD
VA Palo Alto Healthcare System
795 Willow Road
Menlo Park, CA, 94025
United States
Phone: 1 6502486147
Email: cpgray@gmail.com

Abstract

Background: The rapid implementation of virtual care (ie, telephone or video-based clinic appointments) during the COVID-19 pandemic resulted in many providers offering virtual care with little or no formal training and without clinical guidelines and tools to assist with decision-making. As new guidelines for virtual care provision take shape, it is critical that they are informed by an in-depth understanding of how providers make decisions about virtual care in their clinical practices.

Objective: In this paper, we sought to identify the most salient factors that influence how providers decide when to offer patients video appointments instead of or in conjunction with in-person care.

Methods: We conducted semistructured interviews with 28 purposefully selected primary and specialty health care providers from the US Department of Veteran's Affairs health care system. We used an inductive approach to identify factors that impact provider decision-making.

Results: Qualitative analysis revealed distinct clinical, patient, and provider factors that influence provider decisions to initiate or continue with virtual visits. Clinical factors include patient acuity, the need for additional tests or labs, changes in patients' health status, and whether the patient is new or has no recent visit. Patient factors include patients' ability to articulate symptoms or needs, availability and accessibility of technology, preferences for or against virtual visits, and access to caregiver assistance. Provider factors include provider comfort with and acceptance of virtual technology as well as virtual physical exam skills and training.

Conclusions: Providers within the US Department of Veterans Affairs health administration system consider a complex set of factors when deciding whether to offer or continue a video or telephone visit. These factors can inform the development and further refinement of decision tools, guides, and other policies to ensure that virtual care expands access to high-quality care.

(J Med Internet Res 2022;24(8):e38826) doi:10.2196/38826

KEYWORDS

virtual care; decision-making; qualitative; virtual visits; web-based; carer; video; telephone; telemedicine; appointments; caregiver

Introduction

The COVID-19 pandemic spurred rapid and widespread implementation of virtual care, including both video- and telephone-based visits, to address acute and chronic needs of patients. Many welcomed the availability of virtual care given benefits such as increased access to care, less travel time for patients, and often lower costs for both patients and health care systems [1]. Providers and patients have also reported unexpected advantages such as greater convenience and the ability to assess patients in their home environments [2].

While in-person care delivery has largely resumed, virtual care continues to play a major role in how health care systems deliver care to patients [3]. Because of this, identifying optimal approaches to virtual care delivery that ensure patient safety, satisfaction, quality of care, and equitable access will remain a critical challenge facing health care organizations [4,5]. Due to the rapid implementation of virtual care during the pandemic, many providers were asked to provide virtual care with little or no formal training and without clinical guidelines and tools to assist with decision-making. Currently, these guidelines and tools are beginning to take shape, and it is critical that they are informed by an in-depth understanding of *how* providers make decisions about virtual care in their clinical practices.

Currently, there is a plethora of qualitative studies describing provider and patient attitudes toward virtual care, as well as perceived barriers and facilitators to virtual care implementation and adoption [2,6-8]. Well-known barriers include lack of institutional support and the infrastructure to support the technology needed for virtual care, low levels of digital literacy among both patients and providers, and poor integration of virtual modalities into existing clinical workflows, to name a few [2,8,9]. Though these studies may improve uptake and implementation of virtual care, they sometimes lack specifics on how and when to provide virtual care as a substitute or adjunct to in-person care. Fewer studies have examined how non–mental health care providers more generally make decisions about when to use virtual modalities [10,11]. Systematic identification of the factors that providers consider when assessing the suitability of virtual care for a given patient and clinical need may inform the aforementioned tools and guidelines necessary for accessible, high-quality care.

The US Department of Veterans Affairs (VA) has been a leading health care organization in the use of virtual care modalities, even before the onset of the COVID-19 pandemic [12,13]. Through a qualitative assessment of VA providers from diverse clinical settings, we sought to identify the most salient factors that impact providers' decisions about when to offer patients virtual care.

Methods

Participants and Study Design

As part of a study of VA's implementation of virtual care and distribution of video-enabled tablets to veterans with access barriers, we conducted a qualitative study of a national sample of VA clinicians. We used a combination of administrative data and provider referral to purposefully sample participants. A majority of providers had only begun offering virtual care in the last 1 to 2 years at the time of interviews. Rather than focus on a single specialty, we strove for variation in our interview sample [14] and selected providers from the following 4 diverse areas of clinical practice: primary care, cardiology, spinal cord injury, and palliative care. With guidance from VA's Office of Connected Care, these specialties were chosen due to higher use of virtual care. Additionally, they offer diverse types of services offered during a clinical visit. Through administrative data, we identified practitioners who had frequently used video visits in the previous calendar year relative to their peers in similar clinical practices. Providers were recruited from several US geographic regions—West, Midwest, South, and Northeast. The participants were sent messages to their institutional email addresses explaining the purpose of the study and asked if they would like to participate. A total of 26 physicians and 2 nurse practitioners agreed to participate in an interview, for a total of 28 providers. Participant characteristics are summarized in Table 1.

Table 1. Characteristics of health care providers (N=28).

Characteristics	Values, n (%)
Clinical specialty or practice	
Primary care	11 (39)
Cardiology	7 (25)
Spinal cord injury	5 (18)
Palliative care	5 (18)
Gender	
Women	16 (57)
Men	12 (43)
Setting	
Rural	5 (18)
Urban or suburban	23 (82)
Years of practice	
Less than 5 years	4 (14)
Over 5 years	24 (86)

Ethics Approval

This quality improvement initiative was reviewed and designated as nonresearch by the supporting VA program office, the Stanford University Institutional Review Board, and VA Research Administration.

Data Collection

Two researchers, a medical sociologist with expertise in qualitative methods (CG) and an internist with qualitative training (RT), conducted interviews with the providers to learn about their experiences offering virtual care prior to and during the COVID-19 pandemic. The interviews took place between December 2020 and June 2021, with each specialty interviewed consecutively to ensure greater consistency. The interview questions focused on circumstances under which providers choose to offer virtual care, preferences for virtual care or in-person care, and perceptions of scenarios where virtual care was inappropriate or less optimal. The providers were also asked to reflect on needed skills and training around virtual care and perceived barriers to providing virtual care more frequently. The interviews were conducted using Microsoft Teams, which lasted approximately 30 minutes, and with permission from interview participants, they were videorecorded and transcribed by a professional transcription service.

Data Analysis

To identify the primary factors informing provider decision-making around virtual care, we employed a qualitative descriptive approach [15], using constant comparison [16,17] to further reduce and synthesize data. First, the research team inductively reviewed 5 transcripts and identified emergent codes, combined these codes with deductive codes derived from the interview questions, and created a codebook used to code all transcripts. The transcripts were uploaded into Atlas.ti (ATLAS.ti Scientific Software Development GmbH), a software that facilitates qualitative data analysis, and coded according to the codebook. After transcripts were coded, the codes and their associated text were reviewed collectively by the team and then grouped together into larger categories. During this process, we identified themes by assessing for repetition and emphasis of specific points. Finally, all team members participated in selecting exemplary quotes and sorting themes into 3 categories of factors that appeared to most impact provider decision-making.

Results

Clinical, Provider, and Patient Factors Impacting the Decision to Use Virtual Care

Thematic analysis revealed that provider decisions about whether to continue with or initiate a virtual visit is driven by clinical, patient, and provider factors (Textbox 1). Although we observed some variation related to specific aspects of the different clinical focus areas, the factors discussed here were noted across all 4 specialties.

Textbox 1. Thematic categorization of factors influencing provider decision-making.

Clinical factors

- Patient acuity
- Need for additional tests or labs
- Change in patient’s status or overall stability
- First visits and patients with no record of recent medical examination

Patient factors

- Patient’s ability to articulate symptoms or needs
- Availability and accessibility of technology
- Preferences regarding virtual visits
- Access to caregiver assistance

Provider factors

- Comfort with and acceptance of virtual technology
- Knowledge about how to conduct physical exam and assessment virtually

Clinical Factors Impacting the Decision to Use Virtual Care

The providers described clinical factors that impact their decisions about whether to see a patient virtually or in person. The common clinical factors cited include patient acuity, a need for additional tests or labs, changes in the patient’s status and overall stability, and a visit with a patient who is new or has no record of recent medical examination.

Patient Acuity

The providers indicated that acute, newly emergent conditions proved most difficult to assess virtually. In particular, they noted that patients’ reports of pain were often challenging to assess virtually, since they were unable to physically examine sensitive areas to help in making a diagnosis. On the other hand, chronic conditions were better suited for virtual management, particularly if patients had already been diagnosed and had an established medical plan. Blood pressure and blood sugar management were characterized as 2 examples of chronic conditions that may be easily managed using virtual care. Elaborating on this observation, a spinal cord provider explained as follows:

> *Most of the time, you cannot [make a diagnosis] without laying hands on the patient. But [when it's] just blood pressure management, blood sugar management, you don't have to have a patient face-to-face encounter. You can do only virtual.*

Need for Additional Tests or Labs

Conditions that required lab draws or imaging to accurately diagnose were described as difficult to manage. The providers noted that patients who were able to have tests performed prior to their virtual visit were much more likely to have a productive visit, but because tests are often completed at the time of the in-person visit, previsit workups were reportedly uncommon. Hence, the providers noted that if a patient needed lab tests, additional virtual visits were often necessary to complete their assessments and ultimately make a diagnosis.

Changes in Health Status

The providers noted that patients who reported changes in clinical status and overall well-being were less appropriate for virtual care. These changes often signaled to providers the need for a comprehensive, in-person physical examination rather than a virtual exam. Some examples of health status changes that clinicians felt warranted an in-person visit included unexpected weight gain or weight loss and fluctuating or inconsistent symptoms accompanying a diagnosed chronic condition. For example, a cardiologist noted the following:

> *[If] I have a visit with a patient that's either a telephone or a [video visit] and identify that there are some factors that are starting to concern me——in general it's weight, shortness of breath, new symptoms that I wish I could have a physical exam or be able to examine the patient—then I will follow those telephone visits … with an in-person visit generally in the next couple of weeks and sometimes more urgently.*

Conversely, the providers indicated that patients who reported a stable and consistent health status made for better candidates for virtual care.

New Patients and Individuals With No Recent Visit

Providers across all specialties maintained the view that first visits and new patients should be seen in person if possible. This view held steady despite the wide variety of conditions being assessed and treated among the providers who participated in the interviews. For example, a physician who treats patients with spinal cord injury stated the following:

> *In terms of pain, you have to have at least the first encounter in person, because you have to do a special test, you have to examine to see specificity, to palpate, to see joints, range of motion.*

Additionally, patients who had not been seen in person for an extended period (2 or more years) were considered less ideal virtual care candidates. However, providers noted that they felt more comfortable offering virtual care when the patients had been recently seen by other providers within the medical system and for whom extensive notes were available.

Patient Factors Impacting the Decision to Use Virtual Care

While the providers largely focused on the clinical needs and circumstances of patients when determining whether a virtual visit would be appropriate, they also described several patient-related factors that influenced decision-making, including a patient's ability to articulate their symptoms and needs, ability to use the technology associated with virtual visits, general preferences for in-person visits, and access to a caregiver to assist with the virtual visit.

Ability to Articulate Symptoms or Needs

The providers explained that patients who were able to communicate their symptoms or needs in a robust and reliable way made for the best virtual visit candidates. Patients who had challenges describing their symptoms, difficulty recalling the timing of certain events or the onset of specific symptoms, or challenges describing physical changes or abnormalities left providers less confident in their virtual assessments. For instance, a primary care provider described how she imparts this advice to residents:

> *I tell the residents as we're seeing patients, one of our first decisions to make is, "Can I safely continue this visit in this fashion, or is there no way I'm going to get enough data by history that I can end at a point where I feel like I've safely cared for the patient?"*

The providers admitted that relying on patients' accounts rather than their own hands-on assessments required a comfort level with virtual assessments, which often took time to develop. In response to this, the providers noted that they had to hone their history taking skills to feel confident with the information patients were relaying to them.

Availability and Accessibility of Technology

The providers indicated that patients needed both personal technology (eg, home computer, tablet, or smart phone) and reliable broadband access to participate in virtual visits. They described many instances of initiating virtual encounters with patients, only to discover that the video or sound quality was poor, and subsequently wasted valuable clinical time troubleshooting these technology-based problems with patients. In such cases, they would either try to follow up by telephone or simply reschedule in-person visits. The providers also noted that individuals with specific clinical or physical characteristics frequently had challenges with virtual visits (eg, older patients with cognitive disabilities such as dementia or patients who experienced sensory loss, namely hearing and visual impairments). A quote from a primary care provider illustrates this point:

> *Like hearing can become a huge problem. If hearing difficulties are too severe, it's really hard to have an appropriate visit. There's something with the tech, the video that I feel like people just can't hear you as*

> *well. I'm not sure if it's the delay and it throws off the mouth reading or something.*

However, despite these challenges, the providers cautioned against assuming that all older patients or patients with sensory loss were inappropriate for virtual care, since they could think of many exceptions to this general observation.

Access to Caregiver Assistance

Finally, the providers noted that having another individual available to assist the patient, typically a caregiver or family member, increased their likelihood of conducting a virtual visit. Particularly among patients with mobility issues, sensory loss, or cognitive impairment, a caregiver was often able to help the veteran troubleshoot technology issues, assist with physical exam maneuvers, or help capture images providers needed to fully assess the patient. Underscoring this point, a palliative care provider explained how they suggest involving caregivers in assessing pain in areas that may be difficult for patients to reach:

> *I always ask the patient "Does it hurt to touch?" And if there's a spouse or another person or a family member or any other person there, I might ask them to touch it.*

Preferences Regarding Virtual Visits

The providers reported that some patients preferred in-person visits to virtual visits and were therefore reluctant to engage through virtual care if an in-person visit could be conducted in a timely manner. They noted that some patients crave face-to-face interactions with their providers and report that the video format fails to replicate that connection. For others, this preference was also attributed to a lack of digital literacy skills and inadequate patient support to help facilitate their use. With additional instruction and digital familiarity, some of these patients could grow more accepting of virtual care.

Nevertheless, the providers speculated that patient's preferences were unlikely to change and that they would continue to opt for in-person visits when given the choice. For instance, a primary care provider reflected as such:

> *My perception of my patients is they're not entirely comfortable never seeing me in person, especially new patients who I've never met. I think most of them feel like, "I'd like to meet you at some point." I think that's always going to be a need there.*

In these scenarios, the providers noted that they would often comply with patients' preferences and opt to see the patient in person rather than virtually. The COVID-19 pandemic, however, necessitated at times that visits be virtual, even when patients preferred in-person care.

Provider Factors Impacting the Decision to Use Virtual Care

While playing less of a role in real-time decision-making around virtual care, the providers also described how factors related to their own acceptance of and comfort with virtual care modalities impacted their decision-making. In addition, they noted that acquiring training on how to assess patients virtually would likely lessen their discomfort and encourage them to provide virtual care more often and to more patients with diverse clinical needs.

Comfort With and Acceptance of Virtual Technology

First, the providers argued that assessing patients virtually required a general acceptance of the format and a recognition that it necessitates a different approach to patient assessment and evaluation. While the providers in our sample largely appreciated virtual care, they described colleagues who lamented the shift to virtual care and found it challenging to adapt their clinical care to the new format. This acceptance provided a foundation for providers to improve their virtual diagnostic skills and increase the likelihood of engaging in a virtual visit with a patient. A primary care provider elaborated on this point as follows:

> *You have to accept the strengths and deficits of video [visits] and don't try to make it into a total replacement for a face-to-face visit, because if you're more comfortable listening to their symptoms, listening to what they tell you and they can relate to you pretty well how much edema they have and where it was before, and if you accept that, then you can get more done.*

Virtual Physical Exam Skills

In most instances, the providers described learning to provide virtual care as a process of "just figuring it out," while also drawing on the fundamentals of their clinical training. In this process, many acquired new skills and adopted new strategies for conducting virtual physical exams, including asking patients to engage in specific maneuvers or provide information not typically asked for in a face-to-face visit. The providers described how conducting virtual exams increased their awareness of the observations they make about patients and their physical health during in-person exams. Virtual exams required deliberate attention to those missing elements. For example, one primary care provider explained:

> *You're assessing the speed they're getting up and moving around, so you have to make sure to ask them to walk around. And so, I think that there is a potential to miss things if you haven't gone through the process of saying, okay, what are the things that I'm likely to miss as a provider given this particular modality, and then how can I try to counter those with just some things on your internal checklist that you want to make sure to ask about?*

In this last example, asking patients to stand and walk around while on video was one way to assess gait and movement. The providers' confidence in and acquisition of these skills increased the likelihood that they would opt to treat and assess a patient virtually.

Discussion

Principal Findings

In this qualitative study of VA providers, we found that a complex set of clinical, patient, and provider factors influences a provider's decision about whether to provide care virtually or in person. Many of the providers in our study referenced scenarios where virtual visits had already been scheduled and initiated, but through examination of patients, they realized an in-person visit would be more clinically appropriate. Such instances added an additional visit for both patients and providers, contributing to potential waste and redundant services. This highlights the value for providers of knowing a priori which scenarios and which patients might be more appropriate for virtual care. Here, we detail several ways that these findings may be used to optimize the use of both virtual and in-person care.

First, the providers noted that, in many cases, a high-quality virtual visit requires some collection of information or data from the patient. As the providers have made the leap to virtual care, many have mourned the loss of data that would be more easily accessible in a traditional in-person visit, such as vital signs and physical exam findings [18]. Some providers have found solutions in home devices such as blood pressure cuffs, blood glucose monitors, pulse oximeters, and scales, all of which can help them to form a more complete picture of patients' vital signs and other important information for decision-making [19,20]. The increasing availability of wearable and other patient-facing digital technologies, including exercise monitors (eg, FitBits and Apple Watches), smartphone-associated portable electrocardiograms, and home-based lab testing may offer additional opportunities to collect key information outside the in-person visit, although there is still a need for evidence about the reliability and consistency of data in different circumstances [21,22]. Augmenting a virtual visit with these technologies may mitigate the risks that the providers in our study noted when they must rely on imperfect or incomplete patient-provided histories.

Second, specific skills and training are required to conduct effective virtual visits and spare providers from "figuring it out as [they] go." Several efforts are underway to develop and disseminate training and instruction on virtual care and to integrate these domains into standard medical school and residency curricula [7,23]. Additional training resources should target mid- or later-career clinicians, since they are less likely to be exposed to interventions geared toward medical trainees. The American Medical Association as well as other organizations and societies have developed resources to help clinicians build telemedicine physical exam skills [24] and communication skills such as "digital empathy" [25]. These resources include guidebooks as well as informational webinar series and videos. Others have created helpful guides for conducting patient-assisted physical exams [26].

Third, there is a need for guidelines to help determine whether a specific visit should be scheduled in person or virtually. The providers largely maintained that patients without recent visits or presenting with new or higher acuity problems might not be best served by virtual visits. Both for building patient rapport and for ensuring a more complete mental model of a patient's condition, it may well be best for an initial patient visit to occur in person for most patients and clinical situations [25]; some have suggested newly diagnosed patients should always be seen in person, at least initially, until medication regimens can be safely established [27]. Nevertheless, there is mounting evidence that with proper training and protocols in place, even high-acuity clinical circumstances can be safely assessed virtually and may even decrease overall rates of emergency care use. For example, Wray et al [28] demonstrated how a "tele-urgent care program" that provided care for a variety of clinical scenarios was safe, effective, and led to the decreased use of emergency departments. Such findings may provide further confidence to virtual care providers that virtual care can provide safe access to care in a variety of clinical scenarios.

Finally, attention to equity is needed to ensure that all patients have opportunities to build digital literacy skills [29], have access to the technology and receive the support they need to participate in virtual visits. Failure to attend to these issues may contribute to further inequity in health care provision and outcomes [30]. Health care organizations have attempted to respond to this digital divide in a variety of ways. For instance, the VA initiated a tablet distribution program, in which at-risk, high-need patients are provided with video-enabled tablets equipped with internet service. This program has resulted in improved access and continuity of care, with high satisfaction rates among Veterans [31,32]. In addition, studies have found that providing patients with hands-on instruction on how to use new technologies may further ameliorate a lack of digital literacy skills [33]. Moreover, functional limitations (eg, loss of eyesight and hearing as well as dementia) also created barriers to patients' use of virtual care in this study. Incorporating principles of universal design, which advocates for designing products and services that can be used by all individuals to the greatest extent possible, may abet some of these issues and ensure accessibility for all patients [30,34,35].

Even with the array of tools and strategies described above, it is unlikely that all combinations of providers, patients, and clinical scenarios will ultimately prove ideal for virtual care. Given the dramatic expansion of virtual care since the onset of the COVID-19 pandemic, the existence of virtual care is effectively a foregone conclusion; what is essential to uncover at this stage is how and when to best use the various visit modalities at provider and patients' disposal. The observations providers shared in this study are useful for generating hypotheses on how to integrate virtual and in-person care.

Limitations

This study has several limitations. First, the sample was limited to providers at the VA; thus, provider experiences may not apply to other settings, particularly those with different reimbursement models. Fee-for-service systems, the predominant mode of health care delivery in the United States, may reimburse virtual visits differently from in-person visits and pose additional incentives or disincentives to use virtual care. However, we were able to assess a broad array of providers across a variety of geographic regions, improving the transferability of our findings. Second, this qualitative study about provider perceptions does not assess the impact of these factors on quality of care and patient outcomes, which would provide value in discussions about the degree to which these criteria should inform guidelines and protocols. For example, while the providers noted that remote monitoring devices and other technology increased their comfort and confidence in virtual examination, the actual impact of this factor on quality and safety of care warrants further evaluation. A final and perhaps most significant limitation is that this study does not assess decision-making in real time and instead relies on providers' reflections on their decisions, an inherent limitation of qualitative interviews. Though challenging to carry out, direct observations of clinical practices may offer a more realistic account of provider decision-making around virtual care.

Conclusion

This qualitative study found that providers within the VA consider a complex array of factors when deciding whether to offer or continue with a virtual visit. Clinical factors were the most dominant, but patient and provider factors also influenced the decision process. These findings can inform health system policies to ensure accessible, high-quality care, as well as policy maker considerations when adjudicating reimbursement levels for virtual care visits. Further development of tools, resources, and guidelines is needed to facilitate real-time provider decision-making about when to offer a patient virtual care.

Acknowledgments

Office of Rural Health Enterprise Wide Initiative and Quality Enhancement Research Initiative (Award Number PEC 18-205; principal investigator: DZ) from the US Department of Veterans Affairs Health Services Research and Development Service of the VA Office of Research and Development.

Disclaimer

The views expressed are those of the authors, and the contents do not represent the views of the US Department of Veterans Affairs or the US Government.

Conflicts of Interest

None declared.

References

1. Mehrotra A, Ray K, Brockmeyer D, Barnett M, Bender J. Rapidly converting to "virtual practices": outpatient care in the era of Covid-19. NEJM Catal Innov Care Deliv 2020:1-5. [doi: 10.1056/CAT.20.0091]
2. Gomez T, Anaya YB, Shih KJ, Tarn DM. A Qualitative Study of Primary Care Physicians' Experiences With Telemedicine During COVID-19. J Am Board Fam Med 2021 Feb 23;34(Supplement):S61-S70. [doi: 10.3122/jabfm.2021.s1.200517]
3. Mann D, Chen J, Chunara R, Testa P, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
4. Herzer KR, Pronovost PJ. Ensuring Quality in the Era of Virtual Care. JAMA 2021 Feb 02;325(5):429-430. [doi: 10.1001/jama.2020.24955] [Medline: 33528544]
5. Zulman DM, Verghese A. Virtual Care, Telemedicine Visits, and Real Connection in the Era of COVID-19: Unforeseen Opportunity in the Face of Adversity. JAMA 2021 Feb 02;325(5):437-438. [doi: 10.1001/jama.2020.27304] [Medline: 33528520]
6. Almathami HKY, 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]
7. Alkureishi MA, Choo Z, Lenti G, Castaneda J, Zhu M, Nunes K, et al. Clinician Perspectives on Telemedicine: Observational Cross-sectional Study. JMIR Hum Factors 2021 Jul 09;8(3):e29690 [FREE Full text] [doi: 10.2196/29690] [Medline: 34184994]
8. Anthony Jnr B. Implications of telehealth and digital care solutions during COVID-19 pandemic: a qualitative literature review. Inform Health Soc Care 2021 Mar 02;46(1):68-83. [doi: 10.1080/17538157.2020.1839467] [Medline: 33251894]
9. Wong A, Bhyat R, Srivastava S, Boissé Lomax L, Appireddy R. Patient Care During the COVID-19 Pandemic: Use of Virtual Care. J Med Internet Res 2021 Jan 21;23(1):e20621 [FREE Full text] [doi: 10.2196/20621] [Medline: 33326410]
10. Segal JB, Dukhanin V, Davis S. Telemedicine in Primary Care: Qualitative Work Towards a Framework for Appropriate Use. J Am Board Fam Med 2022 May 31;35(3):507-516 [FREE Full text] [doi: 10.3122/jabfm.2022.03.210229] [Medline: 35641038]
11. Segal JB, Davis S, Dukhanin V. Working Framework for Appropriate Use of Virtual Care in Primary Care. J Am Board Fam Med 2022 May 31;35(3):629-633 [FREE Full text] [doi: 10.3122/jabfm.2022.03.210469] [Medline: 35641060]
12. Ferguson J, Jacobs J, Yefimova M, Greene L, Heyworth L, Zulman D. Virtual care expansion in the Veterans Health Administration during the COVID-19 pandemic: clinical services and patient characteristics associated with utilization. J Am Med Inform Assoc 2021 Mar 01;28(3):453-462 [FREE Full text] [doi: 10.1093/jamia/ocaa284] [Medline: 33125032]

13. Heyworth L, Kirsh S, Zulman D, Ferguson J, Kizer K. Expanding access through virtual care: The VA's early experience with Covid-19. NEJM Catal Innov Care Deliv 2020 Jun 01:1-11.
14. Sandelowski M. Sample size in qualitative research. Res Nurs Health 1995 Apr;18(2):179-183. [doi: 10.1002/nur.4770180211] [Medline: 7899572]
15. Sandelowski M. Whatever happened to qualitative description? Res. Nurs. Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g]
16. Boeije H. A purposeful approach to the constant comparative method in the analysis of qualitative interviews. Qual Quant 2002;36(4):391-409. [doi: 10.1023/A:1020909529486]
17. Glaser BG. The Constant Comparative Method of Qualitative Analysis. Social Problems 1965 Apr;12(4):436-445. [doi: 10.1525/sp.1965.12.4.03a00070]
18. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: Capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
19. Bailey L. Wearable Internet of Things Healthcare Systems, Virtual Care, and Real-Time Clinical Monitoring in Assessing and Treating Patients with COVID-19 Symptoms. Am. J. Med. Res 2021;8(1):91. [doi: 10.22381/ajmr8120219]
20. Mbunge E, Muchemwa B, Jiyane S, Batani J. Sensors and healthcare 5.0: transformative shift in virtual care through emerging digital health technologies. Global Health Journal 2021 Dec;5(4):169-177. [doi: 10.1016/j.glohj.2021.11.008]
21. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
22. Wu M, Luo J. Wearable technology applications in healthcare: a literature review. Online J Nurs Inf 2019;23(3).
23. Onorato S, Joshi A, Schwartz AW. Lights, Camera, Action: Optimizing Virtual Video Visits to Provide High-Quality Care. J Gen Intern Med 2021 Jun 27;36(6):1751-1754 [FREE Full text] [doi: 10.1007/s11606-020-06278-z] [Medline: 33111243]
24. Telemedicine five minute moments. Society of Bedside Medicine. URL: https://bedsidemedicine.org/telemedicine-5m2 [accessed 2022-08-09]
25. Telehealth Implementation Playbook. American Medical Association. 2022. URL: https://www.ama-assn.org/system/files/ama-telehealth-playbook.pdf [accessed 2022-08-09]
26. Benziger CP, Huffman MD, Sweis RN, Stone NJ. The Telehealth Ten: A Guide for a Patient-Assisted Virtual Physical Examination. Am J Med 2021 Jan;134(1):48-51 [FREE Full text] [doi: 10.1016/j.amjmed.2020.06.015] [Medline: 32687813]
27. Car J, Koh GC, Foong PS, Wang CJ. Video consultations in primary and specialist care during the covid-19 pandemic and beyond. BMJ 2020 Oct 20;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
28. Wray CM, Junge M, Keyhani S, Smith JE. Assessment of a multi-center tele-urgent care program to decrease emergency department referral rates in the Veterans Health Administration. J Telemed Telecare 2021 Jun 21:1357633X2110248. [doi: 10.1177/1357633x211024843]
29. Oh SS, Kim K, Kim M, Oh J, Chu SH, Choi J. Measurement of Digital Literacy Among Older Adults: Systematic Review. J Med Internet Res 2021 Feb 3;23(2):e26145. [doi: 10.2196/26145]
30. Lyles CR, Wachter RM, Sarkar U. Focusing on Digital Health Equity. JAMA 2021 Nov 09;326(18):1795-1796. [doi: 10.1001/jama.2021.18459] [Medline: 34677577]
31. Slightam C, Gregory AJ, Hu J, Jacobs J, Gurmessa T, Kimerling R, et al. Patient Perceptions of Video Visits Using Veterans Affairs Telehealth Tablets: Survey Study. J Med Internet Res 2020 Apr 15;22(4):e15682 [FREE Full text] [doi: 10.2196/15682] [Medline: 32293573]
32. Jacobs JC, Hu J, Slightam C, Gregory A, Zulman DM. Virtual Savings: Patient-Reported Time and Money Savings from a VA National Telehealth Tablet Initiative. Telemed J E Health 2020 Sep 01;26(9):1178-1183. [doi: 10.1089/tmj.2019.0179] [Medline: 31880502]
33. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catal Innov Care Deliv 2020 May 04:1-13. [doi: 10.1056/CAT.20.0123]
34. Lopez AM, Lam K, Thota R. Barriers and Facilitators to Telemedicine: Can You Hear Me Now? American Society of Clinical Oncology Educational Book 2021 Jun(41):25-36. [doi: 10.1200/edbk_320827]
35. Kirschner KL, Curry RH. Educating health care professionals to care for patients with disabilities. JAMA 2009 Sep 23;302(12):1334-1335. [doi: 10.1001/jama.2009.1398] [Medline: 19773571]

Abbreviations

VA: US Department of Veterans Affairs

Edited by G Eysenbach; submitted 18.04.22; peer-reviewed by S Azer, L Seuren; comments to author 27.05.22; revised version received 29.06.22; accepted 18.07.22; published 24.08.22.

Please cite as:
Gray C, Wray C, Tisdale R, Chaudary C, Slightam C, Zulman D
Factors Influencing How Providers Assess the Appropriateness of Video Visits: Interview Study With Primary and Specialty Health Care Providers
J Med Internet Res 2022;24(8):e38826
URL: https://www.jmir.org/2022/8/e38826
doi:10.2196/38826
PMID:36001364

Original Paper

Usability Testing of a Digital Assessment Routing Tool for Musculoskeletal Disorders: Iterative, Convergent Mixed Methods Study

Cabella Lowe[1], MSc; Mitchell Browne[1], MSc; William Marsh[2], PhD; Dylan Morrissey[1,3], PhD

[1]Centre for Sports & Exercise Medicine, William Harvey Research Institute, Queen Mary University of London, London, United Kingdom

[2]Risk and Information Systems Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom

[3]Department of Physiotherapy, Barts Health NHS Trust, London, United Kingdom

Corresponding Author:
Cabella Lowe, MSc
Centre for Sports & Exercise Medicine
William Harvey Research Institute
Queen Mary University of London
Mile End Hospital
Bancroft Road
London, E1 4DG
United Kingdom
Phone: 44 07976315105
Email: c.lowe@qmul.ac.uk

Abstract

Background: Musculoskeletal disorders negatively affect millions of patients worldwide, placing significant demand on health care systems. Digital technologies that improve clinical outcomes and efficiency across the care pathway are development priorities. We developed the musculoskeletal Digital Assessment Routing Tool (DART) to enable self-assessment and immediate direction to the right care.

Objective: We aimed to assess and resolve all serious DART usability issues to create a positive user experience and enhance system adoption before conducting randomized controlled trials for the integration of DART into musculoskeletal management pathways.

Methods: An iterative, convergent mixed methods design was used, with 22 adult participants assessing 50 different clinical presentations over 5 testing rounds across 4 DART iterations. Participants were recruited using purposive sampling, with quotas for age, habitual internet use, and English-language ability. Quantitative data collection was defined by the constructs within the International Organization for Standardization 9241-210-2019 standard, with user satisfaction measured by the System Usability Scale. Study end points were resolution of all grade 1 and 2 usability problems and a mean System Usability Scale score of ≥80 across a minimum of 3 user group sessions.

Results: All participants (mean age 48.6, SD 15.2; range 20-77 years) completed the study. Every assessment resulted in a recommendation with no DART system errors and a mean completion time of 5.2 (SD 4.44, range 1-18) minutes. Usability problems were reduced from 12 to 0, with trust and intention to act improving during the study. The relationship between eHealth literacy and age, as explored with a scatter plot and calculation of the Pearson correlation coefficient, was performed for all participants (r=–0.2; 20/22, 91%) and repeated with a potential outlier removed (r=–0.23), with no meaningful relationships observed or found for either. The mean satisfaction for daily internet users was highest (19/22, 86%; mean 86.5, SD 4.48; 90% confidence level [CL] 1.78 or –1.78), with nonnative English speakers (6/22, 27%; mean 78.1, SD 4.60; 90% CL 3.79 or –3.79) and infrequent internet users scoring the lowest (3/22, 14%; mean 70.8, SD 5.44; 90% CL 9.17 or –9.17), although the CIs overlap. The mean score across all groups was 84.3 (SD 4.67), corresponding to an *excellent* system, with qualitative data from all participants confirming that DART was simple to use.

Conclusions: All serious DART usability issues were resolved, and a good level of satisfaction, trust, and willingness to act on the DART recommendation was demonstrated, thus allowing progression to randomized controlled trials that assess safety and effectiveness against usual care comparators. The iterative, convergent mixed methods design proved highly effective in fully evaluating DART from a user perspective and could provide a blueprint for other researchers of mobile health systems.

International Registered Report Identifier (IRRID): RR2-10.2196/27205

(J Med Internet Res 2022;24(8):e38352) doi:10.2196/38352

KEYWORDS

mobile health; mHealth; eHealth; digital health; digital technology; musculoskeletal; triage; physiotherapy triage; usability; acceptability; mobile phone

Introduction

Background

Musculoskeletal disorders (MSDs) are prevalent across all ages, have shown an increase in the global disease burden over the past decade [1-3], and are associated with increased life expectancy and reduced activity [4,5]. MSDs are leading contributors to years lived with disability, early work retirement, and reduced ability to participate socially [5]. In many countries, they present the most significant proportional reason for lost productivity in the workplace, leading to significant impacts on the Gross Domestic Product and health care costs [6,7].

In the United Kingdom, the MSD burden of care poses a significant financial challenge to the National Health Service (NHS), costing £4.76 billion (US $3.84 billion) of resources and using up to 30% of primary care physician visits annually [8,9]. A freedom of information request has revealed that the average waiting times for NHS musculoskeletal outpatient physiotherapy services exceeded 6 weeks in the year to April 2019, with some patients waiting 4 months for routine physiotherapy appointments [10]. Longer waiting times can result in delays to physiotherapy services, with detrimental effects on pain, disability, and quality of life for waiting patients [11,12], highlighting the need for a targeted policy response [3,13].

Reducing inconsistency in clinical pathway delivery, including unwarranted secondary care consultations and investigations, forms part of the “Getting It Right First Time (GIRFT)” national program implemented within the UK NHS and has demonstrated cost reduction across the musculoskeletal pathway, particularly relevant in overburdened health care systems [14] Musculoskeletal triage as a single point of entry is effective in improving user satisfaction, diagnostic agreement, appropriateness of referrals, and reduction in patient waiting times [15], where it has been demonstrated to be effective using several methods by a range of clinicians [16-18]. However, using clinicians to provide MSD triage carries its own challenges [19].

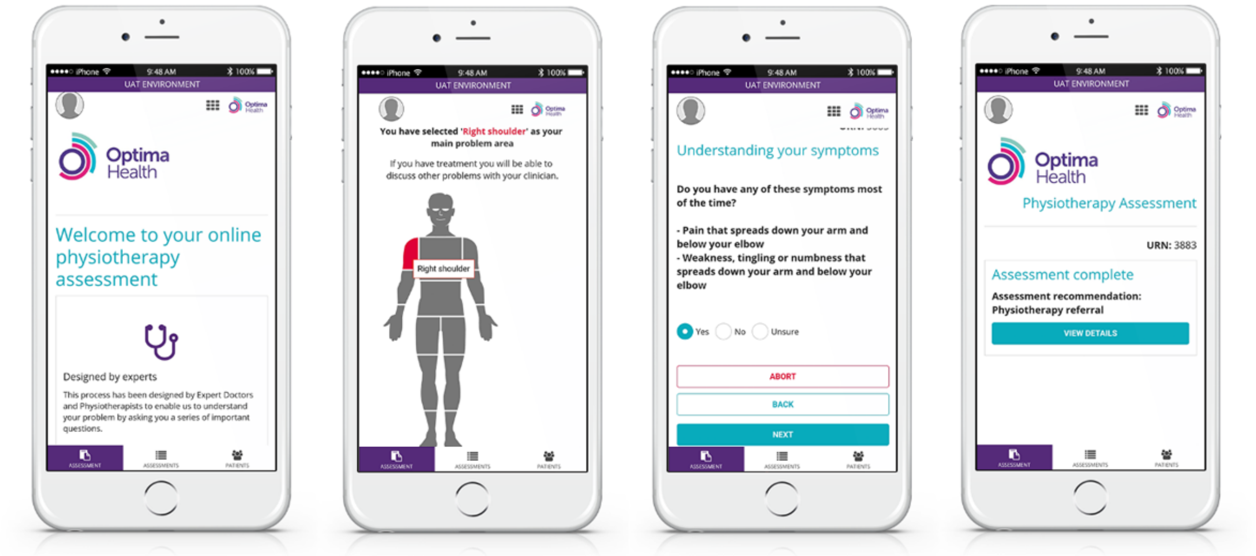
Mobile health (mHealth), defined by the World Health Organization as a medical or public health practice that is supported by mobile devices [20], has seen rapid evolution and adoption, and currently, smartphone apps have the potential to make the treatment and prevention of diseases cost-efficient and widely accessible [21,22]. Optima Health has developed the mHealth Digital Assessment Routing Tool (DART) specifically for triaging MSDs, delivering a narrower but deeper assessment than that found with more generic symptom checkers. A digital alternative to clinician-led triage, which is able to replicate the same stratification of care and reduction in costs, is a desirable objective, although some mHealth tools have not demonstrated cost-effectiveness or have merely shown a shift in spending to another part of the health system [23]. It is also recognized that many mHealth apps fail to scale up from a prototype to successful implementation, with inattention to usability during the design and testing phases being identified as a potential cause of the high abandonment rate [24-27]. Although acknowledging usability is crucial in the design, development, testing, and implementation of mHealth apps [28-32], a consistent approach to testing has not yet been established, with researchers using a combination of different study methodologies [33].

An iterative, convergent mixed methods design was used to assess the usability of DART, using cyclical evaluation and improvement plus mixed methods to provide richness while quantifying use, maximizing usability, and therefore supporting system adoption [34]. The testing protocol for this study has been described in detail in a previous publication [35].

DART Overview

DART is a first contact mHealth system comprising an algorithm distinguished by 9 body areas, providing the patient with a recommendation for the most appropriate level of intervention based on their responses (Figure 1). Screening for serious pathologies is completed at the start of the assessment, with less urgent medical referrals being identified as the patient passes through the questioning. The referrals recommended by the algorithm are configured to match the service provider’s local MSD pathways. DART typically signposts emergency or routine medical assessments, specific condition specialists, physiotherapy, self-management programs, and psychological support services.

Figure 1. The Digital Assessment Routing Tool mobile health system.



Welcome to your online physiotherapy assessment

Designed by experts

This process has been designed by Expert Doctors and Physiotherapists to enable us to understand your problem by asking you a series of important questions.

You have selected 'Right shoulder' as your main problem area

If you have treatment you will be able to discuss other problems with your clinician.

Understanding your symptoms

Do you have any of these symptoms most of the time?

- Pain that spreads down your arm and below your elbow
- Weakness, tingling or numbness that spreads down your arm and below your elbow

● Yes ○ No ○ Unsure

Physiotherapy Assessment

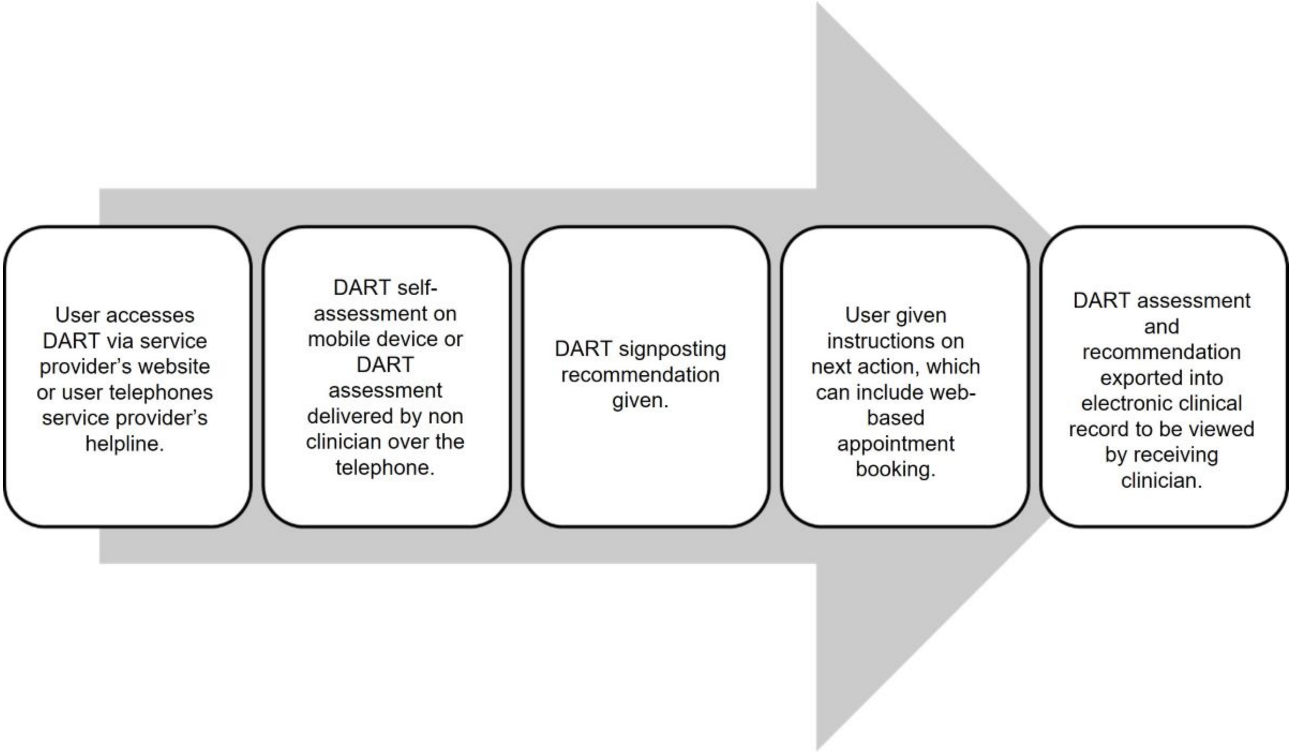
URN: 3883

Assessment complete

Assessment recommendation: Physiotherapy referral

Integration of DART with the provider's clinical record system means that assessment data and recommendations can be made instantly available to the receiving clinician. Using a link on the clinical provider's website, DART can be accessed 24/7 using a mobile device or computer, directing users to care at an earlier stage of their injury than would be possible via a traditional clinician-led triage process (Figure 2). Alternatively, DART can be delivered over the telephone by a nonclinician. Reduction in treatment waiting times and reallocation of triage clinical resources to more complex assessments and management could hold significant benefits for the user and health care system.

Figure 2. Integration of the DART mobile health system within an existing musculoskeletal disorder pathway. DART: Digital Assessment Routing Tool.



Previous Work

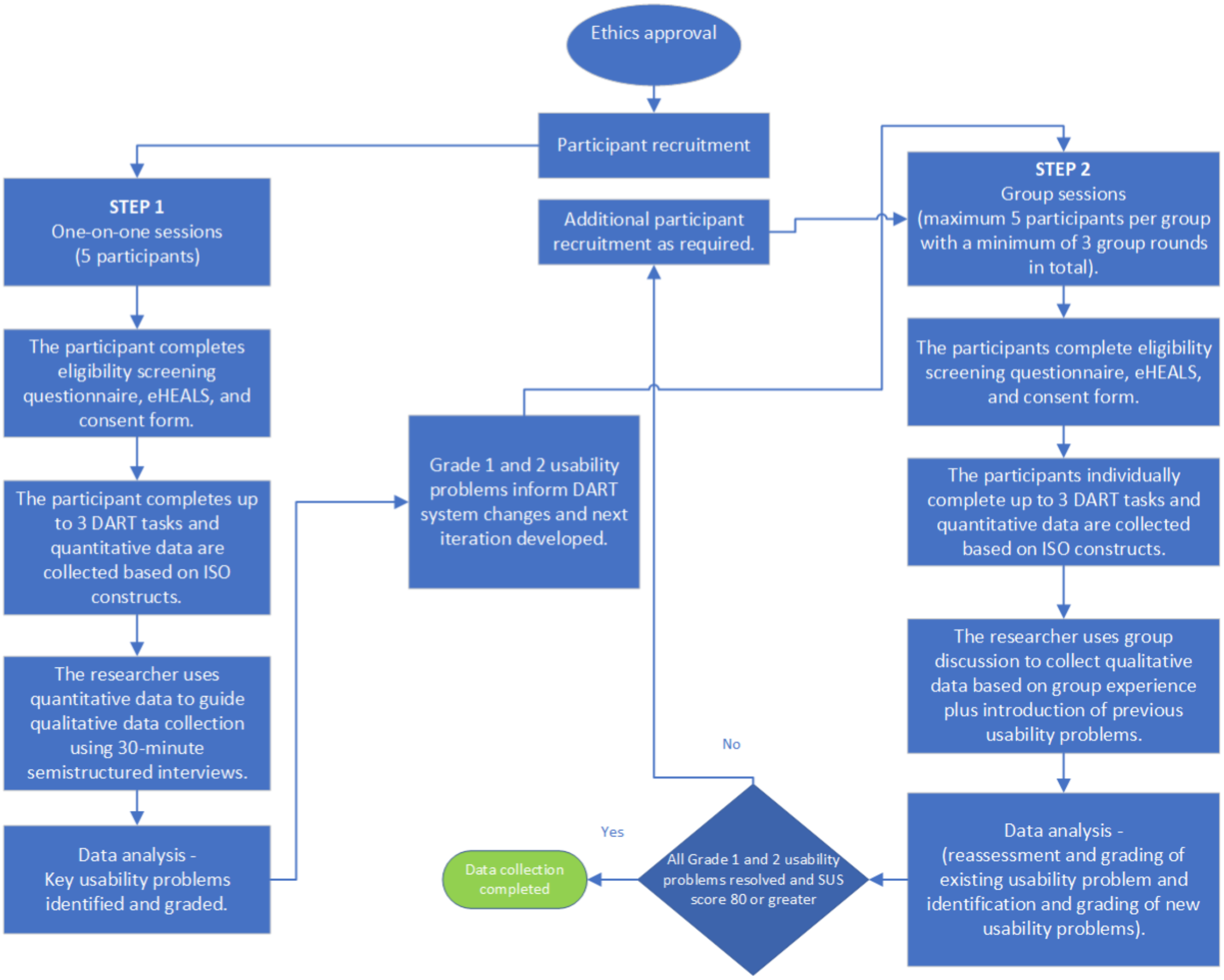
This usability study is part of a larger project, bringing DART from concept to implementation through a series of clinical and academic research work packages. Clinical algorithm validity was assessed by a panel of clinical experts using vignettes incorporating common MSD presentations, as well as red flags and complex presentations, with the panel deeming the validity to be sufficient to allow DART to proceed to further research studies. The protocol devised for this usability study went through a series of iterations within an internal review process, comprising the research project team and DART system developers to arrive at the final version [35]. The objective of this study was to optimize usability before evaluating the safety and effectiveness of DART through a randomized controlled trial, the pilot protocol for which has been published [36].

Methods

Study Design

This study used an iterative, convergent mixed methods design, the protocol for which has been published elsewhere [35]. Step 1 involved in-depth interviews with 5 participants to identify key usability issues, followed by step 2, where group sessions captured greater diversity of data from a potential DART user population (Figure 3). Quantitative data collection was defined by the constructs of effectiveness, efficiency, and satisfaction within the International Organization for Standardization (ISO) 9241-210-2019 standard [30] and provided researchers with a focus for qualitative data collection during both steps. Accessibility was monitored throughout the testing process following the principles described in ISO 30071-1-2019 for embedding inclusion within the design process [31]. Mixed methods data collection and analysis continued cyclically through all rounds of testing until the fourth DART mHealth system iteration was found to perform according to the agreed criteria and the study end points of all grade 1 and 2 usability problems being resolved, as well as a mean System Usability Scale (SUS) score of ≥80, were achieved. The relationship between the likelihood to recommend a system and the mean SUS score has been found to be strongly correlated, and a score of ≥80 was chosen as a study end point as achievement of this threshold is considered to increase the probability of users recommending the system to a friend, therefore positively affecting adoption [37].

Figure 3. DART usability study iterative, convergent mixed methods design. New participants were recruited for each testing round. Participants raising specific issues in previous rounds were invited individually to review and provide feedback on changes. DART: Digital Assessment Routing Tool; eHEALS: eHealth Literacy Scale; ISO: International Organization for Standardization.



Participant Recruitment

A stratified purposive sampling method was used to gather information from participants by using a sampling matrix and quotas [38], categorized by participant age, internet use, sex, and English for speakers of other languages (ESOL) groups—all of which are subgroups that have shown to contribute small differences in internet use [39]. For this study, “daily internet users” were defined as individuals who access the internet every day or almost every day, and “infrequent internet users” were those who were not daily users but had accessed the internet within the past 3 months [39]. Recruitment was conducted via flyers and emails to local community groups, Optima Health's existing client base of employers and staff, and Queen Mary University of London students, as well as via social media. In the latter stages, snowballing yielded participants with characteristics of interest; study recruitment continued throughout the study process until the study end points were reached. Potential participants expressing an interest were sent a patient information sheet and consent form and had the opportunity to review this material before consenting to join the study. A total of 33 individuals expressed an interest in participating, of whom 22 (67%) enrolled in the study after meeting the screening criteria.

Inclusion and Exclusion Criteria

The study participant inclusion criteria were as follows: (1) adults aged >18 years; (2) able to speak and read English; (3) living in the United Kingdom; (4) accessed the internet at least once every 3 months; (5) access to a smartphone, tablet, or laptop; and (6) current or previous experience of a musculoskeletal condition.

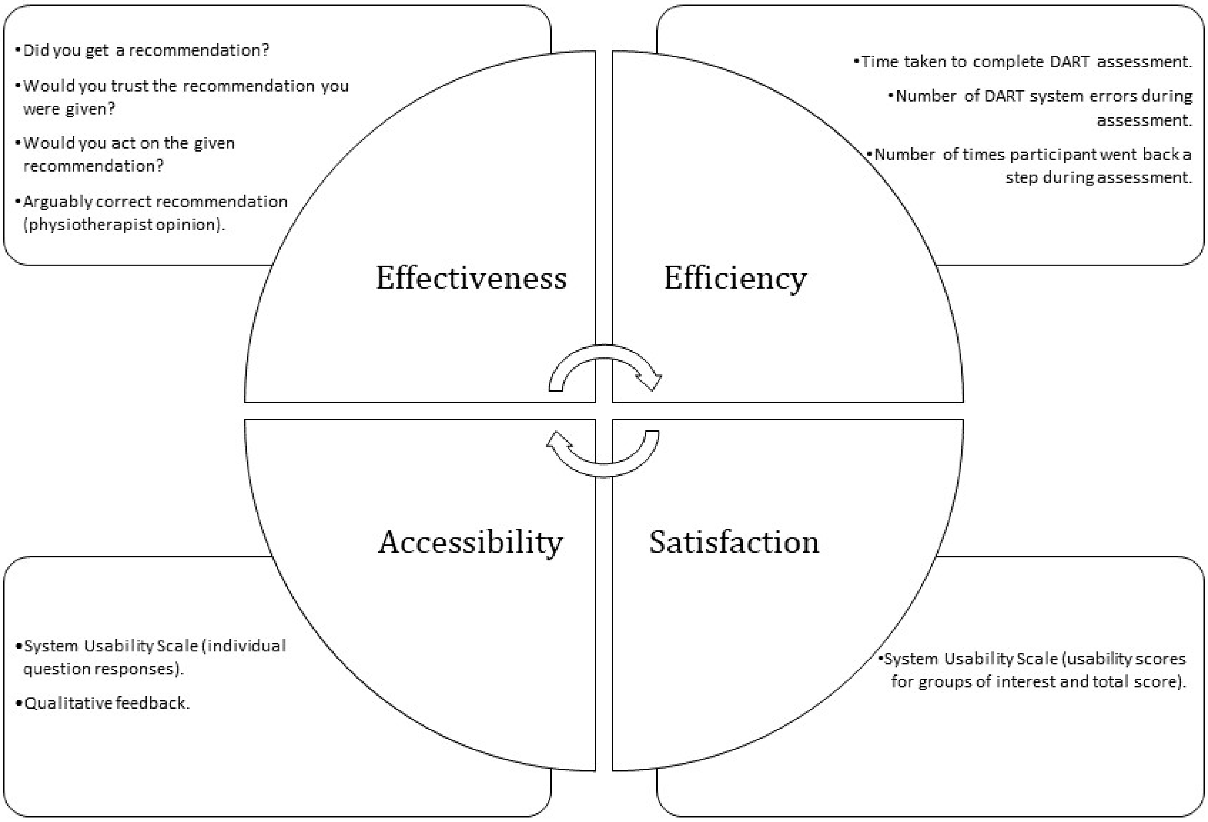
The study participant exclusion criteria were as follows: (1) significant visual or memory impairment sufficient to affect the ability to answer questions and recall information in an individual or group discussion setting; (2) medically trained musculoskeletal health care professional, such as a physician or physiotherapist; (3) relatives or friends of the researchers; and (4) Optima Health employees.

Data Collection

Following the attainment of consent, participants completed a short questionnaire, including the eHealth Literacy Scale (eHEALS) [40], to provide demographic data and were given instructions by the researcher on how to log into the DART system test site. The first 5 participants in step 1 attended one-on-one video call interviews lasting up to 60 minutes where they could choose up to 3 existing or previous musculoskeletal conditions to complete assessments while being encouraged to give feedback using the concurrent think-aloud method [41]. Participant choice was not limited to specific body sites as usability features were synonymous across all 9 body sites. The participants in step 2 tested DART individually and then attended 30-minute video call group discussion sessions facilitated by the researcher.

Assessing DART performance using satisfaction scales alone was not considered adequate; thus, data collection parameters were defined using the ISO constructs (Figure 4). Following their DART assessments, all participants completed a questionnaire and the SUS [42-44]. The researcher (physiotherapist with postgraduate MSD qualifications) assessed the clinical accuracy of the DART recommendation based on the diagnosis the participant had been given by their treating clinician. Quantitative data were also taken from the DART system itself.

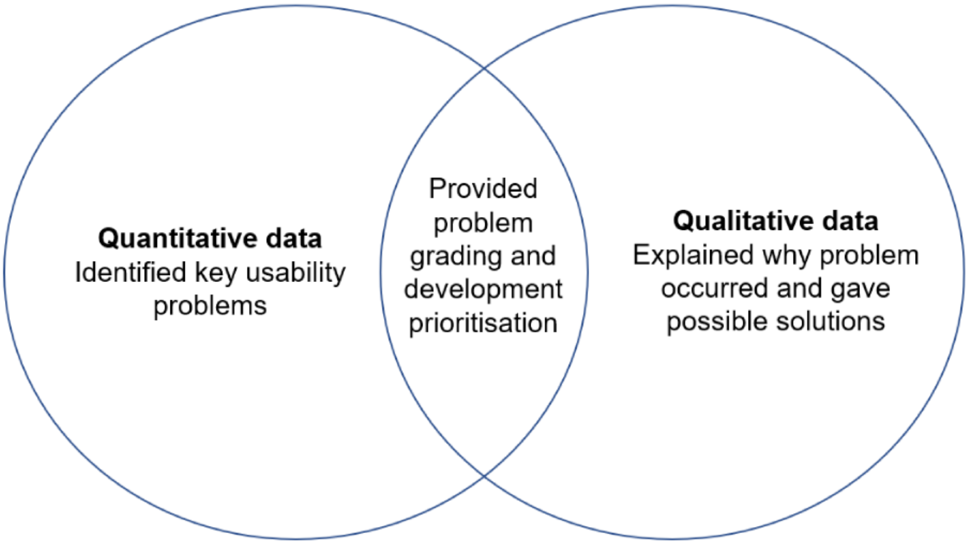
Figure 4. Data collection methods used to assess DART performance against the International Organization for Standardization 9241-210-2019 and International Organization for Standardization 30071-1-2019 standards constructs of effectiveness, efficiency, and satisfaction [30] and accessibility [31]. DART: Digital Assessment Routing Tool.



Quantitative data aligned to the ISO 9241-210-2019 standard constructs were generated from participant questionnaires and DART system data, as shown in Figure 4. These informed the researcher's qualitative data collection, aided by the use of a visual joint display, merging both types of data to illuminate not only usability problem themes but also potential system improvements (Multimedia Appendix 1). Qualitative data recorded during the interviews and group sessions were transcribed verbatim using the Otter transcription software (Otter.ai; automated video and audio transcription software) and checked for accuracy against the original recording. During group sessions, previous usability problems were introduced to participants to assess the impact of changes made to the previous iteration. In addition, users who raised specific issues in previous rounds were invited individually to review and feedback on changes. In addition to usability problems, any participant feedback on accessibility or positive aspects of DART was recorded.

Data analysis occurred after each round of testing and leveraged the strengths of the convergent mixed methods design to identify usability issues and inform the changes required for subsequent DART iterations (Figure 5). Of particular importance was the thematic analysis of qualitative data provided by real-world users, which ensured their views were included in the DART system development to improve usability.

Figure 5. The convergent mixed methods design, where both data types are collected simultaneously to allow the analysis and grading of usability problems, thus informing the next system iteration.



Data analysis was performed to identify the overall satisfaction score and differences between groups (mean score, SD, and confidence level [CL]). Statistical analyses examined the relationship between participant age and eHealth literacy using Pearson correlations.

Restrictions imposed during the COVID-19 pandemic led to all data collection sessions being conducted remotely using Microsoft Teams videoconferencing software and web-based questionnaires.

Ethics Approval

This study received approval from the Queen Mary University of London Ethics of Research Committee (QMREC2018/48/048) in June 2020.

Data Analysis

Extending the convergent mixed methods design from data collection to analysis, the reporting used a weaving approach where usability problems were brought together on a theme-by-theme basis and presented through joint displays [45].

Quantitative data from web-based questionnaires and measures of efficiency from the DART system were analyzed and reported to identify key usability issues. Participant SUS raw scores were converted and analyzed by groups of specific interest (daily internet users, infrequent internet users, and ESOL), and the amalgamated mean score across all participants was converted into a percentile score to provide benchmarking against other web-based systems [46].

To minimize bias, quantitative data were collected by an independent researcher during the initial 5 semistructured interviews, and web-based questionnaires were used for the group sessions. Using a thematic analysis approach, qualitative data derived from transcripts of interviews and group sessions were reviewed and analyzed systematically by the 2 researchers independently. Patterns and clusters of meaning within the data were identified and labeled according to the area of system functionality. Data not directly related to the overall research question were excluded. The 2 researchers then worked together to agree and create a thematic framework with higher-order key usability themes able to address the research objective [47,48]. Data were indexed into usability problems of key importance to the study and quotes extracted for each subtheme, thus providing the details required to make the system changes needed to remove or mitigate grade 1 and 2 usability issues. The researchers, working independently initially and then together, arrived at a consensus and allocated a problem severity grade to each usability problem. This was obtained by considering the impact and frequency of the problem, leading to a decision on the risk of not addressing the problem versus the reward of correcting it [49] (Table 1). Once problems had been graded, matched system developments were passed to the DART system developers to guide the next iteration. Actions to address all grade 1 and 2 usability problems were completed for the next iteration, together with closely associated grade 3 and 4 problems if they fell within the scope of the development work. All usability problems remained on record and were reassessed after each round and, if necessary, regraded. Positive feedback about the system was also reported.

Table 1. Usability problem grading criteria, adapted from guidance issued by The Food and Drug Administration [49].

Grade	Impact	Frequency	Implications	Action
1	High	High, moderate, or low	Prevents effective use of the system	Address in next study iteration
2	Moderate or low	High or moderate	Affects the quality of system delivery	Address in next study iteration
3	Moderate or low	Low or moderate	Minor issues for several users or a small number of users highlighting concerns important to them	Document and address in later development
4	Low	Low	Small issues that, if resolved, could improve user satisfaction	Document and address in later development

Statistical Analysis

The relationship between participant age and eHealth literacy was analyzed using Pearson correlations in Microsoft Excel (a spreadsheet with statistical analysis functionality) to identify user groups less likely to use DART successfully.

Differences in satisfaction scores were present between groups, with expert internet users having the highest mean score (mean 86.5, SD 4.48; 90% CL 1.78).

Results

Overview

A total of 22 participants were enrolled and completed the study (Table 2). The first testing round comprised 23% (5/22) of participants who completed qualitative "think-aloud" data collection led by a researcher familiar with the system and with training in the use of the method. It has been suggested that this relatively small number of participants is sufficient to expose 75% of usability issues, including all catastrophic problems, with further testing of subsequent iterations using new participants to identify less serious problems [50]. This proved to be the case, and data sufficiency was achieved. This was supported by a narrow study aim and the quality of dialog with the first 5 participants. The final sample size was not predefined and was re-evaluated after each round of results [51].

There was representation from all the groups of interest; however, not all quotas were met, and small sample sizes, especially infrequent internet users, resulted in a skew of data in favor of daily internet users. This compromised detailed statistical analyses across groups (Table 3).

Table 2. Participant characteristics (N=22).

Characteristic	Daily internet users	Infrequent internet users	ESOL[a,b]	All groups
Total sample, n (%)	19 (86)	3 (14)	6 (27)	22 (100)
Age (years)				
Values, mean (SD)	47.6 (15.7)	55 (11.4)	41 (8.5)	48.6 (15.2)
Values, range	20-77	47-68	31-55	20-77
Sex (male), n (%)	9 (41)	1 (5)	3 (14)	10 (46)
eHEALS[c] score				
Values, mean (SD)	29 (8)	25 (4)	26 (12.3)	28.8 (7.8)
Values, range	8-38	21-29	8-37	(8-38)

[a] ESOL: English for speakers of other languages.

[b] All ESOL participants were also daily internet users.

[c] eHEALS: eHealth Literacy Scale.

Table 3. Recruitment matrix showing minimum quotas and number of participants recruited by characteristics of interest (N=22)[a].

Characteristic	Daily internet user (n=19)		Infrequent internet user (n=3)	
	Quota	Enrolled, n (%)	Quota	Enrolled, n (%)
Age (years)				
18-54	2-4	7 (37)	1-3	2 (67)
55-74	2-4	10 (53)	1-3	1 (33)
≥75	1-3	1 (5)	2-4	0 (0)
Sex				
Male	Minimum 6	7 (37)	Minimum 4	1 (33)
Female	Minimum 6	10 (53)	Minimum 4	2 (67)
ESOL[b]				
Non-ESOL	Minimum 6	15 (79)	Minimum 6	3 (100)
ESOL	Minimum 2	6 (32)	Minimum 2	0 (0)

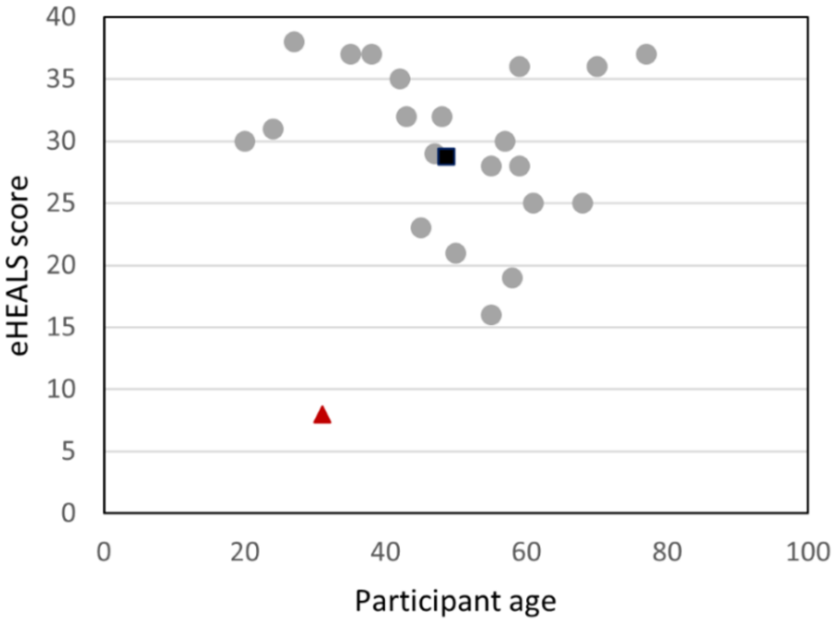
[a] Total study participants quota was 20.

[b] ESOL: English for speakers of other languages.

We were interested to know whether the frequency of internet use, age, eHealth literacy, or being a speaker of English as a second language would affect DART usability, as these factors have been highlighted as potential variables in mHealth adoption [39]. There was a wide range of eHEALS scores across participants (mean 28.8, SD 7.8; 95% CI 25.1-32.3), with the highest score of 38/40 achieved by a daily internet user aged 27 years and the lowest score of 8/40 achieved by an ESOL daily internet user aged 31 years. The oldest participant (aged 77 years) achieved a score of 37/40, and the youngest participant (aged 20 years) scored 30/40.

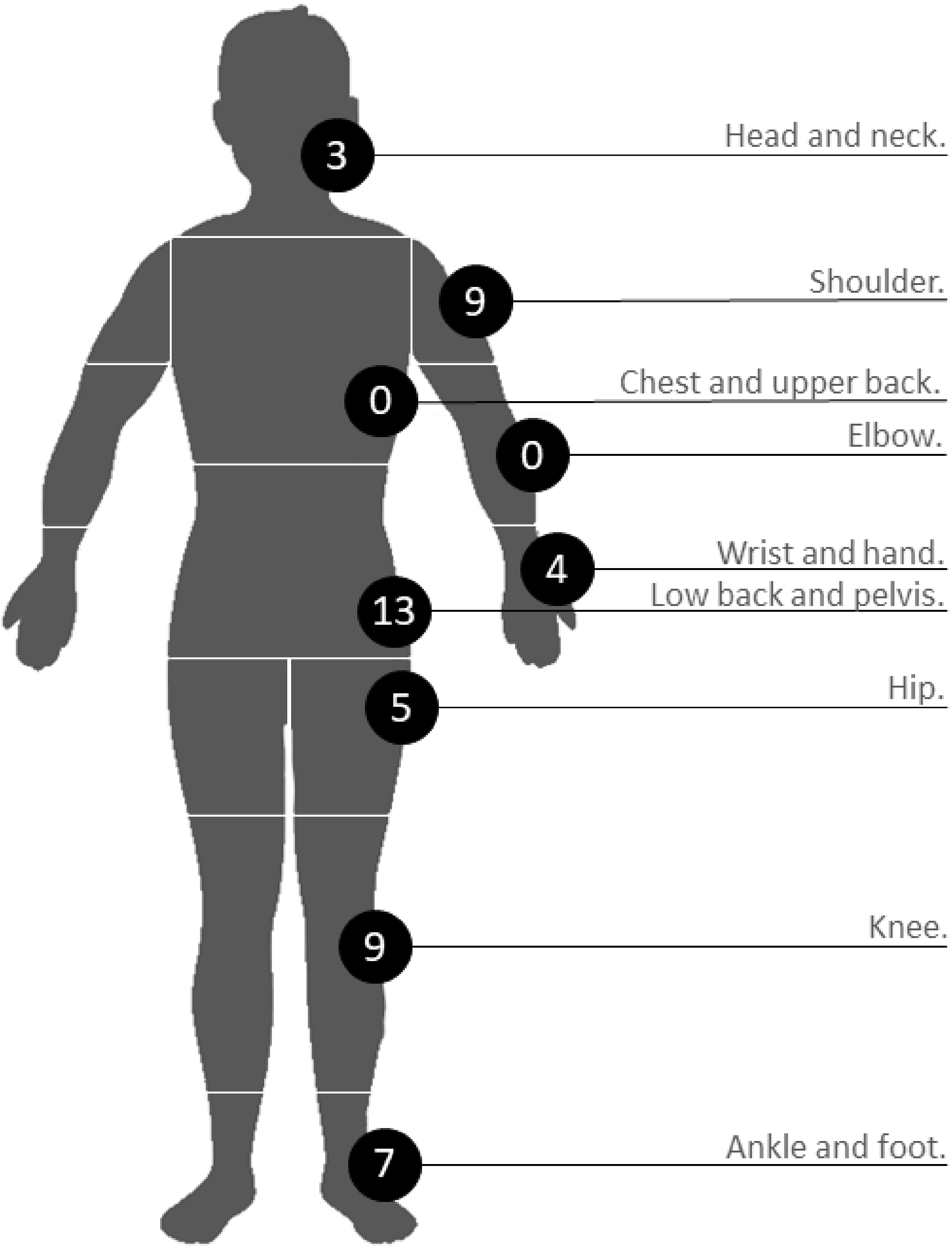
The relationship between eHealth literacy and age, as explored with a scatterplot and calculation of Pearson correlation coefficients, was performed for all participants (20/22, 91%; *r*=–0.2) and repeated with the potential outlier removed, as indicated in Figure 6 in red (19/22, 86%; *r*=–0.23), with no meaningful relationship observed or found for either.

Figure 6. Relationship between age and eHealth literacy scores across all participants. (eHEALS scores range from 0 to 40 scale, with higher scores being better). eHEALS: eHealth Literacy Scale.



A total of 50 assessments were completed by the 22 participants across a possible 9 body sites (Figure 7). The most frequently chosen body site was the low back and pelvis (13/22, 26%), followed by shoulder and knee (both 9/22, 18%). Two body sites were not selected by participants for testing: chest and upper back and elbow. Within a typical MSD triage service, these are often the least occurring body sites. However, the usability features are consistent with those of the other body regions; thus, it is unlikely that any new problems would have been identified through the selection of these pathways.

Figure 7. Body sites selected by participants. The number of selections represents the total of the front, back, and either side of a given body site. The Digital Assessment Routing Tool algorithms are designed to assess for musculoskeletal disorder conditions that occur or refer to pain in the selected body site.



Usability Problems

A total of 19 individual usability problems were identified across all 5 rounds of testing, of which 12 (63%) were initially classified as grade 1 or 2. These grades were either reduced or resolved over the iterations. DART iteration 4 was reviewed by participants during testing round 4, and no grade 1 or 2 usability problems were found. This was validated during testing round 5, and the study end points were achieved (Figure 8).

Within the grade 1 and 2 usability problems, 3 main themes and 7 contributory subthemes were identified (Figure 9).

Over each of the 5 testing rounds, grade 1 and 2 usability problems were discussed with the participants and regraded (Table 4).

Figure 8. Number of usability problems across testing rounds by grade. The incidence and problem grading changed over the 5 rounds of testing, with grade 1 and 2 problems being negated or reduced to a lower grade. All grade 3 and 4 issues were documented, reviewed, and prioritized for future Digital Assessment Routing Tool development.

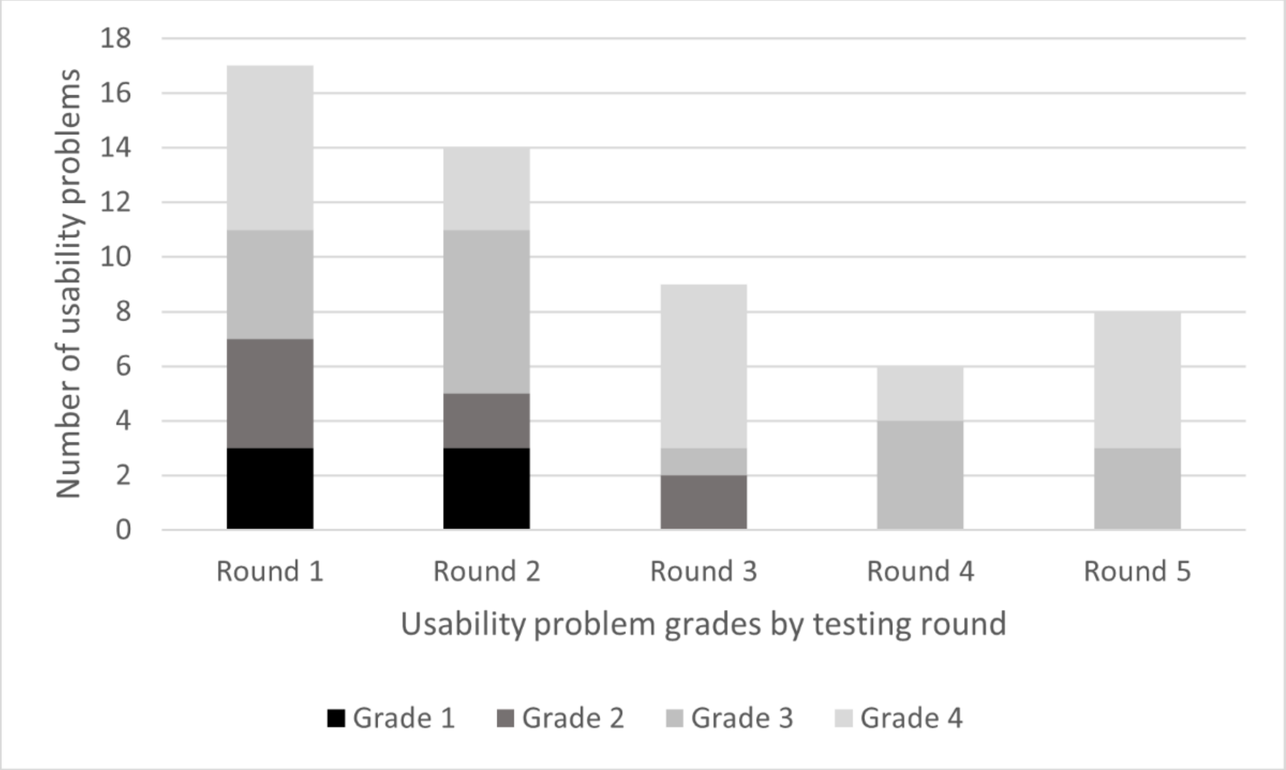
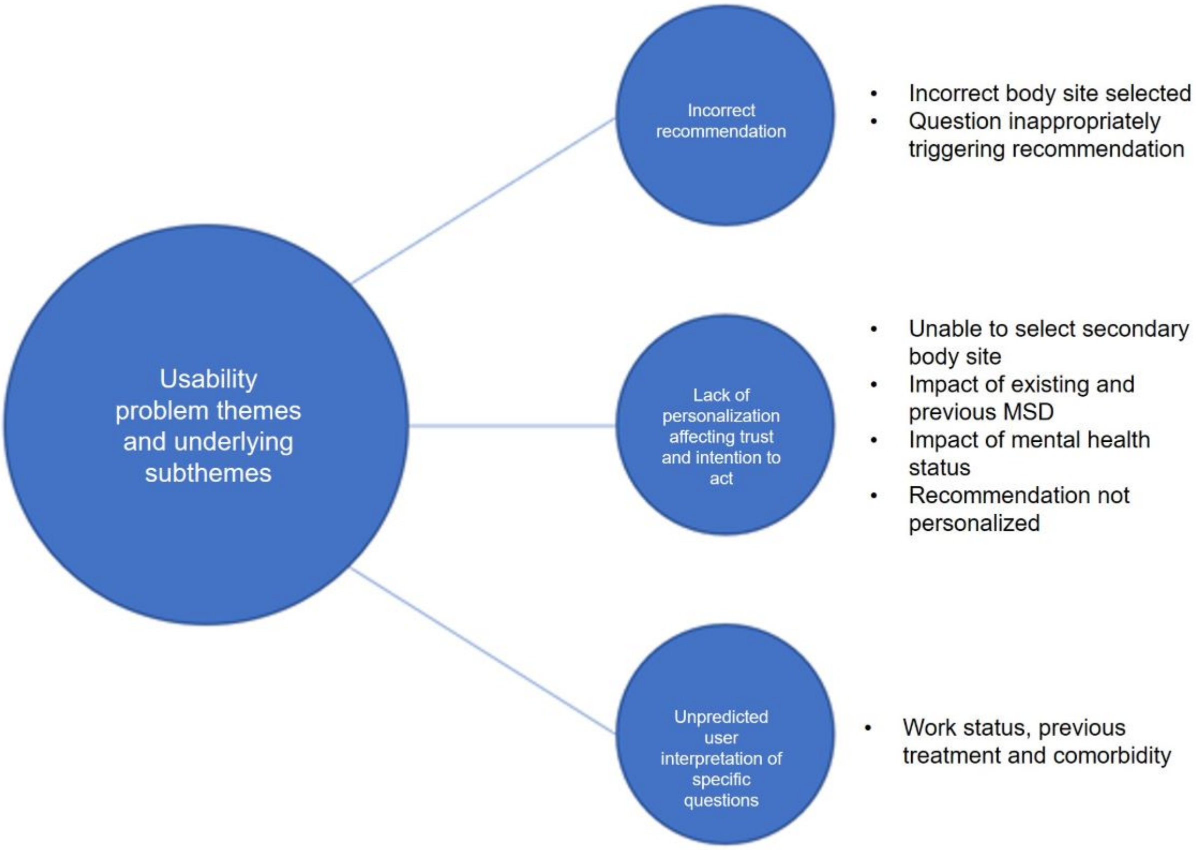


Figure 9. Grade 1 and 2 usability problem themes and underlying subthemes. MSD: musculoskeletal disorder.



- Incorrect body site selected
- Question inappropriately triggering recommendation

- Unable to select secondary body site
- Impact of existing and previous MSD
- Impact of mental health status
- Recommendation not personalized

- Work status, previous treatment and comorbidity

Table 4. Grade 1 and 2 usability problem themes, subthemes, and participant quotes over 5 rounds of testing[a].

Underlying theme and subthemes	Usability problem grade of subtheme (testing round)	Participant quotes
Theme 1: incorrect DART[b] recommendation compared with expert opinion		
Incorrect body site selected	• 1 (1) • 2 (2) • 2 (3) • 4 (4) • 4 (5)	• "If people are like me, they don't read things properly, especially at the beginning. Was there an option to start again, because people might mess up?" [DART005] • "The only time that I felt slightly lacking in confidence was on the body site. That was the only time I wasn't sure the system would grab the information I clicked." [DART002]
Question inappropriately triggering recommendation—systemic inflammatory disease and central nervous system condition	• 1 (1) • 1 (2) • 0 (3) • 0 (4) • 0 (5)	• "I think that came because I said that I'm stiff in the morning, that eight minutes or whatever." [DART006] • "So I read weakness, not severe weakness. So, I can stand on it and support my weight, but it just hurts like hell, rather than not being able to support myself." [DART015]
Theme 2: lack of personalization affecting participants' trust and intention to act		
Unable to select secondary body site	• 0 (1) • 2 (2) • 2 (3) • 0 (4) • 0 (5)	• "I think if I could have put more evidence in, then I'd be more likely to follow the recommendation at the end because I think it was relevant to me." [DART006] • "I suppose you could differentiate slightly more between the source point of the pain and the consequences for your other limbs like, you know, I knew very well that it was bad at the back that was causing my inability to walk. So maybe distinction between primary pain and a secondary or referred pain might be useful." [DART010]
Impact of existing and previous MSD[c]	• 2 (1) • 0 (2) • 0 (3) • 0 (4) • 4 (5)	• "I don't want to waste the GP's time or my time waiting for an appointment to be told what I already know. So, in my two cases, it wasn't so much about diagnosis is more of an okay, this has returned. We know the course of action." [DART019] • "I wasn't sure whether sometimes we're talking about what it's like when it's really bad, or what it's like in general." [DART005]
Impact of mental health status	• 0 (1) • 2 (2) • 3 (3) • 3 (4) • 3 (5)	• "I think one way of making it better is also seeing how it affects someone psychologically as well. I think that this is something which can sometimes be overlooked, but I think it's important to see how it is impacting on someone's emotional wellbeing?" [DART001] • "When you seek sort of medical advice, or you have a condition that gives you worry and anxiety, probably you expect a little bit more than just sort of self-treatment." [DART016]
Recommendations not sufficiently personalized	• 4 (1) • 2 (2) • 0 (3) • 0 (4) • 0 (5)	• "I guess, it might need to be a little bit more personalized recommendations depending on what people choose." [DART001] • "I suppose the only thing that might dissuade people, would be that if they were users of it, and it came up with the same sort of end page every time." [DART002]
Theme 3: participant difficulty in interpreting questions		
Specific questions (work status, previous treatment, and comorbidities)	• 2 (1) • 3 (2) • 4 (3) • 0 (4) • 4 (5)	• "You were distinguishing between people who were employed, and people who are not employed. It just seemed to me as though there was quite a large category of people lumped together in that one box and maybe it would be better to differentiate them a bit more, so that they did actually tick retired or they ticked student" [DART010] • "Where you were asked whether you'd had surgery or physio, it just was rather a broad question. I thought maybe it should have been a tick box for that to show which one you'd had." [DART004] • "You're a little bit unsure about whether it's really registered to your osteoporosis." [DART010]

---

[a] Usability problems were clustered into subthemes based on specific areas of DART functionality. Problem grades were reduced in severity over testing rounds as problems were negated or reduced during DART iterations (grade 1 is the most severe, and grade 4 is the least severe).

[b] DART: Digital Assessment Routing Tool.

[c] MSD: musculoskeletal disorder.

Construction of a joint display showed how different types of data were combined to assess performance against the ISO 9241-210-2019 constructs (Tables 5, 6, and 7).

Table 5. Display of qualitative data by International Organization for Standardization 9241-210-2019 standard constructs (effectiveness, efficiency, and satisfaction)[a].

Construct and goal	Participant quotes
Construct 1: effectiveness	
Assessment results for a recommendation being given	• "I found it really user friendly and I found I could read the questions quite quickly and just give an answer and move on." [DART018]
Assessment results for a correct clinical recommendation;	• "I expected the area [selected body part] that I chose to change color, I would do it a different color, red or something like that." [DART005]
Assessment of whether the participant would trust	• "It might be easier if you just say have a secondary field to sort of like give your secondary issues as well. You know, sometimes it's just not, it's like the neck runs into the arm and lower parts, but it can be different things as well." [DART014] • "It might make people feel a bit more confident that they've done it right." [DART015]
Assessment of whether the participant would act upon	• "I think if I could have put more evidence in, then I'd be more likely to follow the recommendation at the end, because I think it was relevant to me." [DART006]
Construct 2: efficiency	
Time taken to reach recommendation (minutes)	• "It was very quick. And I quite like that it has one thing for one page, which is a very short question, it gives you a few options, and then you answer so you don't have to go through long text questions, one after the other. So, it just takes you very quickly step by step. And it's quite, I don't know, for me, it was super easy and clear to answer questions." [DART017]
DART[b] system errors	• "I found it really simple system to use very, very easy and had no problems at all." [DART010B]
DART system backsteps	• "That was a question about whether I'd been off work for a long time and if I'm employed or self-employed, something that I didn't find quite straightforward." [DART020]
Construct 3: satisfaction	
System Usability Scale score per round	• "If I had this actual system, I would have saved £150 in cash and probably three months of pain had I been able to access it when I had my problems with my back." [DART013] • "It's done me a favor actually, because I was in two minds whether to try and get a private injection, whether to go to an osteopath or physio. I think it might save me money in the long run." [DART014]

[a] Participant quotes provide a deeper understanding of system performance and usability problems.

[b] DART: Digital Assessment Routing Tool.

Table 6. Display of quantitative data by International Organization for Standardization 9241-210-2019 standard constructs (effectiveness, efficiency, and satisfaction) over 5 testing rounds[a].

Construct, goal, and testing round	Result
Construct 1: effectiveness[a]	
Assessment results for a recommendation being given; participants in testing round achieving construct theme (%)	
Round 1	13 (100)
Round 2	11 (100)
Round 3	11 (100)
Round 4	10 (100)
Round 5	5 (100)
Assessment results for a correct clinical recommendation; participants in testing round achieving construct theme (%)	
Round 1	11 (85)
Round 2	5 (45)
Round 3	11 (100)
Round 4	10 (100)
Round 5	10 (100)
Assessment of whether the participant would trust; participants in testing round achieving construct theme (%)	
Round 1	13 (100)
Round 2	9 (82)
Round 3	11 (100)
Round 4	8 (80)
Round 5	5 (100)
Assessment of whether the participant would act upon, n (%)	
Round 1	12 (92)
Round 2	8 (73)
Round 3	11 (100)
Round 4	8 (80)
Round 5	4 (80)
Construct 2: efficiency[b]	
Time taken to reach recommendation (minutes)	
Round 1	
Values, mean (SD)	Not recorded
Values, range	Not recorded
Round 2	
Values, mean (SD)	5.7 (5.35)
Values, range	1-18
Round 3	
Values, mean (SD)	5.4 (4.54)
Values, range	1-15
Round 4	
Values, mean (SD)	3.5 (1.5)
Values, range	1-5
Round 5	
Values, mean (SD)	7.4 (2.13)

Construct, goal, and testing round	Result
Values, range	3-17
All groups	
Values, mean (SD)	5.2 (4.44)
Values, range	1-18
DART[c] system errors	
Round 1	0
Round 2	0
Round 3	0
Round 4	0
Round 5	0
DART system backsteps	
Round 1	1
Round 2	2
Round 3	2
Round 4	2
Round 5	6
Construct 3: satisfaction[d]	
System Usability Scale score per round[e]	
Round 1	
Values, n (%)	5 (23)
Values, mean (SD)	91.6 (4.23)
Margin of error	4.46 or –4.46
Round 2	
Values, n (%)	6 (27)
Values, mean (SD)	87 (10.23)
Margin of error	12.72 or –12.72
Round 3	
Values, n (%)	5 (23)
Values, mean (SD)	79.5 (16.91)
Margin of error	21.02 or –21.02
Round 4	
Values, n (%)	2 (9)
Values, mean (SD)	78.8 (18.75)
Margin of error	N/A[f]
Round 5	
Values, n (%)	4 (18)
Values, mean (SD)	78.8 (5.73)

Construct, goal, and testing round	Result
Margin of error	9.11 or –9.11

[a] Quantitative data show the number of participants in each round and the percentage that achieved the construct theme.

[b] Time taken to complete an assessment (time taken to reach a disposition was not measured during round 1, as the “think-aloud” method of data capture was prioritized at this stage); number of system errors where the participant was unable to navigate to the end of the assessment because of a system technical error; backsteps where the participant moved back to the previous question.

[c] DART: Digital Assessment Routing Tool.

[d] System Usability Scale scores by round, group of interest, and across all groups.

[e] Responses were scored on a 5-point Likert scale (1=strongly disagree and 5=strongly agree) and converted to a score of between 0 and 4, with 4 being the most positive usability rating. Converted scores for all participants are multiplied by 2.5 to give a range of possible total values from 0 to 100. We used 90% CI to allow the benchmarking of the overall DART System Usability Scale score with other studies using this value [46].

[f] N/A: not applicable.

Table 7. System Usability Scale score per group for construct 3 (satisfaction) of the International Organization for Standardization 9241-210-2019 standard.

System Usability Scale[a] score per group	Daily internet users (n=19)	Infrequent internet users (n=3)	ESOL[b] internet users (n=6)	All participants (n=22)
Values, mean (SD)	86.5 (4.48)	70.8 (5.44)	78.1 (4.60)	84.3 (12.73)
Margin of error	1.78 or –1.78	9.17 or –9.17	3.79 or –3.79	4.67 or –3.79

[a] Responses were scored on a 5-point Likert scale (1=strongly disagree and 5=strongly agree) and converted to a score of between 0 and 4, with 4 being the most positive usability rating. Converted scores for all participants are multiplied by 2.5 to give a range of possible total values from 0 to 100. We used 90% CI to allow the benchmarking of the overall Digital Assessment Routing Tool System Usability Scale score with other studies using this value [46].

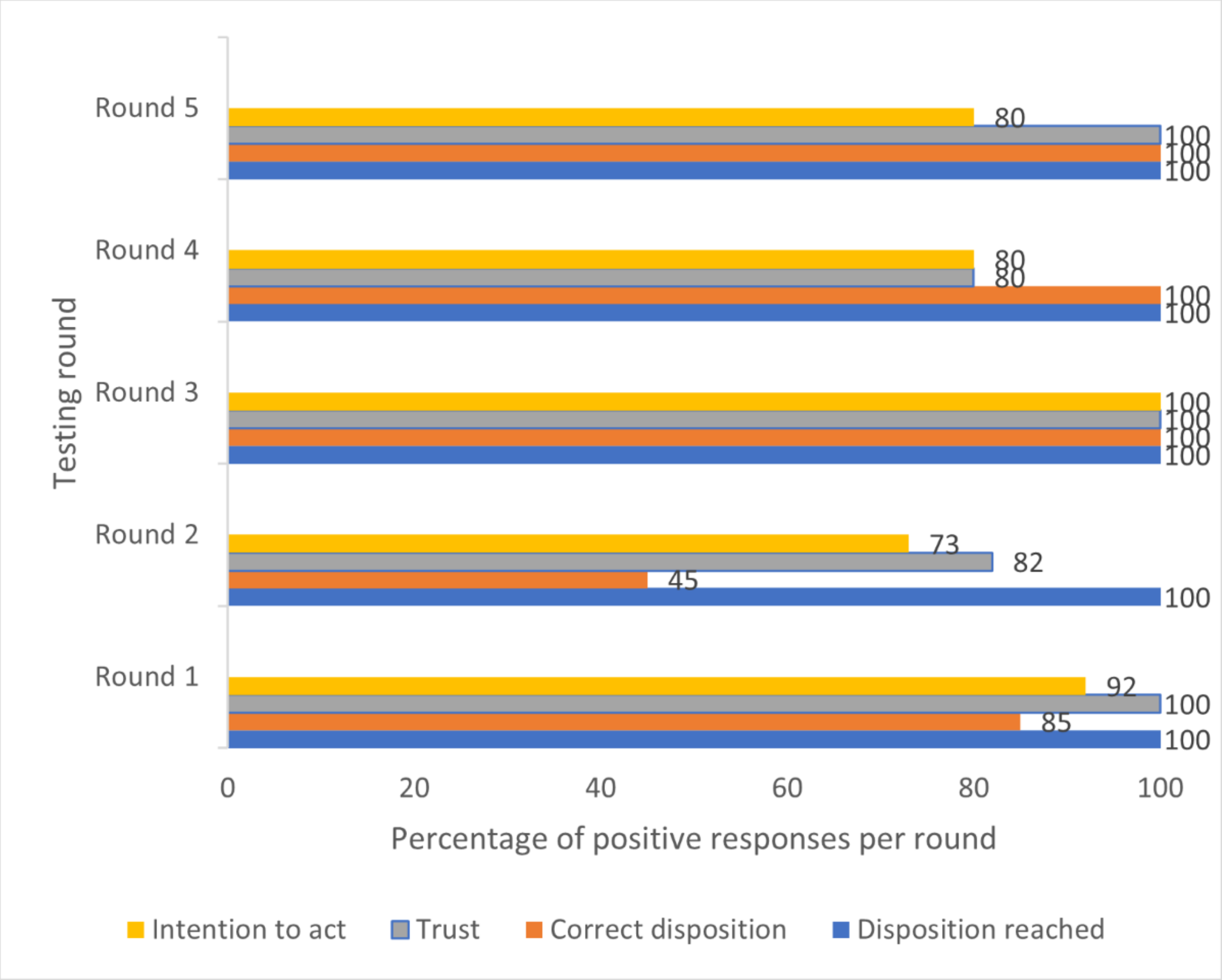
[b] ESOL: English for speakers of other languages.

Effectiveness

All assessments resulted in a recommendation. Other measures of effectiveness improved over the DART iterations, culminating in a high degree of efficiency being reached (Figure 10).

Of the 50 assessments, 8 (16%) resulted in an incorrect recommendation being given, equating to a grade 1 usability issue. Qualitative data revealed that the selection of the incorrect body site at the start of the assessment was responsible for one of these errors. A total of 7 inappropriate clinical escalations were triggered by 2 specific screening questions for systemic inflammatory disease (SID) and central nervous system conditions. Both were reviewed against the evidence base, rewritten, and incorporated into iteration 3. Subsequent testing rounds, including inviting the participants who revealed this problem to retest, confirmed that this usability issue was solved. Participants said their trust and willingness to act would increase if all their symptoms are considered, and this could be achieved by adding a text box on the body site page where they could enter information about problems in other body areas. A related theme was participants wanting to personalize their assessment by adding additional information, and DART iteration 3 included the addition of a free text box at the end of each page. This improved both trust and intention to act, with all participants during testing round 5 arriving at a correct recommendation that they would trust, with just one assessment where the participant said they would not act on the recommendation related to their previous experience of their MSD resolving spontaneously. A small number of participants felt that the lack of personalization of the DART recommendation page made them less likely to act on the advice. This was a result of a test version being used for the study, containing a simple generic recommendation rather than the detailed advice and next actions that would be found on a production version. However, the importance of this feedback was noted and will guide the final DART version to be deployed.

Figure 10. Combined measures of effectiveness by testing round. Results are displayed as the percentage of total assessments to allow comparison, as there were different numbers of participants and assessments in each round. The percentage of assessments in each round resulted in a positive response to the following queries: (1) whether a disposition was achieved, (2) whether it was a clinically correct disposition, (3) whether the participant would trust the disposition, and (4) whether the participant would act on the disposition.



Efficiency

Quantitative indicators of efficiency remained high throughout testing, reinforced by qualitative data.

Round 5 had a larger number of ESOL participants, and it was noted the mean time for this group was slightly longer (6 minutes). The longest time (18 minutes) taken to complete an assessment was by an ESOL participant (Figure 11). All participants without exception said that the time taken to complete an assessment was acceptable and that the format of the questions was clear and supported their ability to make decisions easily.

System errors were defined as DART technical errors, such as presenting the user with an error message or the system timing out. No system errors were encountered during any testing rounds.

The number of times a participant moved back in the pathway to review the previous question remained consistent across the first 4 testing rounds, even with the introduction of participants who used the internet less frequently and ESOL users. Backsteps increased to 6 in round 5 and were linked to one of the ESOL participants, who had the lowest recorded eHEALS score (Figure 12). He told us some of the questions were more challenging to answer and required him to reread them.

Figure 11. Efficiency (time to complete an assessment). Time taken for participants to complete assessments. A total of 16 participants completed 34 assessments in total across rounds 2 to 5. Time was not recorded in round 1, as participants were encouraged to use the “think-aloud” technique.

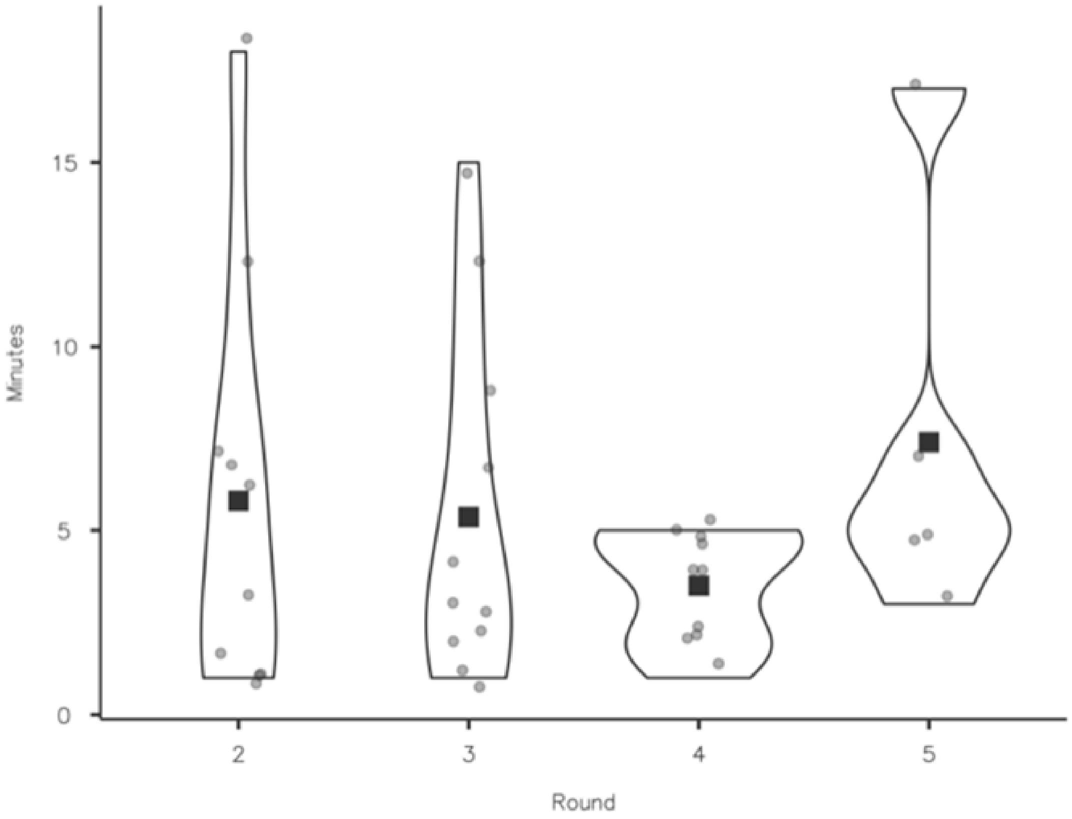
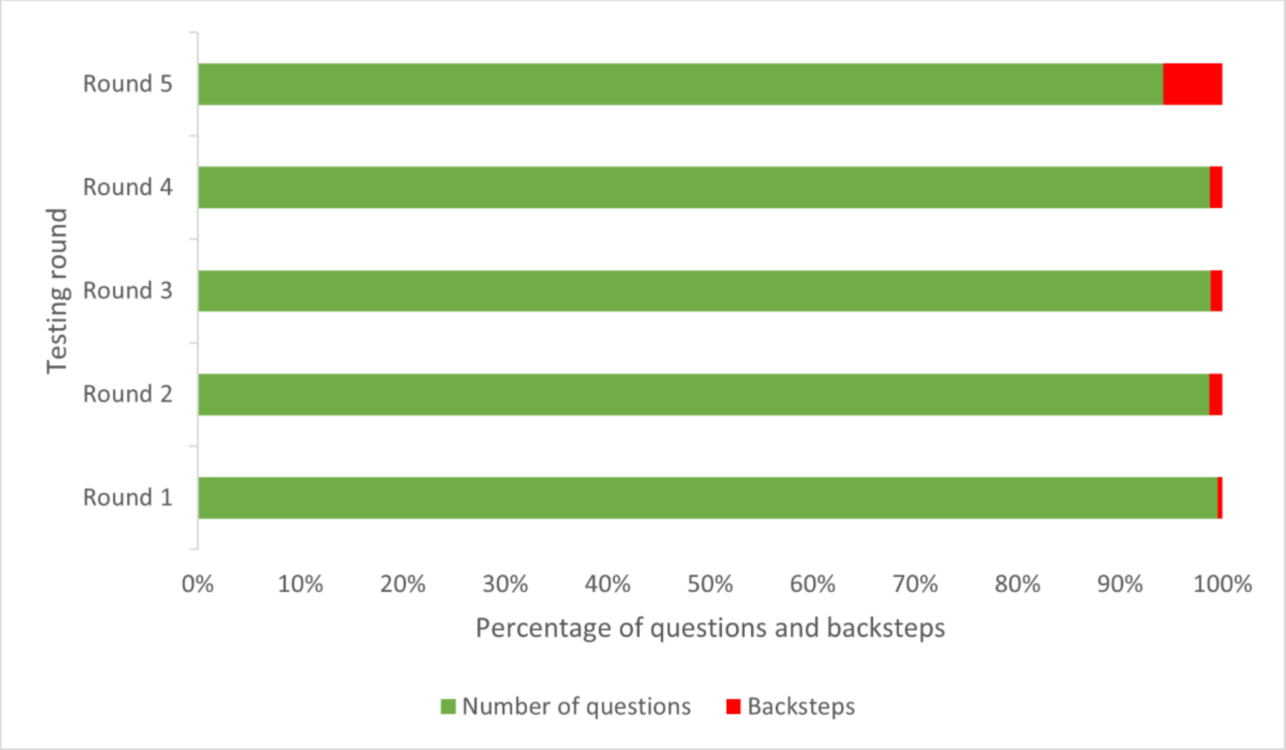


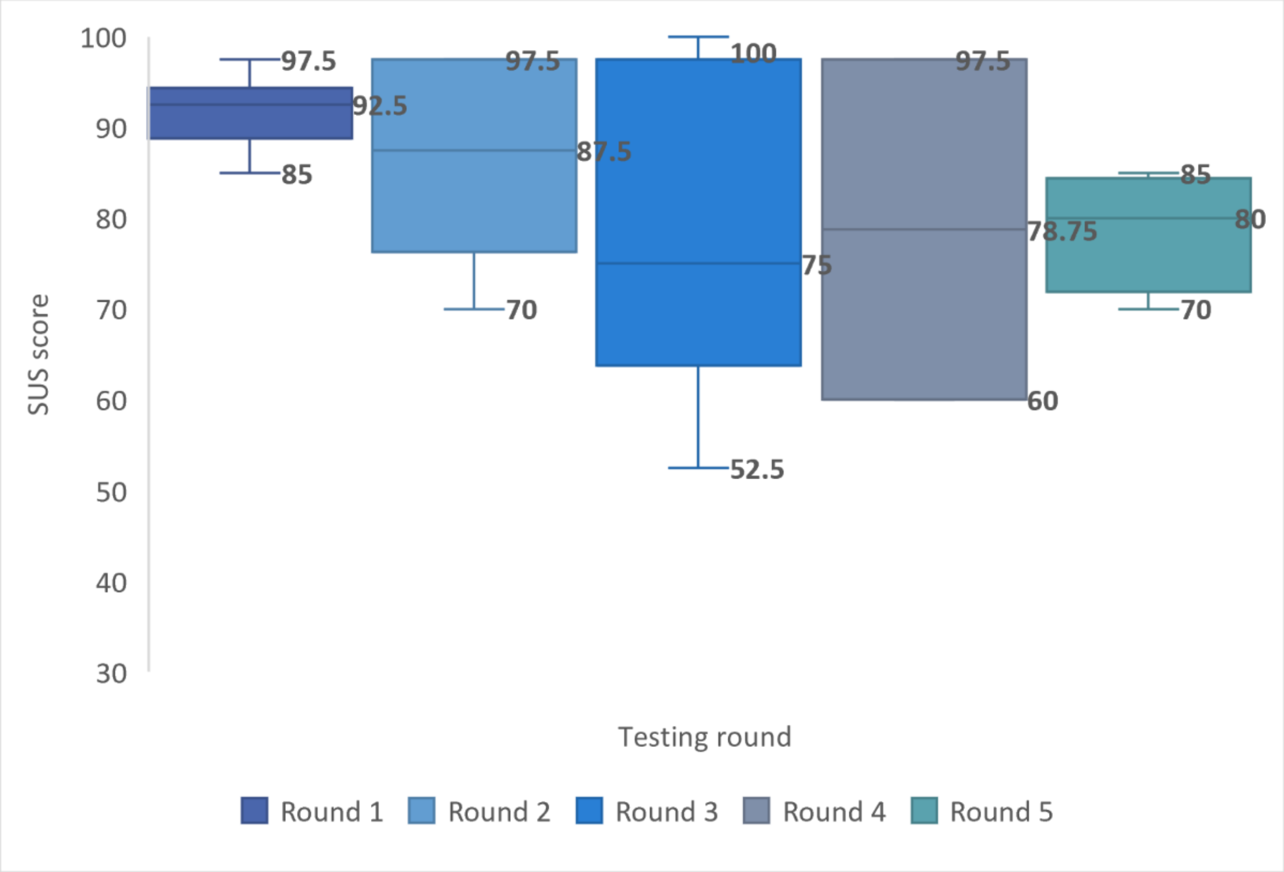
Figure 12. Measures of efficiency (backsteps) by testing round. Number of times a participant moved back a step in the question set to review their previous question and response, with backsteps shown as a percentage of the total number of questions in the assessment.



Satisfaction

Satisfaction was measured quantitively across all groups using the SUS, with qualitative data providing deeper insights into specific question responses. Although high levels of system satisfaction were prevalent throughout testing, cumulative satisfaction scores reduced with each round as infrequent internet users and ESOL participants were recruited (Figure 13). However, the final mean SUS score of 84.3 (SD 12.73; 90% CL 4.67) across all groups achieved the predefined study end point of a score of ≥80, representing a “good” or better system and associated with an increase in the probability that users would recommend DART to a friend [37].

Figure 13. Measure of satisfaction by testing round. SUS: System Usability Scale.



Differences in mean satisfaction scores were present between groups, with daily internet users scoring highest (19/22, 86%; mean 86.5, SD 4.48; 90% CL 1.78 or –1.78), and nonnative English speakers (6/22, 27%; mean 78.1, SD 4.60; 90% CL 3.79 or –3.79) and infrequent internet users scoring the lowest (3/22, 14%; mean 70.8, SD 5.44; 90% CL 9.17 or –9.17), although the CIs overlap.

Although care should be taken in examining individual SUS item responses as external validity only exists on aggregated scores [51], analysis of the highest scoring questions did reveal some useful insights. All participant groups scored highest when saying that they would not need to learn many things before they could use DART. Both daily and infrequent internet users agreed they would not need technical support to use DART, with ESOL participants agreeing to a slightly lesser extent.

All groups did not feel they would use DART frequently or that the functions in the system were well integrated. This was not an unexpected finding as DART is intended for single-time use to determine the correct level of intervention for the user’s MSD and, therefore, would not be used frequently. In contrast to most other mHealth systems, DART is not designed to provide an MSD intervention, with no requirement for the user to navigate to additional features within the system.

Satisfaction adjectives were associated with each participant’s individual total score to aid in explaining the results to non–human factor professionals [52], with 91% (20/22) of participant scores equating DART as a “good,” “excellent,” or “best imaginable system.” The remaining 2 (8%) participants rated DART as “fair,” with none rating it as “poor” or “worst imaginable.”

Using the normalizing process described by Sauro [46], DART ranks within the 96th to 100th percentile (SUS score 84.1-100) of systems tested using the SUS, with an associated adjective rating of “Excellent” [53]. Benchmarking of the DART SUS score against the mean score of 67 (SD 13.4) from 174 studies assessing the usability of public-facing websites utilizing the SUS revealed that DART was among the highest scoring systems assessed in this way [46].

Accessibility

Accessibility has been central to the design of the DART user interface, and the Appian platform on which DART is constructed includes features supporting accessibility for a wide range of users, including those with disabilities who use assistive technologies such as screen readers.

DART's simplicity of use was recognized in the qualitative data by all participants from all groups, who felt that it was sufficiently simple to use and liked the fact that all assessments resulted in a disposition being given. This supports the theory of a low barrier to entry for DART, provided users have internet access:

> *It is so simple, its one most simple of the things, websites, I've engaged with.* [DART016]
>
> *It was very quick. And I quite like that it has one thing for one page, which is a very short question, it gives you a few options, and then you answer so you don't have to go through long text questions, one after the other. So, it just takes you very quickly step by step. And it's quite, I don't know, for me, it was super easy and clear to answer questions.* [DART017]

We asked participants whether they could think of anyone who may not be able to use DART:

> *I wonder how somebody like my mom's age would cope with it and I actually thought there wouldn't be too many who wouldn't.*

One of the ESOL participants tested DART with the help of her daughter, which she told us was her normal practice when she needed to use the internet and common practice within her community:

> *No it's easy to do. My daughter is helping me. Little bit, I understand most of the things, but little bit some questions, what can I say? So, my daughter guides me and help me.* [DART012]

When asked how other ESOL users would use DART, she said the following:

> *I think they need somebody's help, their partner or their children, somebody, or some of their friends, some can help them and then they can do it.*

Overall, participants felt that DART would be easily accessible to people who are familiar with using the internet but that a telephone-delivered alternative would be required for some users. Additional benefits of reducing the time to receiving a diagnosis or treatment and financial savings were also mentioned:

> *If I had this actual system, I would have saved £150 in cash and probably three months of pain had I been able to access it when I had my problems with my back.* [DART013]

Discussion

Principal Findings

The use of the iterative, convergent mixed methods design proved effective; rich data provided objective measures of system performance together with identification of serious usability problems and solutions by real-world users. The results from this study indicate that through a series of iterations, DART usability reached a sufficiently high standard to proceed to further safety and effectiveness trials.

Theme 1: Factors Leading to an Incorrect Recommendation Being Given

Selection of the appropriate body site is crucial to driving matched clinical algorithms within DART, and failure to do so accounted for 8 incorrect recommendations. In addition, participant confidence in DART being able to recognize their body site selection was considered important to most participants, being related to their wider trust in the system and associated intention to act on the recommendation they received. As a result, the body site diagram was refined across all iterations.

A specific DART question designed to screen for SID triggered false-positive recommendations for participants describing mechanical pain. Correct identification of SID can be problematic for primary care clinicians because of varied symptom presentation and overlap with more common osteoarthritic joint conditions [53-55]. A study using patient-entered responses showed that osteoarthritis was diagnosed in 38% of SID cases [56], a result that is clinically significant, given the interdependency of early recognition of SID, minimizing a poor patient outcome [57,58]. During DART testing, these participants prioritized the presence of pain characteristics of osteoarthritis over the hot and swollen multiple joint symptoms presented in SID. This problem was addressed by a detailed review of the literature on the differential diagnosis of SID, rewriting, and inclusion of new questions within all algorithms with associated linked age logic. Subsequent testing rounds, including participants who revealed this problem, confirmed success in negating this usability issue. The implications of creating false positives are often underestimated, with most symptom checker development taking a conservative approach, resulting in systems typically being more risk averse than health care practitioners [59-61]. However, the failure to provide an accurate routing decision can affect user trust and system adoption [62], as well as the creation of unnecessary referrals to urgent or emergency services.

Theme 2: Impact of Assessment Personalization on Participant Trust and Intention to Act on the Given Recommendation

Personalization of assessments was perceived by some participants as a key advantage of a patient-clinician triage interaction over an mHealth system. It has been shown that lack of personalization can affect trust and intention to act, with implications for system adoption [63]. Our participants wanted to "tell their story," not feeling involved in the assessment process unless they could provide information personal to them, including entering details of secondary body site areas. It has been estimated that 8% of patients with MSD presenting to a primary care physician have a problem in >1 body site [64]; thus, an additional comment box was added to the body site diagram page, inviting users to enter other problem areas, something our participants said would address this problem.

In all but 1 assessment, participants who said they would trust DART also said that they would act on the recommendation they were given. A direct correlate of eHealth user trust is information quality, defined by knowledgeable, impartial, and expert sources. These are important factors that lead users to believe the system is acting in their best interest, as they feel a clinician would do [65].

Trust factors have a significant direct effect on user intention to act [66], a key requirement for successful DART introduction into an MSD digital pathway and system adoption. However, we found other factors may be at play, such as previous experience in MSD management. One of the participants told us that they would trust the DART recommendation for physiotherapy but would have waited to see whether their problem improved without treatment, as it had before. It has been recognized that when users of eHealth make decisions on system trust and intention to act, especially those with high eHealth literacy, they will often corroborate information using other web-based content to “triangulate” advice, particularly if the primary source is not familiar to them [63,66]. Interestingly, it has been shown that eHealth users in the United Kingdom with access to free NHS health care are less likely to use health corroboration than users in other countries with private health systems [66]. The NHS website is considered a trusted source of information for many citizens in the United Kingdom, and deploying DART within an NHS pathway may enhance a user’s trust in the given recommendations.

Qualitative data revealed a usability problem that was not considered during development—that serious condition-screening questions on DART had the potential to cause user anxiety in some individuals, who otherwise would not have considered the potential for their problem to be serious. On the basis of this feedback, we placed these less frequently occurring conditions in the context of their incidence to allay unnecessary user anxiety. This is an important consideration for developers of mHealth triage systems, as although the creation of false positives is well recognized and largely accepted as a prudent conservative approach to medical risk management [62], it is also suggested that a significant proportion of potential users would reject using a symptom checker for fear of receiving a wrong diagnosis or an assessment that could cause them anxiety [60].

Theme 3: Impact of User Interpretation When Answering Specific DART Questions

A small number of questions provoked some unexpected participant responses attributable to individual interpretation, likely influenced by their personal experiences. This theme was not identified during previous validation work completed by the panel of expert clinicians. It was only possible to reveal and understand this important usability factor by using a convergent mixed methods design with real-world users, reinforcing the advantage of this methodology over the common practice of using vignettes constructed or delivered by clinicians. An expert clinician recognizes conditions by virtue of pattern recognition, bypassing the conscious, effortful cognitive requirements demanded of a nonclinician user to interpret questions and make decisions on how to respond [67]. Moreover, clinicians are highly educated and not representative of a real-world system user population, including people with eHealth literacy challenges. This study concluded that diversity of user personal experiences can influence how real-world users respond to questions presented by an mHealth system and, ultimately, the recommendation they receive, thus presenting a challenge to developers. For this reason, it is suggested that clinical testing of mHealth systems using vignettes is best used as a precursor to real-world usability testing comprising a representative sample of potential system users.

We found no relationship between age and eHealth literacy, with older participants equally able to arrive at a recommendation as the younger participants. Although this finding should be treated with caution because of the small number of older participants, it could suggest that the perceived ability to seek and use health information is more related to the frequency of internet use rather than age and that differences in eHealth literacy are less likely to be between user group demographics but rather socioeconomic variables between individuals within them [68]. A recent report showed continued growth of internet use in the United Kingdom, with a 6% increase in households with internet access between 2018 and 2020. In the same period, the increase in the number of households with a single adult aged >65 years who accessed the internet within a 3-month period rose from 59% to 80% [39], challenging perceptions about potential mHealth user demographics. An ESOL participant who had assistance from her daughter told us she often sought help from family or neighbors to use the internet, that this was common practice within her community, and that DART could be used effectively in this way. Web-based “surrogate seeking” is now a widespread practice, with significant numbers of internet health information seekers accessing advice on behalf of someone else [69]. However, some studies still link the use of web-based symptom checkers to younger and more highly educated populations [60] and self-referrals for the assessment of musculoskeletal conditions generally [70].

Limitations

Recruitment during the COVID-19 pandemic proved challenging, particularly for people who were not daily users of the internet, as they typically do not engage with social media or advertisements sent via email. All data had to be gathered remotely, affecting the recruitment of people not confident in using web-based video call technology. Although the full recruitment quota was not met for infrequent internet users, this was partially addressed, and the feedback from these participants was particularly valuable in highlighting usability problems. This recruitment challenge could be an indicator of self-selection for DART user adoption related to internet use.

During the DART tests, most participants recalled past conditions that had been resolved or that had been present for some time and had changed since the first onset. At times, this created a problem for participants regarding how to respond to questions; for example, current symptoms versus symptoms they had at the beginning of their problem when they first sought clinical advice. This could be addressed in future studies by only recruiting participants with current problems who had not received medical advice; however, this could potentially exclude participants with chronic MSDs and more complex conditions, thus limiting generalizability.

Although a generic internet system assessment tool, the SUS was chosen as a measure of DART user satisfaction in the absence of a more specific validated mHealth usability measure. As a result, not all the questions were matched to DART in its role as a single-use assessment system with no additional integrated functionality. Other usability assessment tools were considered, including those that measure domains such as loyalty, trust, credibility, and appearance; however, these were designed for the assessment of transactional business systems and included questions inappropriate for DART, such as purchasing and confidence in concluding business [37]. Other tools measured the usability of mHealth systems that support the therapeutic management of conditions over time, with repeated patient use and different integrated functions and features, meaning that the domains assessed were not directly applicable [71].

Future Work

The purpose of this usability study was to optimize usability before proceeding to a trial evaluating the safety and effectiveness of DART against a usual care comparator. A protocol for an initial pilot study has been published and will explore the key aspects of the trial methodology; assess the procedures; and collect exploratory data to inform the design of a definitive, randomized, crossover, noninferiority trial to assess DART safety [36]. DART is currently deployed in a controlled live clinical environment where we use system data, as well as user and clinician feedback, to further refine the algorithms and system usability. A quality improvement study, where DART is integrated into an existing public health service, is also in the design phase.

Conclusions

This study suggests the DART mHealth system has the potential to be offered as an alternative to primary care physician–led or physiotherapist-led triage as part of an MSD pathway. Participants found DART easy to use and would trust and act on the routing recommendation they were given. With all significant usability problems addressed, DART can proceed to the next stage of validation—a randomized controlled trial to assess the safety and effectiveness against a usual care comparator. The inclusion of real-world participants revealed important usability problems and solutions that were not identified during the development or expert panel review stages and highlights the importance of a more sophisticated approach to mHealth system usability testing. The iterative, convergent mixed methods design proved to be highly effective for system development and evaluation and could provide a blueprint for other researchers of mHealth systems.

Acknowledgments

This study was funded by Optima Health.

Authors' Contributions

CL and DM designed the study. MB collected quantitative data, whereas CL collected qualitative data, and both performed data analysis and interpretation. CL drafted the manuscript. DM and WM reviewed the manuscript and provided the final approval. CL takes responsibility for the integrity of the data analysis.

Conflicts of Interest

Optima Health has developed the Digital Assessment Routing Tool system and is the owner of the associated intellectual property. The principal investigator (CL) is an employee of Optima Health and a PhD research student at Queen Mary University of London.

Multimedia Appendix 1
Visual joint display (researcher version).
[DOCX File , 25 KB - jmir_v24i8e38352_app1.docx]

References

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
2. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1859-1922 [FREE Full text] [doi: 10.1016/S0140-6736(18)32335-3] [Medline: 30415748]
3. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: a systematic analysis of the World Health Organization Burden of Diseases Database. Ann Rheum Dis 2019 Jun;78(6):844-848. [doi: 10.1136/annrheumdis-2019-215142] [Medline: 30987966]

4. Musculoskeletal health. World Health Organization. 2022 Jul 14. URL: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions [accessed 2022-08-22]
5. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
6. The Impact of Musculoskeletal Disorders on Americans – Opportunities for Action. Bone and Joint Initiative USA. 2016. URL: http://www.boneandjointburden.org/docs/BMUSExecutiveSummary2016.pdf) [accessed 2022-08-22]
7. Board assurance prompt: improving the effectiveness of musculoskeletal health for NHS providers. Good Governance Institute. 2019. URL: https://www.good-governance.org.uk/wp-content/uploads/2019/04/MSK-BAP-draft-providers-AW-GW.pdf [accessed 2022-08-22]
8. Ahmed N, Ahmed F, Rajeswaran G, Briggs TR, Gray M. The NHS must achieve better value from musculoskeletal services. Br J Hosp Med (Lond) 2017 Oct 02;78(10):544-545. [doi: 10.12968/hmed.2017.78.10.544] [Medline: 29019731]
9. Musculoskeletal conditions. National Health Service England. 2020. URL: https://www.england.nhs.uk/ourwork/clinical-policy/ltc/our-work-on-long-term-conditions/musculoskeletal/ [accessed 2020-05-31]
10. Equipsme. 2021. URL: https://www.equipsme.com/blog/up-to-four-months-to-see-a-physiotherapist/#:~:text=The%20shortest%20wait%20for%20physiotherapy%20was%20in%20Northern,and%20just%201%20day%20for%20an%20urgent%20appointment [accessed 2022-08-22]
11. Deslauriers S, Déry J, Proulx K, Laliberté M, Desmeules F, Feldman DE, et al. Effects of waiting for outpatient physiotherapy services in persons with musculoskeletal disorders: a systematic review. Disabil Rehabil 2021 Mar;43(5):611-620. [doi: 10.1080/09638288.2019.1639222] [Medline: 31304824]
12. Lewis AK, Harding KE, Snowdon DA, Taylor NF. Reducing wait time from referral to first visit for community outpatient services may contribute to better health outcomes: a systematic review. BMC Health Serv Res 2018 Nov 20;18(1):869 [FREE Full text] [doi: 10.1186/s12913-018-3669-6] [Medline: 30458761]
13. Steel N, Ford JA, Newton JN, Davis AC, Vos T, Naghavi M, et al. Changes in health in the countries of the UK and 150 English Local Authority areas 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018 Nov 03;392(10158):1647-1661 [FREE Full text] [doi: 10.1016/S0140-6736(18)32207-4] [Medline: 30497795]
14. Briggs T. A national review of adult elective orthopaedic services in England: Getting It Right First Time. British Orthopaedic Association. 2015 Mar. URL: https://gettingitrightfirsttime.co.uk/wp-content/uploads/2017/06/GIRFT-National-Report-Mar15-Web.pdf [accessed 2022-08-22]
15. Joseph C, Morrissey D, Abdur-Rahman M, Hussenbux A, Barton C. Musculoskeletal triage: a mixed methods study, integrating systematic review with expert and patient perspectives. Physiotherapy 2014 Dec;100(4):277-289. [doi: 10.1016/j.physio.2014.03.007] [Medline: 25242531]
16. Oakley C, Shacklady C. The clinical effectiveness of the extended-scope physiotherapist role in musculoskeletal triage: a systematic review. Musculoskeletal Care 2015 Dec;13(4):204-221. [doi: 10.1002/msc.1100] [Medline: 25758358]
17. Bornhöft L, Larsson ME, Thorn J. Physiotherapy in Primary Care Triage - the effects on utilization of medical services at primary health care clinics by patients and sub-groups of patients with musculoskeletal disorders: a case-control study. Physiother Theory Pract 2015 Jan;31(1):45-52. [doi: 10.3109/09593985.2014.932035] [Medline: 24988315]
18. Bornhöft L, Thorn J, Svensson M, Nordeman L, Eggertsen R, Larsson ME. More cost-effective management of patients with musculoskeletal disorders in primary care after direct triaging to physiotherapists for initial assessment compared to initial general practitioner assessment. BMC Musculoskelet Disord 2019 May 01;20(1):186 [FREE Full text] [doi: 10.1186/s12891-019-2553-9] [Medline: 31043169]
19. Transforming musculoskeletal and orthopaedic elective care services: a handbook for local health and care systems. National Health Service England. 2017. URL: https://www.england.nhs.uk/wp-content/uploads/2017/11/msk-orthopaedic-elective-care-handbook-v2.pdf [accessed 2022-08-22]
20. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. Geneva, Switzerland: World Health Organization; 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2022-08-22]
21. Personalised Health and Care 2020 – Using Data and Technology to Transform Outcomes for Patients and Citizens: A Framework for Action. National Information Board. 2014 Nov. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf [accessed 2022-08-22]
22. Babatunde OO, Bishop A, Cottrell E, Jordan JL, Corp N, Humphries K, et al. A systematic review and evidence synthesis of non-medical triage, self-referral and direct access services for patients with musculoskeletal pain. PLoS One 2020 Jul 6;15(7):e0235364 [FREE Full text] [doi: 10.1371/journal.pone.0235364] [Medline: 32628696]
23. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does not decrease spending. Health Aff (Millwood) 2017 Mar 01;36(3):485-491. [doi: 10.1377/hlthaff.2016.1130] [Medline: 28264950]
24. 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]

25. Jones RB, Ashurst EJ, Trappes-Lomax T. Searching for a sustainable process of service user and health professional online discussions to facilitate the implementation of e-health. Health Informatics J 2016 Dec;22(4):948-961 [FREE Full text] [doi: 10.1177/1460458215599024] [Medline: 26324052]
26. 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]
27. 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]
28. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2019 Mar. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2022-08-22]
29. Guidance: a guide to good practice for digital and data-driven health technologies. Department of Health and Social Care, United Kingdom Government. 2021 Jan 19. URL: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology [accessed 2022-08-22]
30. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems (ISO 9241-210:2019). International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2022-08-22]
31. Information technology — Development of user interface accessibility — Part 1: Code of practice for creating accessible ICT products and services (ISO/IEC 30071-1:2019). International Organization for Standardization. 2019. URL: https://www.iso.org/standard/70913.html [accessed 2022-08-22]
32. Green Paper on mobile health ("mHealth") – SWD135 final. European Commission. 2014. URL: https://ec.europa.eu/newsroom/dae/redirection/document/5147 [accessed 2022-08-22]
33. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
34. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e11656 [FREE Full text] [doi: 10.2196/11656] [Medline: 31025951]
35. Lowe C, Hanuman Sing H, Browne M, Alwashmi MF, Marsh W, Morrissey D. Usability testing of a digital assessment routing tool: protocol for an iterative convergent mixed methods study. JMIR Res Protoc 2021 May 18;10(5):e27205 [FREE Full text] [doi: 10.2196/27205] [Medline: 34003135]
36. Lowe C, Hanuman Sing H, Marsh W, Morrissey D. Validation of a musculoskeletal digital assessment routing tool: protocol for a pilot randomized crossover noninferiority trial. JMIR Res Protoc 2021 Dec 13;10(12):e31541 [FREE Full text] [doi: 10.2196/31541] [Medline: 34898461]
37. Sauro J, Lewis JR. Standardized usability questionnaires. In: Sauro J, Lewis JR, editors. Quantifying the User Experience. 2nd edition. Burlington, MA, USA: Morgan Kaufman; 2016:185-248.
38. Ritchie J, Lewis J, McNaughton Nicholls CM, Ormston R. Qualitative Research Practice: A Guide for Social Science Students and Researchers. 2nd edition. Thousand Oaks, CA, USA: Sage Publications; 2013.
39. Internet access – households and individuals, Great Britain: 2020. Office for National Statistics. 2020. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2020 [accessed 2021-10-12]
40. 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]
41. Alhadreti O, Mayhew P. Rethinking thinking aloud: a comparison of three think-aloud protocols. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, Canada p. 1-12. [doi: 10.1145/3173574.3173618]
42. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. London, UK: CRC Press; 1996:189-194.
43. 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 [FREE Full text] [doi: 10.1080/10447310802205776]
44. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
45. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
46. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Scotts Valley, CA, USA: CreateSpace; 2012.
47. Joffe H. Thematic analysis. In: Harper D, Thompson AR, editors. Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners. Hoboken, NJ, USA: Wiley Online Library; 2021:209-223. [doi: 10.1002/9781119973249.ch15]
48. Spencer L, Ritchie J, Ormston R, O'Connor W, Barnard M. Analysis: principles and processes. In: Ritchie J, Lewis J, McNaughton Nicholls C, Ormston R, editors. Qualitative Research Practice: A Guide for Social Science Students and Researchers. Thousand Oaks, CA, USA: Sage Publications; 2014:269-295.

49. Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Food and Drug Administration Staff. U.S. Department of Health and Human Services, Food and Drug Administration. 2016 Feb 3. URL: https://www.fda.gov/media/80481/download [accessed 2022-08-22]
50. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. Int J Human Comput Stud 1994 Sep;41(3):385-397 [FREE Full text] [doi: 10.1006/ijhc.1994.1065]
51. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
52. Sauro J, Dumas JS. Comparison of three one-question, post-task usability questionnaires. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI '09; April 4-9, 2009; Boston, MA, USA p. 1599-1608. [doi: 10.1145/1518701.1518946]
53. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
54. Suresh E. Diagnosis of early rheumatoid arthritis: what the non-specialist needs to know. J R Soc Med 2004 Sep;97(9):421-424 [FREE Full text] [doi: 10.1258/jrsm.97.9.421] [Medline: 15340020]
55. Rheumatoid arthritis | Causes, symptoms, treatments. Versus Arthritis. 2022. URL: https://www.versusarthritis.org/media/24663/rheumatoid-arthritis-information-booklet-2022.pdf [accessed 2022-08-22]
56. Diagnosis of rheumatoid arthritis (RA). National Rheumatoid Arthritis Society. 2022. URL: https://nras.org.uk/resource/diagnosis/ [accessed 2022-08-22]
57. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Aug 24;17(1):362 [FREE Full text] [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
58. Monti S, Montecucco C, Bugatti S, Caporali R. Rheumatoid arthritis treatment: the earlier the better to prevent joint damage. RMD Open 2015 Aug 15;1(Suppl 1):e000057 [FREE Full text] [doi: 10.1136/rmdopen-2015-000057] [Medline: 26557378]
59. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. JAMA 2018 Oct 02;320(13):1360-1372. [doi: 10.1001/jama.2018.13103] [Medline: 30285183]
60. Semigran HL, Linder JA, 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]
61. 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]
62. 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;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
63. Using Technology to Ease The Burden on Primary Care. Healthwatch Enfield. 2019. URL: https://healthwatchenfield.co.uk/wp-content/uploads/2019/01/Report_UsingTechnologyToEaseTheBurdenOnPrimaryCare.pdf [accessed 2022-08-22]
64. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in Internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]
65. Arthritis Research UK National Primary Care Centre, Keele University. Consultations for selected diagnoses and regional problems. Musculoskeletal Matters. 2010. URL: https://viewer.joomag.com/musculoskeletal-matters-2-consultations-for-selected-diagnoses-regional/0458165001492524063 [accessed 2022-08-20]
66. Sillence E, Blythe JM, Briggs P, Moss M. A revised model of trust in Internet-based health information and advice: cross-sectional questionnaire study. J Med Internet Res 2019 Nov 11;21(11):e11125 [FREE Full text] [doi: 10.2196/11125] [Medline: 31710297]
67. Gruppen LD. Clinical reasoning: defining it, teaching it, assessing it, studying it. West J Emerg Med 2017 Jan;18(1):4-7 [FREE Full text] [doi: 10.5811/westjem.2016.11.33191] [Medline: 28115999]
68. Berkowsky RW. Exploring predictors of eHealth literacy among older adults: findings from the 2020 CALSPEAKS survey. Gerontol Geriatr Med 2021 Dec 14;7:23337214211064227 [FREE Full text] [doi: 10.1177/23337214211064227] [Medline: 34926723]
69. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
70. Igwesi-Chidobe CN, Bishop A, Humphreys K, Hughes E, Protheroe J, Maddison J, et al. Implementing patient direct access to musculoskeletal physiotherapy in primary care: views of patients, general practitioners, physiotherapists and clinical commissioners in England. Physiotherapy 2021 Jun;111:31-39 [FREE Full text] [doi: 10.1016/j.physio.2020.07.002] [Medline: 33316866]
71. Garcia A. UX Research | Standardized Usability Questionnaire. User Research. 2013 Nov 27. URL: https://arl.human.cornell.edu/linked%20docs/Choosing%20the%20Right%20Usability%20Questionnaire.pdf [accessed 2022-08-22]

Abbreviations

CL: confidence level
DART: Digital Assessment Routing Tool
eHEALS: eHealth Literacy Scale
ESOL: English for speakers of other languages
ISO: International Organization for Standardization
mHealth: mobile health
MSD: musculoskeletal disorder
NHS: National Health Service
SID: systemic inflammatory disease
SUS: System Usability Scale

Edited by R Kukafka; submitted 29.03.22; peer-reviewed by F Burstein, P Worthy; comments to author 30.05.22; revised version received 19.07.22; accepted 29.07.22; published 30.08.22.

Please cite as:
Lowe C, Browne M, Marsh W, Morrissey D
Usability Testing of a Digital Assessment Routing Tool for Musculoskeletal Disorders: Iterative, Convergent Mixed Methods Study
J Med Internet Res 2022;24(8):e38352
URL: https://www.jmir.org/2022/8/e38352
doi:10.2196/38352
PMID:36040787

Original Paper

Usability Evaluation of a Noninvasive Neutropenia Screening Device (PointCheck) for Patients Undergoing Cancer Chemotherapy: Mixed Methods Observational Study

Ganimete Lamaj[1], BSc; Alberto Pablo-Trinidad[1], PhD; Ian Butterworth[1], MSc; Nolan Bell[1], BSc; Ryan Benasutti[1], BSc; Aurelien Bourquard[1], PhD; Alvaro Sanchez-Ferro[1], MD, PhD; Carlos Castro-Gonzalez[1], PhD; Ana Jiménez-Ubieto[2], PhD; Tycho Baumann[2], MD; Antonia Rodriguez-Izquierdo[2], MD; Elizabeth Pottier[3], MM; Anthony Shelton[3], RN, OCN; Joaquin Martinez-Lopez[2], MD, PhD; John Mark Sloan[3], MD

[1]Leuko Labs, Inc, Boston, MA, United States

[2]Hematology Department, Hospital Universitario 12 de Octubre, Madrid, Spain

[3]Section of Hematology & Medical Oncology, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA, United States

Corresponding Author:
Ganimete Lamaj, BSc
Leuko Labs, Inc
8 St Marys Street
No 613
Boston, MA, 02215
United States
Phone: 1 781 954 0250
Email: ganimete@leuko.com

Abstract

Background: Patients with cancer undergoing cytotoxic chemotherapy face an elevated risk of developing serious infection as a consequence of their treatment, which lowers their white blood cell count and, more specifically, their absolute neutrophil count. This condition is known as neutropenia. Neutropenia accompanied by a fever is referred to as febrile neutropenia, a common side effect of chemotherapy with a high mortality rate. The timely detection of severe neutropenia (<500 absolute neutrophil count/µL) is critical in detecting and managing febrile neutropenia. Current methods rely on blood draws, which limit them to clinical settings and do not allow frequent or portable monitoring. In this study, we demonstrated the usability of PointCheck, a noninvasive device for neutropenia screening, in a simulated home environment without clinical supervision. PointCheck automatically performs microscopy through the skin of the finger to image the blood flowing through superficial microcapillaries and enables the remote monitoring of neutropenia status, without requiring venipuncture.

Objective: This study aimed to evaluate the usability of PointCheck, a noninvasive optical technology for screening severe neutropenia, with the goal of identifying potential user interface, functionality, and design issues from the perspective of untrained users.

Methods: We conducted a multicenter study using quantitative and qualitative approaches to evaluate the usability of PointCheck across 154 untrained participants. We used a mixed method approach to gather usability data through user testing observations, a short-answer qualitative questionnaire, and a standardized quantitative System Usability Scale (SUS) survey to assess perceived usability and satisfaction.

Results: Of the 154 participants, we found that 108 (70.1%) scored above 80.8 on the SUS across all sites, with a mean SUS score of 86.1 across all sites. Furthermore, the SUS results indicated that, out of the 151 users who completed the SUS survey, 145 (96%) found that they learned how to use PointCheck very quickly, and 141 (93.4%) felt very confident when using the device.

Conclusions: We have shown that PointCheck, a novel technology for noninvasive, home-based neutropenia detection, can be safely and effectively operated by first-time users. In a simulated home environment, these users found it easy to use, with a mean SUS score of 86.1, indicating an excellent perception of usability and placing this device within the top tenth percentile of systems evaluated for usability by the SUS.

Trial Registration: ClinicalTrials.gov NCT04448314; https://clinicaltrials.gov/ct2/show/NCT04448314 (Hospital Universitario 12 de Octubre registration) and NCT04448301; https://clinicaltrials.gov/ct2/show/NCT04448301 (Boston Medical Center registration)

(J Med Internet Res 2022;24(8):e37368) doi:10.2196/37368

KEYWORDS

digital health; usability; patient-centered care; remote monitoring; decision support systems; white blood cells; diagnosis; medical device; cancer; chemotherapy; infection; white blood cell; technology

Introduction

Background

One of the most serious side effects of cytotoxic chemotherapy and immunotherapy is neutropenia—a decrease in neutrophils, the most common type of white blood cell (WBC) and the most important cell needed to prevent bacterial infection. The primary clinical consequence of neutropenia is an elevated risk of life-threatening bacterial infection that typically requires immediate admission to the emergency department, hospitalization, and treatment [1-3]. Every year, approximately 850,000 patients with cancer start chemotherapy treatments in the United States [4], and 140,000 (17%) [5] will endure at least one episode of febrile neutropenia (FN), or neutropenia accompanied by a fever. FN typically requires an admission of over 1 week, costing approximately US $30,000 per episode [6,7], with associated mortality rates between 7% to 10% [8]. The timely detection and awareness of severe neutropenia (ie, <500 absolute neutrophil count/µL) [9] can be crucial to prevent and manage FN in the outpatient setting [10,11] and the emergency department [12,13].

In the current standard of care, the risk of FN is evaluated by using a priori scores, such as the Multinational Association for Supportive Care in Cancer score [14], to indicate primary prophylaxis with growth colony stimulating factors or by patients regularly monitoring their temperature at home to seek emergency care when fever ensues [15]. Despite these existing methods, FN still has an important economic and clinical impact in cancer care. The early detection of neutropenia could be used to prevent FN by triggering an early administration of granulocyte colony-stimulating factor or antibiotics [16-19]. Unfortunately, current neutropenia-monitoring options rely on venipunctures in the clinical setting or finger-prick blood samples at the point of care [20]. These technologies either require laboratory infrastructure limited to the hospital setting or are impractical as they cannot be operated by minimally trained users to achieve accurate and reliable results [21-24]. To address this unmet need, this paper presents a usability evaluation of a novel, noninvasive technology that allows automated and frequent neutropenia monitoring by patients from the home setting with minimal training.

Assessing the usability for this kind of technology is crucial in ensuring the accuracy of the results, driving adoption, and improving patient compliance and adherence [25]. According to the International Organization for Standardization (IEC 62366-1:2015), usability is defined as a “characteristic of the user interface that facilitates use and thereby establishes effectiveness, efficiency and user satisfaction in the intended use environment” [26]. These metrics can be measured by gaining insight into patient perspectives regarding user performance, satisfaction, and acceptability while using an intervention [27]. For this study, the standardized System Usability Scale (SUS) survey was chosen as a quantitative method of assessing subjective usability due to evidence that it can be used to assess any technology [28] and has successfully been used in the medical domain to assess home medical devices [29-31].

The early detection of FN risk is essential as it can be associated with a higher chance of survival, more successful treatment, and improved quality of life. Therefore, the need for these technologies to be user friendly to the majority of the patient population subsequently increases, as this can impact the patients’ perception of the technology and their decision to take the test [32]. Additionally, technology-based solutions such as the one presented in this paper can help strengthen the relationship and communication between patients and their doctors, empower the patients’ well-being, and help doctors make better and more informed decisions [33].

Study Objectives

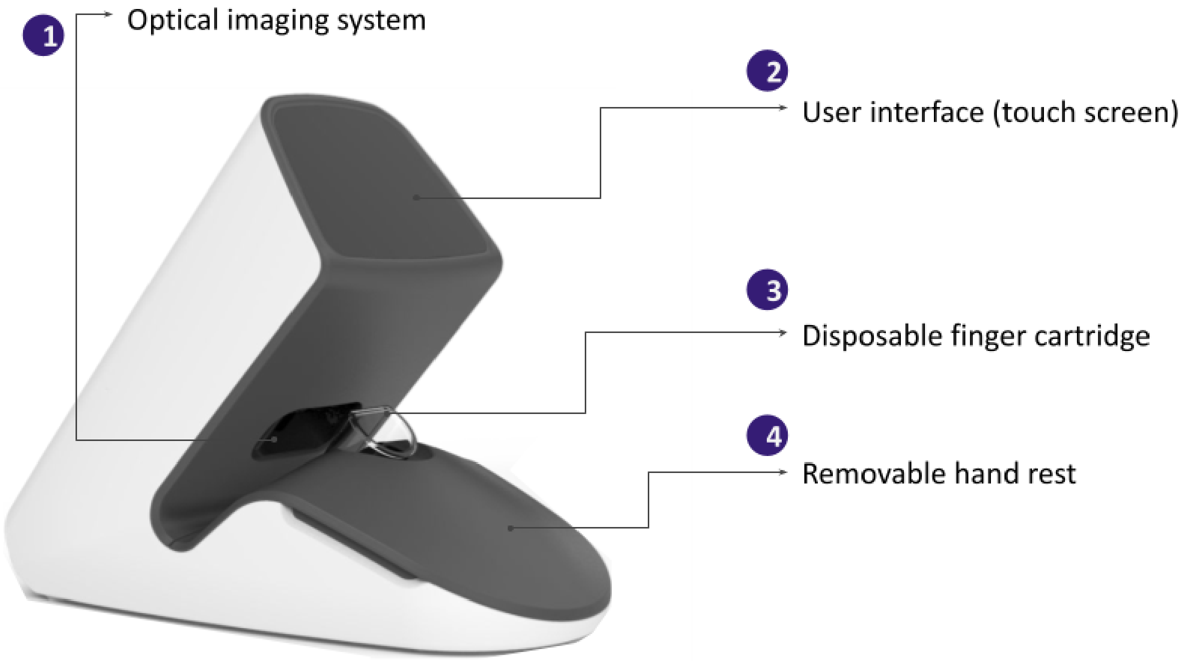
We hypothesized that novice users will consider PointCheck (Leuko Labs), the first noninvasive optical technology for screening severe neutropenia, to be easy to use. The primary study objective was to evaluate the usability of PointCheck with the goal of identifying potential user interface (UI), functionality, and design issues from the perspective of untrained, first-time users in a simulated home environment. The primary end point for the study, defined a priori, was a group mean score of 80.8 on a standardized SUS, indicating a favorable perception of usability and a higher likelihood of adoption.

Methods

Device Description

PointCheck is the first noninvasive device (Figure 1) designed to screen for severe neutropenia in the home setting [34]. By imaging the blood flowing through the capillaries in the finger, PointCheck enables real-time remote monitoring of WBC levels based on optical imaging and without a blood draw [35,36].

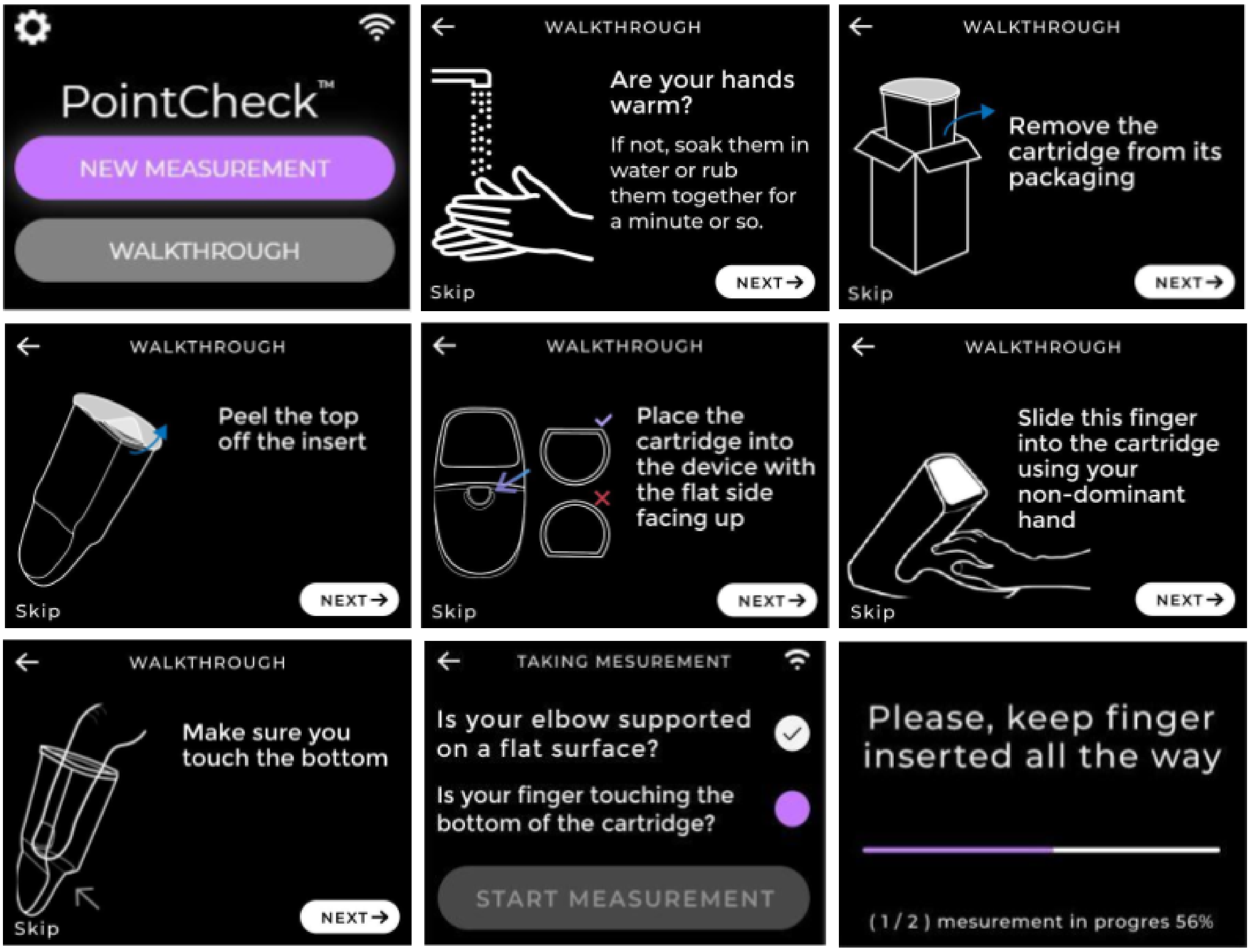
Figure 1. The PointCheck device and its main components.



The device consists of an optics and illumination system, on-board computing electronics, an 8.9-cm touch screen UI, a power cord, and disposable finger cartridges (Figure 1). It uses a camera microscopy system and LEDs to image capillaries in the nailfold region of the finger—typically the nondominant 4th (ring) finger, which has been shown by previous literature to contain the most intact and visible capillaries when compared to other fingers [37]. The finger cartridge is a disposable component that is prefilled with mineral oil and allows for effective optical refractive index coupling to ensure transdermal imaging quality [38,39]. The finger cartridge is designed for 1-time use. The hardware system design resembles the methods used in standard nailfold video capillaroscopy, which is an established technique used by rheumatologists to evaluate capillary morphology and microcirculation [40].

The UI on the touch screen provides a guided walk-through to facilitate the correct use of the device. It prompts the user to warm their hands; open up a new, unused cartridge; properly place the cartridge into the device; and insert their nondominant 4th (ring) finger all the way into the cartridge while properly supporting their arm on a flat, stable surface (Figure 2). A final checklist ensures that the most critical steps have been completed and the user is able to start the 1-minute measurement. The version of the device used in this study was a beta prototype (version 4).

Figure 2. Screenshots of PointCheck’s user interface depicting the user walk-through tutorial in English for taking a measurement and device function via the touch screen interface. Language support for Spanish- and Haitian-speaking populations was implemented to translate the instructions.



Are your hands warm?

If not, soak them in water or rub them together for a minute or so.

Remove the cartridge from its packaging

Peel the top off the insert

Place the cartridge into the device with the flat side facing up

Slide this finger into the cartridge using your non-dominant hand

Make sure you touch the bottom

Is your elbow supported on a flat surface?

Is your finger touching the bottom of the cartridge?

Please, keep finger inserted all the way

(1 / 2) mesurement in progres 56%

Participants and Setting

Usability data was gathered from a cohort including both healthy volunteers and outpatients with cancer receiving chemotherapy. Patients were recruited at both the Boston Medical Center and Hospital Universitario 12 de Octubre before their routine chemotherapy administration. The healthy volunteers were recruited at the Massachusetts Institute of Technology’s Center for Clinical and Translational Research via advertisements displayed on Massachusetts Institute of Technology’s campus and social media and via email lists. The study visits took place in a simulated home environment, and testing was conducted without supervision from a medical professional. No participants had prior experience with the tested device.

A total of 154 participants (85 patients and 69 volunteers) participated in this study. According to standard usability sample size models, this sample size provides a 99% chance of detecting errors with the probability of occurrence of 3% at least once [41].

To ensure the generalizability of the results, we included younger (aged <65 years) and older (aged ≥65 years) adults, patients with diverse cancer types (lymphoma, leukemia, and myeloma, among other tumor types), men and women, and different education levels (≥8th grade or <8th grade). This allowed us to better understand the links between certain characteristics of the potential patients (ie, age, education, technophilia, and health literacy) and the usability [42,43].

Ethics Approval

Institutional review board (IRB) approvals were obtained from the Boston Medical Center IRB (H-39964), Hospital Universitario 12 de Octubre IRB (20/049), and the New England IRB (1290027) to conduct the study prior to recruitment. Participants provided written consent before agreeing to participate in the study according to good clinical practice guidelines (ICH E6:R2) [44].

Study Design

We used a mixed method approach to gather usability data through (1) user observation, (2) a short-answer qualitative e-questionnaire, and (3) a standardized quantitative SUS to assess perceived usability and satisfaction.

Regarding user observation, study coordinators observed participants while they used the device to document any errors that could potentially lead to imaging errors on the device. For example, an unsupported arm or incorrect hand placement could result in too much movement during a measurement and cause an error in the reading. Study coordinators also observed participants to identify and document any points of confusion during the walk-through steps that could be improved. All documented observations were collated into a list to be manually categorized by the type and frequency of occurrence (see Qualitative Results).

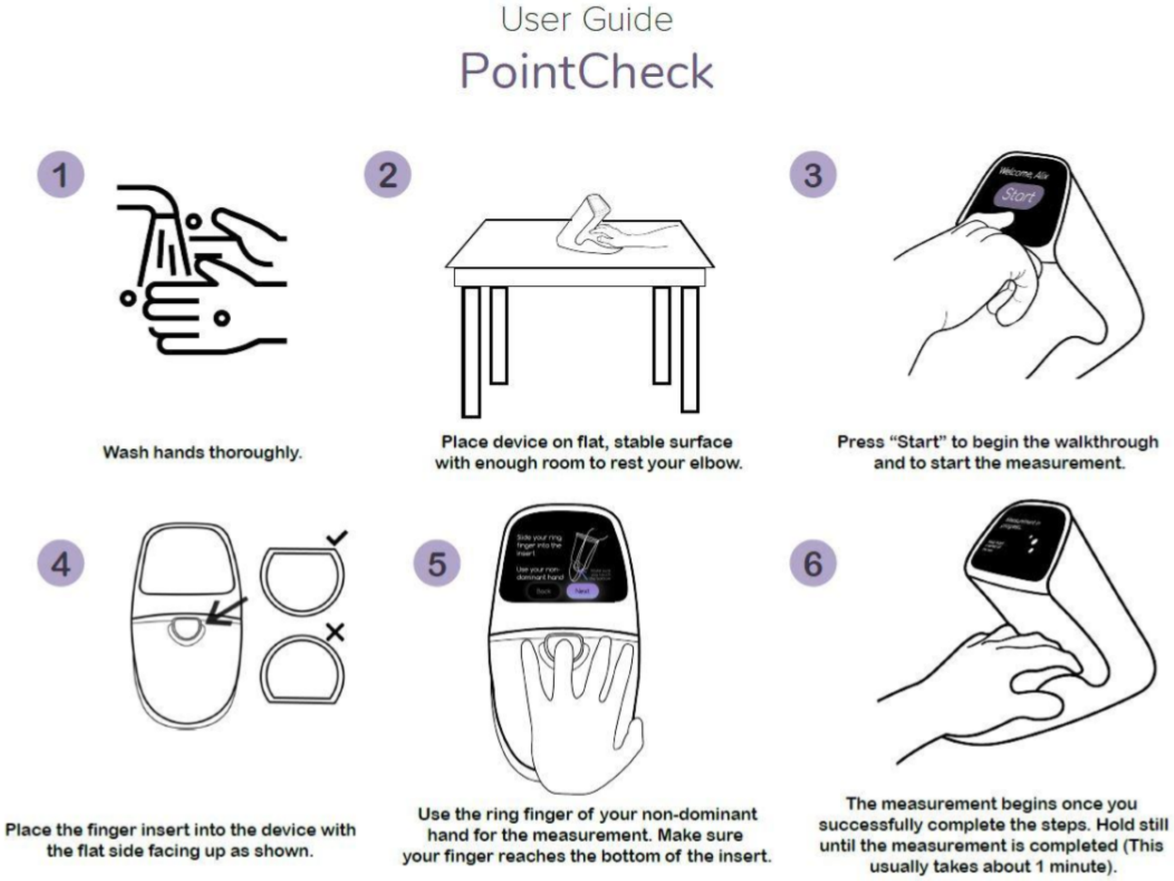
A subset (n=120) of the participants were given the opportunity to give feedback and document their thoughts, feelings, and experience using the device through an e-questionnaire containing 4 questions (Multimedia Appendix 1). We used this questionnaire to assess any potential confusion or difficulties participants may have had using the device or the UI, their attitude toward the product, and any potential features they would like to see added to improve user friendliness. The feedback from the questionnaires was collated into a spreadsheet to be manually categorized into themes (see Qualitative Results).

Finally, the SUS survey was used as a method of assessing subjective usability. The SUS is a Likert-type questionnaire comprising 10 questions with 5 response options ranging from “strongly disagree” to “strongly agree,” allowing for a subjective assessment of usability [45]. Scores range from 0 to 100 with higher scores indicating favorable user perceptions of the device and lower scores indicating low usability. The success criterion for a favorable evaluation in the SUS was defined by a mean group score of greater than 80.8 (see Quantitative Results). This threshold was selected based on previously published cutoffs to define the promoters of a technology [46].

Usability Testing Procedure

Baseline assessments conducted by research staff included a brief physical examination and collection of demographic information. Study coordinators read a short script that provided the participants with information about how to use the device and emphasized that the aim of the study was to test the user friendliness of the device and not the participants’ ability to use the device correctly (Multimedia Appendix 2). In addition, the participants were provided with a 1-page guide containing device instructions before attempting to take a measurement on their own (Figure 3).

Figure 3. One-page quick start guide provided to participants before attempting to take a measurement on their own.



Participants were then asked to follow the instructions presented to them on the device screen, guiding them through the critical steps required to obtain high-quality measurements. The study coordinators did not intervene or answer questions related to device use to reproduce the conditions of unsupervised home use. A second observer monitored and recorded a subset of visits either in person or through the Zoom teleconferencing platform (Zoom Video Communications). Observers documented participant errors, feedback, and tendencies. After completing the initial measurement, participants were immediately asked to complete the SUS and questionnaire to evaluate their first impressions about the user friendliness of the system to prevent any bias introduced from repeating measurements and becoming familiar with the measurement process. Participants performed additional trials, each lasting about 1.5 minutes (1-minute measurement plus 30-second setup and walk-through) for a total of 2 to 6 repeat measurements to evaluate the device precision. These subsequent trials were not used for the perceived usability evaluation and are not reported here.

Data Analysis

Basic demographic characteristics were summarized using descriptive statistics. A final SUS score was computed in accordance with Brooke [47], and responses to the e-questionnaire were tokenized for their content and categorized into themes for qualitative analysis. Statistical comparisons were made between the different group categories stratified by age and literacy to evaluate usability differences using nonparametric techniques (Mann-Whitney *U* test). All quantitative data were processed using RStudio (version 1.3.1093; RStudio Team) [48].

Results

User Statistics

Table 1 represents the breakdown of study participant characteristics by age, education level, gender, and race. Of the 154 participants, 118 (76.6%) were aged <65 years, with an average age of 44.8 (range 18-88) years. A majority (n=102, 66.2%) of the participants had an education level exceeding 8th grade, whereas 43 (27.9%) participants had an education level below 8th grade, and 9 (5.8%) did not provide educational level information.

Table 1. Basic demographics of participants. Educational, race, and ethnicity level data were missing for 9 (5.8%), 11 (7.1%) and 10 (6.5%) out of 154 participants, respectively.

Demographic	Participants (N=154)
Age (years)	
Mean (SD)	44.8 (20.5)
Median (range)	38.3 (18.0-88.5)
Gender, n (%)	
Male	67 (43.5)
Female	87 (56.5)
Educational level, n (%)	
<8th grade	43 (27.9)
≥8th grade	102 (66.2)
Missing	9 (5.8)
Race, n (%)	
American Indian or Alaska Native	1 (0.6)
Asian	28 (18.2)
Black or African American	28 (18.2)
More than 1 race	5 (3.2)
Unknown	11 (7.1)
White	81 (52.6)
Ethnicity, n (%)	
Hispanic or Latino	49 (31.8)
Not Hispanic nor Latino	95 (61.7)
Unknown	10 (6.5)

Quantitative Results

The average SUS score across all participants was 86.1. In total, 70.1% (108/154) of the participants scored above the goal of 80.8 (Tables 2 and 3), which indicated that they would be early promoters and more likely to recommend the device to a friend [49]. When stratifying the SUS results by education level, we found that participants exceeding the 8th grade level scored slightly higher than those with an 8th grade level education and below—but only by a margin of 2.5 points, which was not found to be statistically significant (*P*=.27; Table 4). When stratifying SUS results by age categories, we found that participants aged <65 years also scored higher than participants aged ≥65 years by a margin of 3.4 points, showing a nonstatistically significant trend (*P*=.06). Both groups had a mean score above the predefined threshold of 80.8 (Table 5). When evaluating the SUS results by the individual survey questions, we found that 96% (145/151) of the participants that completed the survey found PointCheck easy to use and felt that they could learn to use it very quickly (Figure 4).

Table 2. Quantitative System Usability Scale (SUS) results across all participants. SUS surveys were incomplete for 3 participants and could not be computed.

SUS score (range 0-100)	Overall (N=154)
Mean (SD)	86.1 (12.2)
Median (range)	87.5 (20.0-100)
Missing, n (%)	3 (1.9)

Table 3. Total percent of promoters (defined as participants scoring >80.8 points on the System Usability Scale). System Usability Scale surveys were incomplete for 3 participants and could not be computed.

Cutoff used (points)	Overall (N=154), n (%)
≤80.8	43 (27.9)
>80.8	108 (70.1)
Missing	3 (1.9)

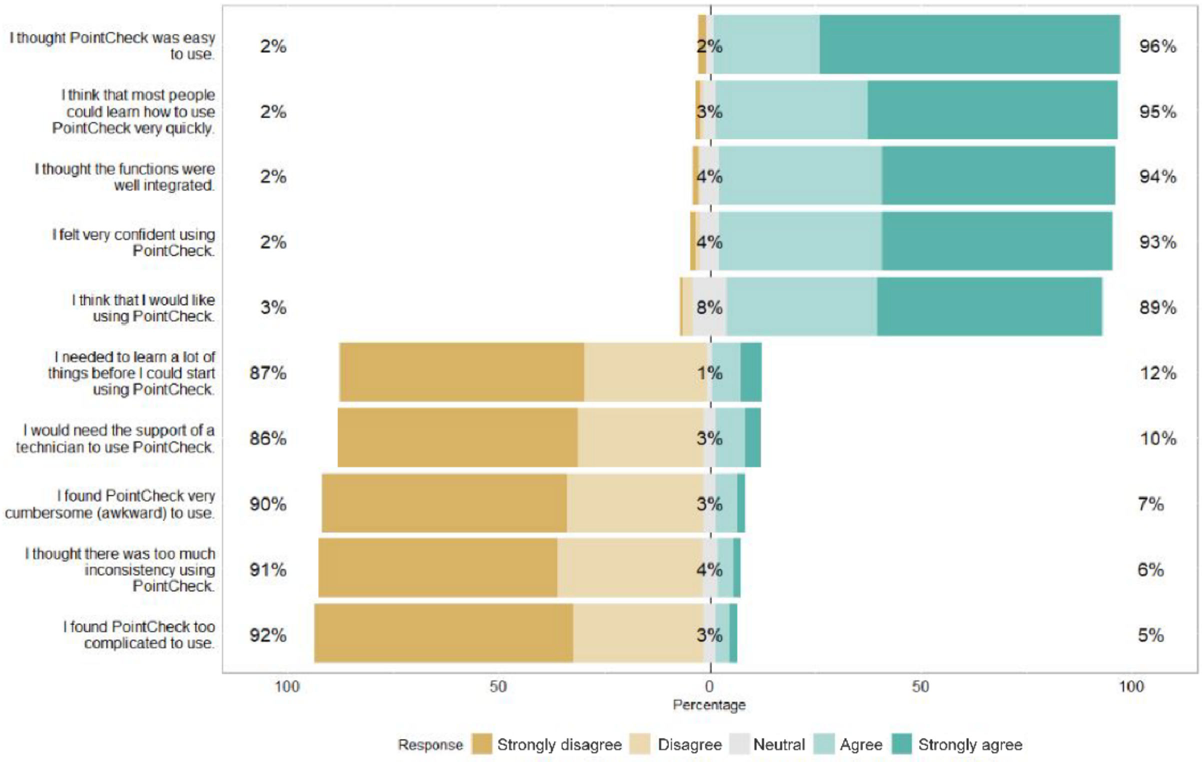
Table 4. System Usability Scale (SUS) results stratified by educational level (<8th grade and ≥8th grade). The results were generated using the data available from a total of 145 participants. Educational level data was missing for 9 participants (N=154, 5.8%).

SUS score (range 0-100)	<8th Grade (N=43)	≥8th Grade (N=102)	Overall (N=154)
Mean (SD)	84.0 (12.7)	86.5 (12.2)	86.1 (12.2)
Median (range)	85.0 (57.5-100)	87.5 (20.0-100)	87.5 (20.0-100)

Table 5. System Usability Scale (SUS) results stratified by age category (<65 years and ≥65 years).

SUS score (range 0-100)	<65 years (N=118)	≥65 years (N=36)	Overall (N=154)
Mean (SD)	86.9 (12.5)	83.5 (11.0)	86.1 (12.2)
Median (range)	90.0 (20.0-100])	82.5 (57.5-100)	87.5 (20.0-100)

Figure 4. SUS survey responses assessed individually. Percentage values are calculated using available data from a total of 151 participants. SUS surveys were incomplete for 3 participants. SUS: System Usability Scale.



Qualitative Results

User Observation

Error observation notes were collated into a list and then manually categorized by the type and frequency of occurrence. The primary error sources included skipping or misreading instructions and on-screen instruction accessibility. The majority (70/86, 81%) of these errors occurred only on the first use and were shown to be correctable by interventional guidance. Such guidance was given after the SUS survey had been completed, and improvement in most cases was demonstrated in subsequent trials. This shows that although the device performs well in independent use, the monitoring of first use by an experienced operator may have further benefit for catching usability errors.

e-Questionnaire Feedback

The feedback from the questionnaires was collated into a spreadsheet and then manually categorized into the following themes: UI/user experience, aesthetic design/logical design, hand rest, cartridge, cleaning/sanitation, and software/bugs. The themes were then broken down into the following subthemes: confidence in use, training effectiveness, UI design/clarity of UI instructions, ergonomic design, foreseeable home use issue, and accessibility. The instances of feedback falling within these subthemes were counted and generated the 3 overarching themes: pretraining effectiveness, user friendliness of PointCheck (related to ease of use, accessibility, and clarity of UI elements), and ergonomic design.

Pretraining Effectiveness

A portion of participants initially expressed some uncertainty when using the device for the first time (Participants #38 and #18; Table 6). With repeated use, however, most participants felt that they could catch on quickly (Participant #40; Table 6).

Table 6. Illustrative quotes for the 3 overarching themes.

Theme/category	Illustrative quote
Pretraining effectiveness	• "Would like a YouTube channel/clip to watch in advance that will explain the device." (Participant #38) • "The most difficult step was probably removing the cartridge from the box. I was not sure if I had to keep the cartridge clean for measurements and the instructions did not tell how I should be holding the cartridge or if I should even be careful or not about touching it too much and getting it dirty" (Participant #18) • "The device was cumbersome (awkward) to use because it was the first time. After the first time, it would be easier to use." (Participant #40)
User friendliness of PointCheck	• "Others can't see as well, may need others to help them if they have dementia or are forgetful." (Participant #53) • "Could not read font of the three step instructions." (Participant #65) • "If I were to use the device on a daily basis, I would be relatively annoyed by the fact the three repeat questions are timed lag to press yes." (Participant #15) • "One thing I was confused about was checking off 'was my finger in all the way.' It was just a circle and I was confused what I was supposed to do on this step." (Participant #21) • "The cartridge lid is shown to be peeled from the flat side (facing down) but then needs to be rotated to be put in with the flat side up. If the peel could be opened in the right orientation, that would have helped." (Participant #24) • "Very crisp, and clear how to start using it...the screen has good contrast, good font choice given resolution and size, and the purple/grey color scheme is also calming." (Participant #30) • "Even though I knew beforehand that I should rest my elbow, I forgot to do so after inserting my finger because that was my main focus, so I really appreciated the reminder to have my elbow rested right after the step where I inserted my ring finger." (Participant #34)
Ergonomic Design	• "The display seems angled a little high, considering that the machine needs to be placed a considerable distance to rest my arm. The steps were fairly intuitive." (Participant #3) • "I was expecting the device to be a bit smaller. I think the device is designed for a very big hand, I think it would be probably better to try to make it the size closer to a computer mouse." (Participant #16) • "My fingers are pretty small and narrow but the soft spiked insides of the capture cylinder still left marks on my finger afterwards. This was completely non-painful but just noting this here for other users who might have thicker fingers! It was a bit bigger than I expected (the size of the device) but it doesn't impact usability. It also produced quite a loud hum but again, doesn't impact usability." (Participant #11) • "I liked the brush-like texture inside the tube. It helped me feel that I had my finger in the right position." (Participant #33) • "My first impression was that it looked pretty compact and that it's very straight to the point in its features--nothing fancy, just functional." (Participant #34)

User Friendliness of UI Design

Older participants (aged ≥65 years) discussed the need for improved screen readability, mentioning increasing the font size or needing additional assistance (Participants #53 and #65; Table 6).

There were mixed opinions on the overall design of the UI, but the majority (145/151, 96%) of the participants found the overall system to be easy to use and that they could learn quickly. Some participants did comment on the elements of UI design, such as buttons, on-screen instructions, or color choices, that made them feel frustrated, confused, or uncertain about whether they were performing the measurement correctly (Participants #15, #21, and #24; Table 6). Other participants expressed satisfaction with the UI design (Participants #30 and #34; Table 6).

Ergonomic Design

Finally, participants also addressed the changes they wished to see in the ergonomic design to better meet the needs of end users (Participants #3, #16, and #11; Table 6). Other participants expressed satisfaction with the ergonomic design (Participants #33 and #34; Table 6).

Discussion

In this study, we aimed to evaluate the usability and design of PointCheck, a novel technology for noninvasive, home-based neutropenia detection. Through a mixed method approach of user observation, questionnaires, and a SUS survey, we have validated the hypothesis that PointCheck is easy to use by first-time users in a simulated home environment with a mean SUS score of 86.1 (Table 2), classified as a score of A (ie, excellent; net promoter score: promoter level) [50,51] and falling within the top tenth percentile of systems as evaluated by the SUS [49].

Although the majority of first-time users expressed high satisfaction with the overall design and user friendliness of PointCheck (Table 3), a number of areas for improvement were identified through the feedback and observation of users and will be implemented into future designs. The main changes to be implemented to enhance the usability of the device and reduce instances of errors include the addition of a tutorial video and walk-through image animations as additional training methods, improved screen readability, improved button design to make them easily identifiable to users, and a modification of the cartridge to be more size inclusive.

In observing the use of the device in context, correct positioning during the use of the device may be more difficult for nonambulatory patients. Ideally, patients will have a training session with their health care professional prior to bringing this device home for normal use. This training would allow patients to familiarize themselves with the device beforehand, ask any questions related to use, and receive the support needed to ensure confidence in using the device alone for weeks at a time.

It is to be noted that a majority of study participants were aged <65 years and have an educational level of ≥8th grade level, both of which are factors that increase the likelihood of technological proficiency and willingness to adopt new technology [52]. Emerging technologies such as smartphones and tablets have raised concerns about their ease of use in older and untrained populations [53]; however, we found that it did not affect the perceived usability of PointCheck, considering that there was no significant differences in the SUS scores among users across educational levels and age categories (Tables 4 and 5). This is consistent with prior literature which has demonstrated that older populations are interested and capable of using modern technologies for managing health and can learn how to use a touch screen after a few tries [53,54]. Although all participants in the study had no previous training or experience using PointCheck, all of them were able to become proficient after 1 or 2 measurements guided by experienced clinicians through repetition by the end of the study visit. This is an indication that training, while necessary for building intuition and confidence when using the device [55], does not need to be extensive and the on-screen walk-through is effective in guiding the user through a measurement alone.

Although this study aimed to evaluate a variety of usability factors in a simulated home environment, a single study cannot claim to assess these factors in all use cases and situations. The perceived usability of PointCheck should be tested further in real-world home environments with users who receive prior training to identify context-related issues in the future.

Overall, this study demonstrated that PointCheck, a novel digital device for noninvasive WBC monitoring, can be easy to use for unsupervised patients in the home setting. By enabling continuous home-monitoring for severe neutropenia, PointCheck has the potential to change the standard of care for patients with cancer and substantially improve their clinical outcomes.

Acknowledgments

Funding for the study was supported through the National Cancer Institute Small Business Innovation Research program (grant R44CA228920), the National Institutes of Health (grant U54HL143541), the Massachusetts Technology Collaborative Digital Health Sandbox Program, and the Centro de Desarrollo Tecnológico e Industrial in Spain (grant SNEO-20191188). We are also grateful to Catherine Ricciardi, DNP, ANP-BC (Nurse Director and Clinical Operations); and Tatiana Urman, RN, MSN (Clinical Research Nurse Coordinator) at the Massachusetts Institute of Technology Center for Clinical and Translational Research. We extend a special thank you to all volunteers and patients who participated in this study.

Conflicts of Interest

GL, APT, IB, NB, RB, AB, ASF, and CCG are current employees and holders of stock options in a privately held company, patents, and royalties of Leuko Labs, Inc. IB and CCG have membership on the board of directors or advisory committees of Leuko Labs, Inc. JML received research funding from Roche, Novartis, Incyte, Astellas, and BMS and consulted for Janssen, BMS, Novartis, Incyte, Roche, GSK, and Pfizer. None of these items are related to this work. JMS has membership on the board of directors or advisory committees of Pharmacosmos and Astra Zeneca and received honoraria from Abbvie and Stemline. None of these items are related to this work. All other authors declared no other conflicts of interest.

Multimedia Appendix 1
Short-answer qualitative e-questionnaire.
[PDF File (Adobe PDF File), 14 KB - jmir_v24i8e37368_app1.pdf]

Multimedia Appendix 2
Study coordinator training script.
[PDF File (Adobe PDF File), 17 KB - jmir_v24i8e37368_app2.pdf]

References

1. Lyman GH, Abella E, Pettengell R. Risk factors for febrile neutropenia among patients with cancer receiving chemotherapy: a systematic review. Crit Rev Oncol Hematol 2014 Jun;90(3):190-199. [doi: 10.1016/j.critrevonc.2013.12.006] [Medline: 24434034]
2. Lyman GH. Impact of chemotherapy dose intensity on cancer patient outcomes. J Natl Compr Canc Netw 2009 Jan;7(1):99-108. [doi: 10.6004/jnccn.2009.0009] [Medline: 19176210]
3. Isidori A, Cerchione C, Daver N, DiNardo C, Garcia-Manero G, Konopleva M, et al. Immunotherapy in acute myeloid leukemia: where we stand. Front Oncol 2021;11:656218 [FREE Full text] [doi: 10.3389/fonc.2021.656218] [Medline: 34041025]
4. Wilson BE, Jacob S, Yap ML, Ferlay J, Bray F, Barton MB. Estimates of global chemotherapy demands and corresponding physician workforce requirements for 2018 and 2040: a population-based study. Lancet Oncol 2019 Jun;20(6):769-780. [doi: 10.1016/S1470-2045(19)30163-9] [Medline: 31078462]
5. Tai E, Guy GP, Dunbar A, Richardson LC. Cost of cancer-related neutropenia or fever hospitalizations, United States, 2012. J Oncol Pract 2017 Jun;13(6):e552-e561 [FREE Full text] [doi: 10.1200/JOP.2016.019588] [Medline: 28437150]
6. Wang W, Li E, Campbell K, McBride A, D'Amato S. Economic analysis on adoption of biosimilar granulocyte colony-stimulating factors in patients with nonmyeloid cancer at risk of febrile neutropenia within the oncology care model framework. JCO Oncol Pract 2021 Aug;17(8):e1139-e1149. [doi: 10.1200/OP.20.00994] [Medline: 33961490]
7. Lyman GH, Poniewierski MS, Crawford J, Dale DC, Culakova E. Cost of hospitalization in patients with cancer and febrile neutropenia and impact of comorbid conditions. Blood 2015 Dec 03;126(23):2089. [doi: 10.1182/blood.v126.23.2089.2089]
8. Kuderer NM, Dale DC, Crawford J, Cosler LE, Lyman GH. Mortality, morbidity, and cost associated with febrile neutropenia in adult cancer patients. Cancer 2006 May 15;106(10):2258-2266 [FREE Full text] [doi: 10.1002/cncr.21847] [Medline: 16575919]
9. Trotti A, Colevas AD, Setser A, Rusch V, Jaques D, Budach V, et al. CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. Semin Radiat Oncol 2003 Jul;13(3):176-181. [doi: 10.1016/S1053-4296(03)00031-6] [Medline: 12903007]
10. Lyman GH, Lyman CH, Agboola O. Risk models for predicting chemotherapy-induced neutropenia. Oncologist 2005;10(6):427-437 [FREE Full text] [doi: 10.1634/theoncologist.10-6-427] [Medline: 15967836]
11. de Naurois J, Novitzky-Basso I, Gill MJ, Marti FM, Cullen MH, Roila F, ESMO Guidelines Working Group. Management of febrile neutropenia: ESMO Clinical Practice Guidelines. Ann Oncol 2010 May;21 Suppl 5:v252-v256 [FREE Full text] [doi: 10.1093/annonc/mdq196] [Medline: 20555092]
12. Kyriacou DN, Jovanovic B, Frankfurt O. Timing of initial antibiotic treatment for febrile neutropenia in the emergency department: the need for evidence-based guidelines. J Natl Compr Canc Netw 2014 Nov;12(11):1569-1573. [doi: 10.6004/jnccn.2014.0157] [Medline: 25361804]
13. Seltzer JA, Frankfurt O, Kyriacou DN. Association of an emergency department febrile neutropenia intervention protocol with time to initial antibiotic treatment. Acad Emerg Med 2022 Jan;29(1):73-82. [doi: 10.1111/acem.14335] [Medline: 34245642]
14. Wijeratne DT, Wright K, Gyawali B. Risk-stratifying treatment strategies for febrile neutropenia-tools, tools everywhere, and not a single one that works? JCO Oncol Pract 2021 Nov;17(11):651-654. [doi: 10.1200/OP.21.00148] [Medline: 33914611]
15. Zimmer AJ, Freifeld AG. Optimal management of neutropenic fever in patients with cancer. J Oncol Pract 2019 Jan;15(1):19-24. [doi: 10.1200/JOP.18.00269] [Medline: 30629902]
16. Meza L, Baselga J, Holmes F, Liang B, Breddy J, Pegfilgrastim Study Group. Incidence of febrile neutropenia (FN) is directly related to duration of severe neutropenia (DSN) after myelosuppressive chemotherapy. Proceedings of the American Society of Clinical Oncology. Vol 21 2002;21:255b.
17. Hartmann LC, Tschetter LK, Habermann TM, Ebbert LP, Johnson PS, Mailliard JA, et al. Granulocyte colony-stimulating factor in severe chemotherapy-induced afebrile neutropenia. N Engl J Med 1997 Jun 19;336(25):1776-1780. [doi: 10.1056/NEJM199706193362502] [Medline: 9187067]
18. Gafter-Gvili A, Fraser A, Paul M, Vidal L, Lawrie TA, van de Wetering MD, et al. Antibiotic prophylaxis for bacterial infections in afebrile neutropenic patients following chemotherapy. Cochrane Database Syst Rev 2012 Jan 18;1:CD004386 [FREE Full text] [doi: 10.1002/14651858.CD004386.pub3] [Medline: 22258955]

19. Egan G, Robinson PD, Martinez JPD, Alexander S, Ammann RA, Dupuis LL, et al. Efficacy of antibiotic prophylaxis in patients with cancer and hematopoietic stem cell transplantation recipients: a systematic review of randomized trials. Cancer Med 2019 Aug;8(10):4536-4546 [FREE Full text] [doi: 10.1002/cam4.2395] [Medline: 31274245]
20. Bachar N, Benbassat D, Brailovsky D, Eshel Y, Glück D, Levner D, et al. An artificial intelligence-assisted diagnostic platform for rapid near-patient hematology. Am J Hematol 2021 Oct 01;96(10):1264-1274. [doi: 10.1002/ajh.26295] [Medline: 34264525]
21. Bond MM, Richards-Kortum RR. Drop-to-drop variation in the cellular components of fingerprick blood: implications for point-of-care diagnostic development. Am J Clin Pathol 2015 Dec;144(6):885-894. [doi: 10.1309/AJCP1L7DKMPCHPEH] [Medline: 26572995]
22. Yang ZW, Yang SH, Chen L, Qu J, Zhu J, Tang Z. Comparison of blood counts in venous, fingertip and arterial blood and their measurement variation. Clin Lab Haematol 2001 Jun;23(3):155-159. [doi: 10.1046/j.1365-2257.2001.00388.x] [Medline: 11553055]
23. Daae LN, Halvorsen S, Mathisen PM, Mironska K. A comparison between haematological parameters in 'capillary' and venous blood from healthy adults. Scand J Clin Lab Invest 1988 Nov;48(7):723-726. [doi: 10.1080/00365518809085796] [Medline: 3201107]
24. Hollis VS, Holloway JA, Harris S, Spencer D, van Berkel C, Morgan H. Comparison of venous and capillary differential leukocyte counts using a standard hematology analyzer and a novel microfluidic impedance cytometer. PLoS One 2012;7(9):e43702 [FREE Full text] [doi: 10.1371/journal.pone.0043702] [Medline: 23028467]
25. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
26. IEC 62366-1:2015(en) medical devices — part 1: application of usability engineering to medical devices. International Electrotechnical Commission. 2015. URL: https://www.iso.org/obp/ui/#iso:std:iec:62366:-1:ed-1:v1:en [accessed 2022-05-31]
27. DE Bleser L, DE Geest S, Vincke B, Ruppar T, Vanhaecke J, Dobbels F. How to test electronic adherence monitoring devices for use in daily life: a conceptual framework. Comput Inform Nurs 2011 Sep;29(9):489-495. [doi: 10.1097/NCN.0b013e31821a1555] [Medline: 21532468]
28. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
29. Kortum P, Peres SC. Evaluation of home health care devices: remote usability assessment. JMIR Hum Factors 2015 Jun 05;2(1):e10 [FREE Full text] [doi: 10.2196/humanfactors.4570] [Medline: 27025664]
30. Chamberlain JJ, Gilgen E. Do perceptions of insulin pump usability impact attitudes toward insulin pump therapy? a pilot study of individuals with type 1 and insulin-treated type 2 diabetes. J Diabetes Sci Technol 2015 Jan;9(1):105-110 [FREE Full text] [doi: 10.1177/1932296814552822] [Medline: 25269659]
31. Arnet I, Rothen JP, Hersberger KE. Validation of a novel electronic device for medication adherence monitoring of ambulatory patients. Pharmacy (Basel) 2019 Nov 20;7(4):155 [FREE Full text] [doi: 10.3390/pharmacy7040155] [Medline: 31756904]
32. Elkefi S, Choudhury A, Strachna O, Asan O. Impact of health perception and knowledge on genetic testing decisions using the health belief model. JCO Clin Cancer Inform 2022 Jan;6:e2100117 [FREE Full text] [doi: 10.1200/CCI.21.00117] [Medline: 34990211]
33. ElKefi S, Asan O. How technology impacts communication between cancer patients and their health care providers: a systematic literature review. Int J Med Inform 2021 May;149:104430 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104430] [Medline: 33684711]
34. Pablo-Trinidad A, Butterworth I, Ledesma-Carbayo MJ, Vettenburg T, Sánchez-Ferro Á, Soenksen L, et al. Automated detection of neutropenia using noninvasive video microscopy of superficial capillaries. Am J Hematol 2019 Aug;94(8):E219-E222 [FREE Full text] [doi: 10.1002/ajh.25516] [Medline: 31120579]
35. Bourquard A, Butterworth I, Sanchez-Ferro A, Giancardo L, Soenksen L, Cerrato C, et al. Analysis of white blood cell dynamics in nailfold capillaries. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:7470-7473 [FREE Full text] [doi: 10.1109/EMBC.2015.7320119] [Medline: 26738019]
36. Bourquard A, Pablo-Trinidad A, Butterworth I, Sánchez-Ferro Á, Cerrato C, Humala K, et al. Non-invasive detection of severe neutropenia in chemotherapy patients by optical imaging of nailfold microcirculation. Sci Rep 2018 Mar 28;8(1):5301 [FREE Full text] [doi: 10.1038/s41598-018-23591-0] [Medline: 29593221]
37. Dinsdale G, Roberts C, Moore T, Manning J, Berks M, Allen J, et al. Nailfold capillaroscopy-how many fingers should be examined to detect abnormality? Rheumatology (Oxford) 2019 Feb 01;58(2):284-288. [doi: 10.1093/rheumatology/key293] [Medline: 30247696]
38. Smith V, Thevissen K, Trombetta AC, Pizzorni C, Ruaro B, Piette Y, EULAR Study Group on Microcirculation in Rheumatic Diseases. Nailfold capillaroscopy and clinical applications in systemic sclerosis. Microcirculation 2016 Jul;23(5):364-372. [doi: 10.1111/micc.12281] [Medline: 27086648]
39. McKay GN, Mohan N, Butterworth I, Bourquard A, Sánchez-Ferro Á, Castro-González C, et al. Visualization of blood cell contrast in nailfold capillaries with high-speed reverse lens mobile phone microscopy. Biomed Opt Express 2020 Apr 01;11(4):2268-2276 [FREE Full text] [doi: 10.1364/BOE.382376] [Medline: 32341882]

40. Chojnowski MM, Felis-Giemza A, Olesińska M. Capillaroscopy - a role in modern rheumatology. Reumatologia 2016;54(2):67-72 [FREE Full text] [doi: 10.5114/reum.2016.60215] [Medline: 27407282]
41. Sauro J, Lewis J. Chapter 4 - did we meet or exceed our goal? In: Inauro J, Lewis JR, editors. Quantifying the User Experience. 2nd ed. Boston, MA: Morgan Kaufmann; 2016:39-60.
42. Toonders SAJ, van Westrienen PE, Konings S, Nieboer ME, Veenhof C, Pisters MF. Patients' perspectives on the usability of a blended approach to an integrated intervention for patients with medically unexplained physical symptoms: mixed methods study. J Med Internet Res 2021 Sep 28;23(9):e19794 [FREE Full text] [doi: 10.2196/19794] [Medline: 34581674]
43. Chaniaud N, Megalakaki O, Capo S, Loup-Escande E. Effects of user characteristics on the usability of a home-connected medical device (Smart Angel) for ambulatory monitoring: usability study. JMIR Hum Factors 2021 Mar 17;8(1):e24846 [FREE Full text] [doi: 10.2196/24846] [Medline: 33729161]
44. E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R1) Guidance for Industry. United States Food and Drug Administration. 2018. URL: https://www.fda.gov/files/drugs/published/E6%28R2%29-Good-Clinical-Practice--Integrated-Addendum-to-ICH-E6%28R1%29.pdf [accessed 2022-07-29]
45. Lewis J, Sauro J. The Factor Structure of the System Usability Scale. In: Lecture Notes in Computer Science. 2009 Jul Presented at: HCD 2009: Human Centered Design; July 19-24, 2009; San Diego, CA p. 94-103. [doi: 10.1007/978-3-642-02806-9_12]
46. Sauro J, Lewis JR. Chapter 8 - standardized usability questionnaires. In: Sauro J, Lewis JR, editors. Quantifying the User Experience. 2nd ed. Boston, MA: Morgan Kaufmann; 2016:185-248.
47. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. London, UK: Taylor and Francis; 1996:1-22.
48. RStudio Team. RStudio: integrated development environment for R. RStudio. 2021. URL: http://www.rstudio.com/ [accessed 2022-05-31]
49. Sauro J. 5 ways to interpret a SUS score. Measuring U. 2018 Sep 19. URL: https://measuringu.com/interpret-sus-score/ [accessed 2022-05-31]
50. Lewis JR, Sauro J. Item benchmarks for the system usability scale. J Usability Stud 2018 May 01;13(3):158-167 [FREE Full text] [doi: 10.5555/3294033.3294037]
51. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009 May 01;4(3):114-123 [FREE Full text] [doi: 10.5555/2835587.2835589]
52. Older Americans and the Internet. Pew Research Center. 2004 Mar 28. URL: https://www.pewresearch.org/internet/2004/03/28/older-americans-and-the-internet/ [accessed 2022-05-23]
53. 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]
54. Yang HH, Yu C, Huang CH, Kuo LH, Yang HJ. Teaching information technology for elder participation: a qualitative analysis of Taiwan retirees. WSEAS Transactions on Information Science and Applications 2010 Sep 01;7(9):1190-1199 [FREE Full text] [doi: 10.5555/1865374.1865382]
55. Keller SC, Gurses AP, Werner N, Hohl D, Hughes A, Leff B, et al. Older adults and management of medical devices in the home: five requirements for appropriate use. Popul Health Manag 2017 Aug;20(4):278-286 [FREE Full text] [doi: 10.1089/pop.2016.0070] [Medline: 28075698]

Abbreviations

FN: febrile neutropenia
IRB: institutional review board
SUS: System Usability Scale
UI: User Interface
WBC: white blood cells

Edited by G Eysenbach; submitted 18.02.22; peer-reviewed by N Chaniaud, S El kefi; comments to author 11.04.22; revised version received 03.06.22; accepted 18.07.22; published 09.08.22.

Please cite as:
Lamaj G, Pablo-Trinidad A, Butterworth I, Bell N, Benasutti R, Bourquard A, Sanchez-Ferro A, Castro-Gonzalez C, Jiménez-Ubieto A, Baumann T, Rodriguez-Izquierdo A, Pottier E, Shelton A, Martinez-Lopez J, Sloan JM
Usability Evaluation of a Noninvasive Neutropenia Screening Device (PointCheck) for Patients Undergoing Cancer Chemotherapy: Mixed Methods Observational Study
J Med Internet Res 2022;24(8):e37368
URL: https://www.jmir.org/2022/8/e37368
doi:10.2196/37368
PMID:35943786

Original Paper

Testing the Effectiveness of an Animated Decision Aid to Improve Recruitment of Control Participants in a Case-Control Study: Web-Based Experiment

Sandro T Stoffel[1,2], MSc, PhD; Jing Hui Law[1,3], MSc, MRes; Robert Kerrison[1,4], MSc, PhD; Hannah R Brewer[5], MSc, PhD; James M Flanagan[5], BSc, PhD; Yasemin Hirst[1,5], MSc, PhD

[1]Department of Behavioural Science and Health, University College London, London, United Kingdom
[2]Institute of Pharmaceutical Medicine, University of Basel, Basel, Switzerland
[3]Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom
[4]School of Health Sciences, University of Surrey, Guildford, United Kingdom
[5]Department of Surgery and Cancer, Imperial College London, London, United Kingdom

Corresponding Author:
Yasemin Hirst, MSc, PhD
Department of Behavioural Science and Health
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 2076791615
Email: y.hirst@ucl.ac.uk

Abstract

Background: Participation in case-control studies is crucial in epidemiological research. The self-sampling bias, low response rate, and poor recruitment of population representative controls are often reported as limitations of case-control studies with limited strategies to improve participation. With greater use of web-based methods in health research, there is a further need to understand the effectiveness of different tools to enhance informed decision-making and willingness to take part in research.

Objective: This study tests whether the inclusion of an animated decision aid in the recruitment page of a study website can increase participants' intentions to volunteer as controls.

Methods: A total of 1425 women were included in a web-based experiment and randomized to one of two experimental conditions: one in which they were exposed to a simulated website that included the animation (animation; n=693, 48.6%), and one in which they were exposed to the simulated website without the animation (control; n=732, 51.4%). The simulated website was adapted from a real website for a case-control study, which invites people to consider taking part in a study that investigates differences in purchasing behaviors between women with and without ovarian cancer and share their loyalty card data collected through 2 high street retailers with the researchers. After exposure to the experimental manipulation, participants were asked to state (1) their intention to take part in the case-control study, (2) whether they would be willing to share their loyalty card for research, and (3) their willingness to be redirected to the real website after completing the survey. Data were assessed using ordinal and binary logistic regression, reported in percentages (%), adjusted odds ratio (AOR), and 95% confidence intervals.

Results: Including the animation in the simulated website did not increase intentions to participate in the study (AOR 1.09; 95% CI 0.88-1.35) or willingness to visit the real study website after the survey (control 50.5% vs animation 52.6%, AOR 1.08; 95% CI 0.85-1.37). The animation, however, increased the participants' intentions to share the data from their loyalty cards for research in general (control 17.9% vs animation 26%; AOR 1.64; 95% CI 1.23-2.18).

Conclusions: While the results of this study indicate that the animated decision aid did not lead to greater intention to take part in our web-based case-control study, they show that they can be effective in increasing people's willingness to share sensitive data for health research.

(J Med Internet Res 2022;24(8):e40015) doi:10.2196/40015

KEYWORDS

animation; research participation; online experiment; case-control; recruitment; decision; effectiveness; epidemiology; recruitment; online; experiment; volunteer; survey; willingness; data; health research; research

Introduction

One of the most effective methods to test for exposure in epidemiological research is to conduct a case-control study in which people who have an illness are compared retrospectively with a matching population without the outcome [1]. Although it is a reliable methodology to test for associations, poor recruitment of population representative controls often undermines such studies [2].

Previous research reports the most common methods of recruiting control participants for case-control studies as follows: door-to-door recruitment, postal invitation, and random digit dialing [1,2]. More recently, with greater access to the internet, many cohort studies have moved their participant management and recruitment online (using unique websites), providing new opportunities to recruit participants, potentially improving diversity and ease of data collection [3,4]. Despite several advantages, however, caution needs to be exercised with the opportunistic recruitment of participants to web-based studies. For example, a recent study reported less than 4% of participants who visited a study recruitment website, after clicking on a targeted social media advertisement campaign, went on to sign up to the research [5]. This indicates that while individuals may be forming some interest to take part in research studies by clicking on a recruitment advertisement, their intention does not always translate to survey completion after they land on the research website.

Evidence on the barriers and facilitators of web-based survey completion primarily relates to the completion of stand-alone web-based surveys, rather than the use of unique websites to recruit participants to case-control studies [6-11]. These studies suggest that individuals' trust in the organization carrying out the research, whether they are early adopters of technology and high in literacy, and whether the research is in line with the individuals' values and beliefs are positive predictors of individual participation. Recommendations to achieve better outcomes include clear communication of the research goals, transparency about how data will be used, and shorter survey length.

Clinical trials have attempted to address some of the above (eg, transparency about how data will be used), using audio-visual decision aids to supplement the process of obtaining informed consent [12]. Communicating information via these mediums (enabled through web-based recruitment strategies), have the potential to reduce the associated cognitive load, facilitate further engagement with the research aims, generate positive attitudes toward the targeted behavior, and subsequently motivate engagement in the behavior itself [13-16]. To our knowledge, the potential impact of animated decision aids on intentions to take part in a web-based case-control study has not previously been investigated. This study aims to measure the effectiveness of an animated decision aid as a supplementary tool on a simulated website of a case-control study to encourage participation.

Methods

Setting

This study comprised a randomized web-based experiment, which assessed the effectiveness of adding an animated decision aid to a simulated website. The simulated website the animation was designed for (or added to) was the recruitment website for the case-control study: Cancer Loyalty Card Study (CLOCS) [17].

Cancer Loyalty Card Study

CLOCS is an observational case-control study that aims to investigate the self-care behaviors of patients with ovarian cancer prior to their cancer diagnosis. It seeks to do this by investigating differences in transactional data (such as medication purchasing) between women with and without ovarian cancer (the transactional data are collected through the loyalty cards of 2 UK-based high street retailers). Cases (ie, women with ovarian cancer) are recruited through participating National Health Service sites, while controls are recruited through the study website. Full details for CLOCS have previously been reported in the study protocol [17].

Animated Decision Aid

A key challenge for CLOCS recruitment has been communicating the research aims clearly, and our previous research highlights that the public often needs further explanations for how individual transactional data can be used in health research [18]. To improve public understanding and engagement with the aims of CLOCS, an animated decision aid was jointly prepared by Science Animated Limited, 2 patient representatives, and the CLOCS research team prior to the initiation of this web-based experiment. It aimed to convey key facts on ovarian cancer, the potential contribution of CLOCS in informing earlier diagnosis, as well as how women in the general population can play a role to aid its efforts (based on the participant information sheet tailored and approved by a National Health Service Research Ethics Committee [19/NW/0427-SA1], included in Multimedia Appendix 1). However, it should be noted that the animation was designed as a supplement to the main study materials, not as a key participant communication material to prompt informed decision-making. The animation features English subtitles and is 123 seconds (2 minutes and 3 seconds) long [19].

Procedure

The randomized web-based experiment was programmed in Survey Monkey. In July 2020, women who were eligible to take part in CLOCS as a control(ie, between the ages of 18 and 70, living in the United Kingdom, and without an ovarian cancer diagnosis) were recruited through a survey vendor (Dynata Limited). For those who were interested in taking part, information about the experiment, including a brief description of CLOCS, was presented and followed by the completion of the study consent form. If participants consented and were eligible, they were randomized (in a 1:1 ratio) to one of the following two experimental conditions: the simulated CLOCS website without the animation (control), or the simulated CLOCS website with the animation (animation) (Multimedia Appendix 2 and 3).

Once everyone viewed the simulated website, they were asked to complete the survey, where they were required to indicate their intention to take part in CLOCS using a 4-point Likert scale (definitely yes, probably yes, probably not, and definitely not) adapted from previous research [20-22]. The study participants were then asked to indicate their loyalty card use by selecting from a list of high street retailers' loyalty cards and whether they would be willing to share their loyalty card data for research purposes. The latter question was adapted from research on willingness to share electronic health data, which uses 4 commonly used models of consent for the use of data [23].

In the next step, the study participants were asked about their educational level, annual household income, and health literacy. For the latter, the eHealth Literacy Scale was used. This scale assesses individuals' retrieval and judgement of health information on the internet [24]. It consists of 8 items rated on a 5-point scale, ranging from 1 to 5, and has demonstrated considerable reliability and validity [24]. The participants' scores across the 8 items on the scale were summed and calculated for a sample mean. Individual scores below and above (or equal to) this sample mean were defined as low and high health literacy, respectively [25,26].

The final survey item was included as a behavior *proxy*. The participants were asked whether they want to be redirected to the actual CLOCS website for more information on how to take part. Those who responded that they would like to visit the website were provided with a link to the CLOCS website at the final page of the survey [22]. The website opened in a new tab for participants who clicked on the link. No data were collected from the study participants for their direct participation in CLOCS associated with the experiment.

Ethics Approval

Ethics approval for this study was granted by the UCL Research Ethics Committee (Project ID: 17823/001).

Data Analysis

A pilot study has been conducted beforehand for the purpose of sample size calculations. Based on the findings from the initial sample of 359 participants, with a 10% difference in intention to take part (definitely yes or probably yes versus definitely no or probably no), we determined the number of participants needed to achieve 95% certainty and 80% power was 650 per trial arm. Data from participants in both the pilot and final sample were combined for analysis.

Sample characteristics were assessed using descriptive statistics (Table S1 in Multimedia Appendix 4). To aid interpretation, the participants' income and educational levels were dichotomized in inferential analyses. For income, we used £30,000 (US $37,000) as the cut-off point, based on the average household income in the United Kingdom (reported by the Office for National Statistics, 2020) [27]. For education, we categorized participants into those with General Certificate of Secondary Education or A Levels and those with a university degree [20-22].

Differences in intentions to take part in CLOCS were assessed using univariate and multivariate ordinal logistic regression. Willingness to visit the actual website and willingness to share loyalty card data for research purposes, between groups, were assessed using univariate and multivariate binary logistic regressions. Odds ratios (ORs), 95% confidence intervals, and P values are presented in the results, with P values below .05 regarded as statistically significant. Participants who spent a short amount of time completing the survey (ie, *survey speeders*) were excluded from the analysis (Multimedia Appendices 5-8, based on the 50% cut-off points provided from previous research using median values) [28]. We report the analysis for the whole sample in Table S2 (Multimedia Appendix 9) and the distribution of the time spent on the survey before and after the exclusion of the *speeders* in Multimedia Appendices 5-8 [29]. Additional analyses for interaction between the intervention and health literacy led to null results and were not reported due to the unbalanced proportion of those with low health literacy and high literacy in the study population.

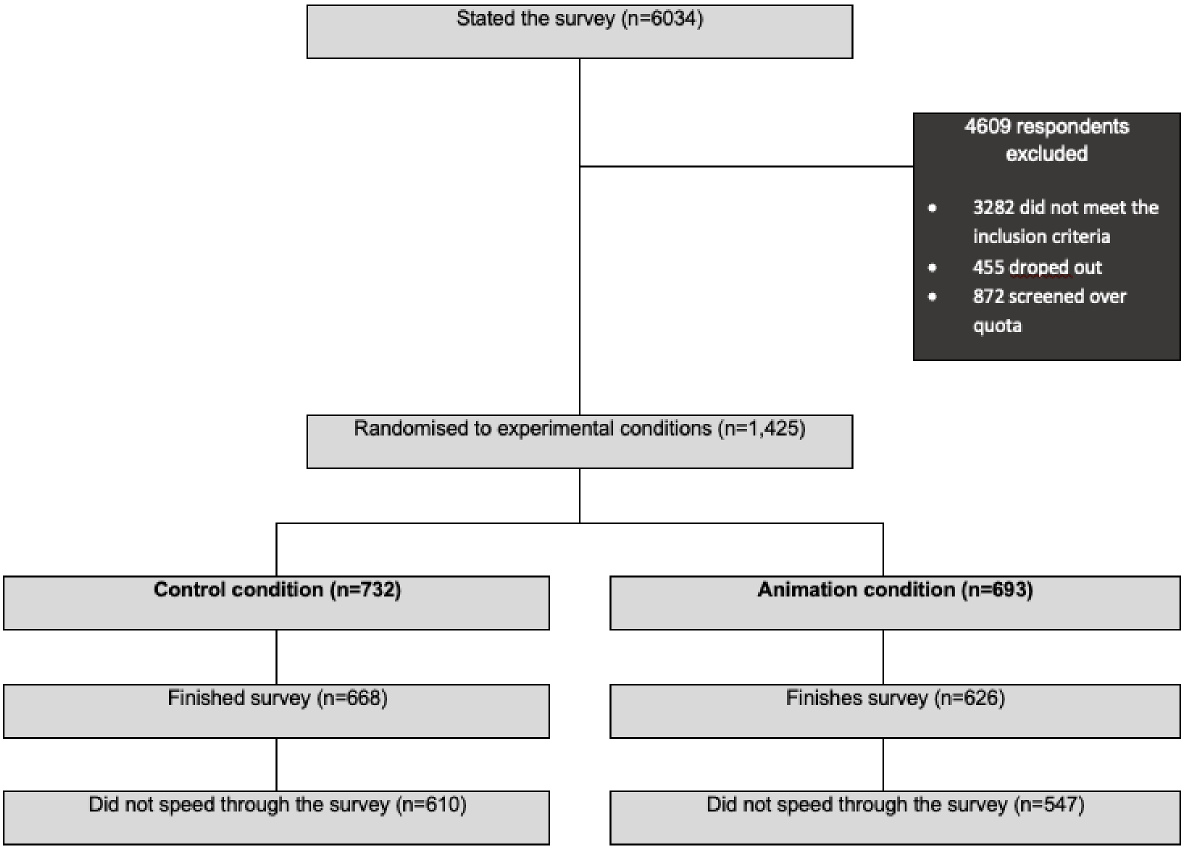
Results

Study Sample

Figure 1 demonstrates the flow of participants through the study. In total, 6034 women were invited to participate, of which 4609 (76.4%) were excluded as they were not eligible, dropped out, or discontinued the initial screening. The remaining 1425 (23.6%) were then randomized to one of the following two experimental conditions: 732 (51.4%) were randomly allocated to the control condition, and 693 (48.6%) to the animation condition. Across conditions, 131 (9.2%) did not finish the survey after randomization. Furthermore, 137 (9.6%) were excluded as they spent less than 50% of the median time. The analytical sample consisted of 1157 women—610 (52.7%) participants in the control condition, and 547 (47.3%) in the animation condition.

Most women in the analytical sample were aged between 55 and 70 years (n=417, 36.9%), did not have a university degree (n=686, 59.3%), and had an annual household income of less than £30,000 (US $37,000; n=622, 53.8%). The mean eHealth Literacy Scale score for the participants was 30.3 out of 40; thus, those who scored below this mean were classified as having low health literacy [20,21]. Post hoc comparisons revealed that sociodemographic characteristics were comparable between the two experimental conditions (Table S1 in Multimedia Appendix 4).

Figure 1. Flow through the study.



Intentions to Participate in CLOCS

Intentions to participate in CLOCS were generally very high, with 69.7% (n=807) of women stating that they would probably or definitely participate. Figure 2 shows the distribution of the intentions to participate in CLOCS after exposure to the simulated website. The ordered logistic regressions in Table 1 show that the inclusion of the animation did not affect participation intentions (OR 1.12; 95% CI 0.90-1.39 and AOR 1.09; 95% CI 0.88-1.35). The regression further shows that older women aged 55-70 years stated lower intentions to participate (OR 0.56; 95% CI 0.42-0.74), while those with an income above average (OR 1.36; 95% CI 1.06-1.63), one or more existing loyalty cards (AOR 2.05; 95% CI 1.16-3.62), and low health literacy (AOR 1.26; 95% CI 1.01- 1.58) had higher intentions to take part in CLOCS.

Figure 2. Distribution for intention to take part in Cancer Loyalty Card Study.

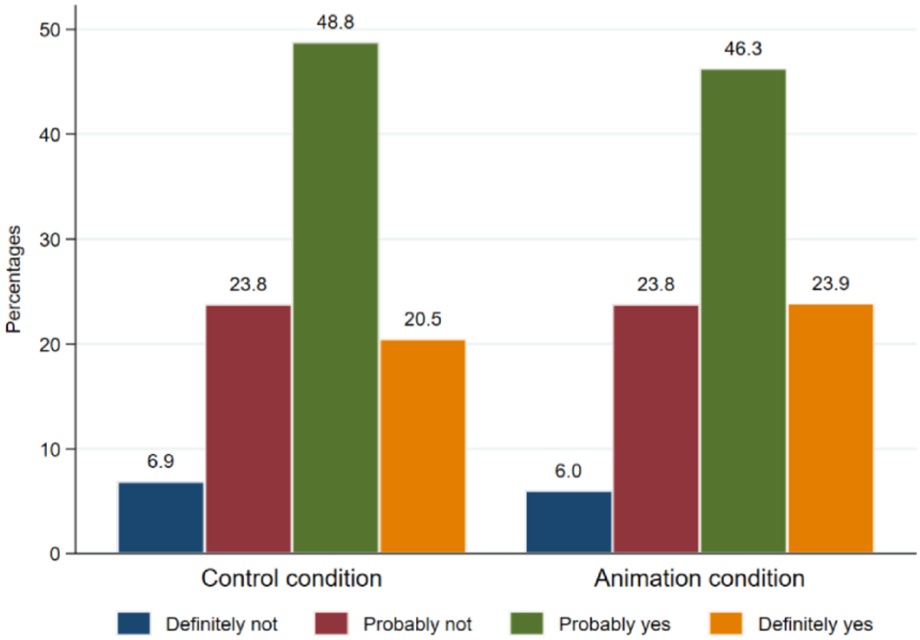


Table 1. Ordered logistic regression on intention to participate in Cancer Loyalty Card Study (N=1157).

Variables	Unadjusted regression			Adjusted regression		
	AOR[a]	95% CI	*P* value	AOR	95% CI	*P* value
Condition						
Control	Reference	—[b]	—	Reference	—	—
Animation	1.120	0.904-1.389	.30	1.088	0.876-1.352	.45
Age (years)						
18-34	Reference	—	—	Reference	—	—
35-44	0.998	0.726-1.371	.99	0.952	0.691-1.311	.76
45-54	0.806	0.577-1.127	.21	0.799	0.569-1.122	.20
55-70	0.556	0.421-0.735	<.001	0.591	0.445-0.786	<.001
Education						
Below or equal to GCSE[c]	Reference	—	—	Reference	—	—
University degree	1.320	1.060-1.643	.01	1.145	0.909-1.441	.25
Income						
Below average	Reference	—	—	Reference	—	—
Above average	1.362	1.097-1.690	.01	1.263	1.006-1.584	.04
Card						
No	Reference	—	—	Reference	—	—
Yes	2.114	1.202-3.719	.009	2.051	1.164-3.616	.01
Health literacy						
High literacy	Reference	—	—	Reference	—	—
Low literacy	1.501	1.125-2.002	.006	1.381	1.029-1.854	.03

[a] AOR: adjusted odds ratio.

[b] Not applicable.

[c] GCSE: General Certificate of Secondary Education.

Willingness to Share Loyalty Card Data for Research

Most study participants stated that their data should be used but that they should have the option of saying no (control condition: n=246, 40.3%; and animation condition: n=181, 33.1%; Figure 3). However, significantly more participants in the animation condition, compared to control, indicated that they would be willing to provide the data if they were needed—as shown in the binary logistic regressions in Table 2 (control: n=109, 17.9% vs animation: n=142, 26.0%; OR 1.61; 95% CI 1.22-2.14 and AOR 1.64; 95% CI 1.23-2.18). Similarly, as with the intentions to participate in CLOCS, women aged 55-70 years were again less likely to state that their data should be used if needed compared to those aged 18-34 years (n=71, 17% vs n=72, 24.8%; AOR 0.62; 95% CI 0.43-0.89).

Figure 3. Distribution for willingness to share loyalty card data .

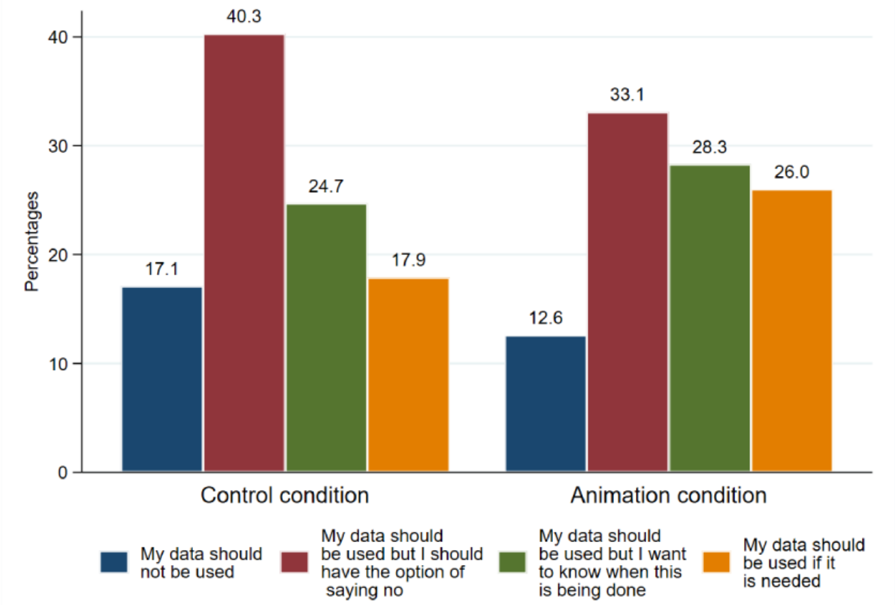


Table 2. Binary logistic regression on agreeing to share data from loyalty cards when needed (N=1157).

Variables	Total, n (%)	Unadjusted regression			Adjusted regression		
		AOR[a]	95% CI	*P* value	AOR	95% CI	*P* value
Overall	251 (21.7)	__[b]	—	—	—	—	—
Condition							
Control	109 (17.9)	Reference	—	—	Reference	—	—
Animation	142 (26.0)	1.612	1.216-2.136	.001	1.640	1.232-2.184	.001
Age (years)							
18-34	72 (24.8)	Reference	—	—	Reference	—	—
35-44	55 (22.9)	0.900	0.602-1.346	.61	0.864	0.575-1.299	.48
45-54	53 (25.2)	1.022	0.678-1.540	.92	1.046	0.688-1.591	.83
55-70	71 (17.0)	0.621	0.430-0.899	.01	0.601	0.411-0.878	.008
Education							
Below or equal to GCSE[c]	147 (21.4)	Reference	—	—	Reference	—	—
University degree	104 (22.1)	1.039	0.782-1.380	.79	0.932	0.690-1.260	.65
Income							
Below average income	130 20.9)	Reference	—	—	Reference	—	—
Above average income	121 (22.6)	1.106	0.836-1.464	.48	1.052	0.782-1.414	.74
Card							
No	9 (20.9)	Reference	—	—	Reference	—	—
Yes	242 (21.7)	1.048	0.496-2.216	.90	1.094	0.511-2.344	.82
Health literacy							
High literacy	214 (22.0)	Reference	—	—	Reference	—	—
Low literacy	37 (20.0)	0.886	0.599-1.309	.54	0.807	0.541-1.204	.30

[a] AOR: adjusted odds ratio.

[b] Not applicable.

[c] GCSE: General Certificate of Secondary Education.

Willingness to Visit the CLOCS Website After the Survey

A slight majority of the study participants (n=596, 51.5%) indicated that they would like to visit the CLOCS website after the end of the survey (Multimedia Appendix 10). Table 3 shows that there was no difference between the two experimental conditions (control: n=308, 50.5% vs animation: n=288, 52.6%; OR 1.09; 95% CI 0.87-1.37 and AOR 1.08; 95% CI 0.85-1.37). Women with low health literacy, in comparison to those with high health literacy (n=111, 60% vs n=485, 49.9%; AOR 1.43; 95% CI 1.03-1.98), and women aged 35-44 years, in comparison to those aged 18-34 years, were more interested in visiting the study website (n=145, 60.4% vs n=148, 51%; AOR 1.49; 95% CI 1.05-2.12).

Table 3. Binary logistic regression on willingness to visit website after the survey (N=1157).

Variables	Total, n (%)	Unadjusted regression			Adjusted regression		
		AOR[a]	95% CI	*P* value	AOR	95% CI	*P* value
Overall	596 (51.5)	—[b]	—	—	—	—	—
Condition							
Control	308 (50.5)	Reference	—	—	Reference	—	—
Animation	288 (52.7)	1.090	0.865-1.374	.46	1.079	0.853-1.367	.53
Age (years)							
18-34 years	148 (51.0)	Reference	—	—	Reference	—	—
35-44 years	145 (60.4)	1.464	1.036-2.071	.03	1.489	1.048-2.115	.03
45-54 years	118 (56.2)	1.231	0.861-1.758	.25	1.303	0.905-1.877	.16
55-70 years	185 (44.4)	0.765	0.566-1.033	.08	0.825	0.606-1.123	.22
Education							
Below or equal to GCSE[c]	337 (49.1)	Reference	—	—	Reference	—	—
University degree	259 (55.0)	1.265	1.000-1.601	.05	1.214	0.946-1.559	.13
Income							
Below average income	322 (51.8)	Reference	—	—	Reference	—	—
Above average income	274 (51.2)	0.978	0.776-1.233	.85	0.865	0.677-1.107	.25
Card							
No	18 (41.9)	Reference	—	—	Reference	—	—
Yes	578 (51.9)	1.498	0.808-2.776	.20	1.521	0.810-2.854	.19
Health literacy							
High literacy	485 (49.9)	Reference	—	—	Reference	—	—
Low literacy	111 (60.0)	1.506	1.094-2.074	.01	1.431	1.031-1.986	.03

[a] AOR: adjusted odds ratio.

[b] Not applicable.

[c] GCSE: General Certificate of Secondary Education.

Discussion

Key Findings

This randomized web-based experiment examined the effectiveness of an animated decision aid to increase the willingness to participate in a case-control study. The results show that the animation did not increase intentions to participate in a real-world case-control study (CLOCS), or willingness to visit the real study website after the survey. However, the animation increased the participants' willingness to share data from their loyalty cards for research. Interestingly, immediately after the completion of this web-based experiment, there was a spike in activity within the case-control study, with over 100 people signing up to participate in CLOCS.

Comparison With Previous Literature

Our findings reflect the mixed evidence currently available in the literature. While there has been some support for the effectiveness of animated decision aids in the context of health behavior research, many studies have focused on how the interventions can improve participants' knowledge of the health behavior concerned, as their main outcome. When assessing participant intention to engage in the behavior (or an objective measurement of the behavior itself), the findings have been more inconsistent [30,31].

An interesting finding in our study is that, among all participants, those with lower health literacy scores were more interested to find out about the study compared with those with high health literacy scores. Our sample size calculations were based solely on the primary outcome; thus, we might have been underpowered to detect the interaction effects of health literacy and outcomes on the behavior *proxy*. However, the previous studies indicate that multimedia interventions are not always significant in individuals with lower educational levels [14] and low health literacy [15,16]. It has been previously shown in the cancer screening literature that gist-based supplementary materials could be used to enhance engagement with the main literature among people with low numeracy [32], and perhaps using animation and other easy-to-read materials could enhance participant recruitment in health research [33]. Furthermore, the positive association between low health literacy and willingness to visit the website may be explained by factors such as wanting to find more information, the salience of the research topic, and other factors that were not included in this web-based experiment. As such, future studies focusing on the comprehension of the materials among people with low health literacy using a think-aloud methodology could explain this outcome.

Strengths and Limitations

This study has some important limitations, which call for follow-up research. First, we did not include a comprehension assessment to check if participants fully understood or watched the animation. Second, we did not measure attitudes toward the simulated website and CLOCS. A recent systematic review on participant comprehension and informed consent in health research further highlighted that while there are efforts to improve participation rates using various methods, there is a lack of assessment of participant readability, literacy, and standardization of recruitment methods in health research for informed consent procedures [33]. While all the participant-facing CLOCS research materials have been reviewed by patient and public representatives, further assessment of participant comprehension prior to the animation experiment would have strengthened our methodology. However, the rationale for the exclusion of these measures was based on the assumptions that people often form immediate decisions about whether something is relevant to them using heuristic decision-making before establishing deliberative decisions [34]. As such, by excluding cognitive measures in our assessment to minimize judgement and bias, we tried to capture individuals' potential reactions to the website as close to their reaction in real life. In this context, further studies using eye-tracking experiments on the simulated website will be highly informative to build a better understanding of the interaction with the website and the contents [35].

On the other hand, exposure to the animation increased the intentions to share loyalty card data for research in general, but not for intentions to participate in the CLOCS study; this suggests that there are study-specific characteristics that did not appeal to individuals (eg, actively signing up to provide information) or that the study participants were not eligible. However, a recent study also shows that only half of the population is willing to share shopping data for health research, highlighting the differences in sociodemographic characteristics of people who are willing to share their data for health research [36]. The characteristics of the participants who are willing to take part in CLOCS in this study mirror the results of this age-stratified survey employed in England, with older women less willing to take part. While self-sampling bias will continue to be a concern of case-control studies based on the differences in characteristics of the people who are willing to take part, recruitment strategies could be stratified and tailored to engage different populations who are less willing to take part in health research based on this evidence and the validation of public acceptability. Our results support this evidence further using experimental design with greater internal validity for potential barriers in recruiting participants to a case-control study.

Implications for Policy and Future Research

While such questions of generalizability are warranted, this study still poses important relevance and implications to current research contexts. Due to the COVID-19 pandemic, many research studies have been forced to consider the possibility of being adapted online. This may require researchers to derive additional strategies to reach web-based samples with different characteristics. Underrepresentation in health research is already an issue for minority populations, people with low literacy, and those with greater deprivation using traditional methods of recruitment, unless they are specifically targeted [37,38]; thus, there is a further need to ensure that web-based strategies can provide means for researchers to attract representative samples.

Future studies should therefore continue exploring web-based methods to facilitate complex decision-making processes for potential participants of health research. The CLOCS animation has not been actively disseminated for participant recruitment following this dearth of evidence; however, unique findings might be obtained for research of a different nature. Other multimedia formats or mediums such as social media can be further explored in future studies, along with the consideration of potentially important variables such as participants' willingness to share data for research purposes.

Conclusion

The results of this study indicate that the animated decision aid did not influence the participants' intention to take part in CLOCS or visit the study website. The animation, however, increased the probability of individuals stating that they would share their loyalty card data for research. Future research should continue exploring methods that can effectively engage participants with low health literacy to participate in complex health research.

Acknowledgments

The authors would like to acknowledge Deborah Tanner and Fiona Murphy who have acted as representatives to patients with ovarian cancer since the inception of the CLOCS project and helped us inform the design and contents of the animation. This study was funded by Cancer Research UK Early Detection and Diagnosis project grant (C38463/A26726).

Authors' Contributions

JHL, STS, RK, and YH developed the study concept and design. YH, HRB, and JMF designed and developed the animation aid for the Cancer Loyalty Card Study (CLOCS). JHL, STS, and YH performed the data analysis and interpretation. JHL, STS, and YH drafted the manuscript. All authors provided critical revisions and approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Cancer Loyalty Card Study participant information sheet - health volunteer version 2.
[DOCX File , 666 KB - jmir_v24i8e40015_app1.docx]

Multimedia Appendix 2
A screenshot for the simulated website without the animation.
[PNG File , 523 KB - jmir_v24i8e40015_app2.png]

Multimedia Appendix 3
A screenshot of the simulated website with the animation.
[PNG File , 553 KB - jmir_v24i8e40015_app3.png]

Multimedia Appendix 4
Sample characteristics.
[DOCX File , 14 KB - jmir_v24i8e40015_app4.docx]

Multimedia Appendix 5
Distribution for time spent on the survey without excluding speeders (control group).
[PNG File , 72 KB - jmir_v24i8e40015_app5.png]

Multimedia Appendix 6
Distribution for time spent on the survey after excluding speeders (Control group).
[PNG File , 78 KB - jmir_v24i8e40015_app6.png]

Multimedia Appendix 7
Distribution for time spent on the survey without excluding speeders (Animation group).
[PNG File , 66 KB - jmir_v24i8e40015_app7.png]

Multimedia Appendix 8
Distribution for time spent on the survey excluding speeders (Animation group).
[PNG File , 77 KB - jmir_v24i8e40015_app8.png]

Multimedia Appendix 9
Adjusted regression models on the whole sample without excluding the speeders.
[DOCX File , 14 KB - jmir_v24i8e40015_app9.docx]

Multimedia Appendix 10
Distribution of willingness to visit the website after the survey.
[PNG File , 18 KB - jmir_v24i8e40015_app10.png]

References

1. Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. The Lancet 2005 Apr;365(9468):1429-1433. [doi: 10.1016/s0140-6736(05)66379-9]
2. Bernstein L. Control recruitment in population-based case-control studies. Epidemiology 2006 May;17(3):255-257. [doi: 10.1097/01.ede.0000209440.94875.42] [Medline: 16617272]
3. Brøgger-Mikkelsen M, Ali Z, Zibert JR, Andersen AD, Thomsen SF. Online Patient Recruitment in Clinical Trials: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Nov 04;22(11):e22179 [FREE Full text] [doi: 10.2196/22179] [Medline: 33146627]
4. Lane TS, Armin J, Gordon JS. Online Recruitment Methods for Web-Based and Mobile Health Studies: A Review of the Literature. J Med Internet Res 2015 Jul 22;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
5. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012 Feb 01;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
6. Loxton D, Harris ML, Forder P, Powers J, Townsend N, Byles J, et al. Factors Influencing Web-Based Survey Response for a Longitudinal Cohort of Young Women Born Between 1989 and 1995. J Med Internet Res 2019 Mar 25;21(3):e11286 [FREE Full text] [doi: 10.2196/11286] [Medline: 30907739]
7. Keusch F. Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection. Manag Rev Q 2015 Jan 9;65(3):183-216. [doi: 10.1007/s11301-014-0111-y]
8. Fang J, Shao P, Lan G. Effects of innovativeness and trust on web survey participation. Computers in Human Behavior 2009 Jan;25(1):144-152. [doi: 10.1016/j.chb.2008.08.002]
9. Fan W, Yan Z. Factors affecting response rates of the web survey: A systematic review. Computers in Human Behavior 2010 Mar;26(2):132-139. [doi: 10.1016/j.chb.2009.10.015]
10. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and Efficiency of Online and Offline Recruitment Methods: A Web-Based Cohort Study. J Med Internet Res 2017 Mar 01;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]
11. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res 2016 Nov 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
12. Synnot A, Ryan R, Prictor M, Fetherstonhaugh D, Parker B. Audio-visual presentation of information for informed consent for participation in clinical trials. Cochrane Database Syst Rev 2014 May 09(5):CD003717 [FREE Full text] [doi: 10.1002/14651858.CD003717.pub3] [Medline: 24809816]
13. Visscher BB, Steunenberg B, Heijmans M, Hofstede JM, Devillé W, van der Heide I, et al. Evidence on the effectiveness of health literacy interventions in the EU: a systematic review. BMC Public Health 2018 Dec 29;18(1):1414 [FREE Full text] [doi: 10.1186/s12889-018-6331-7] [Medline: 30594180]
14. Lin Y, Chen C, Lee W, Cheng Y, Lin T, Lin C, et al. Educational video-assisted versus conventional informed consent for trauma-related debridement surgery: a parallel group randomized controlled trial. BMC Med Ethics 2018 Mar 09;19(1):23 [FREE Full text] [doi: 10.1186/s12910-018-0264-7] [Medline: 29523129]
15. Housten AJ, Kamath GR, Bevers TB, Cantor SB, Dixon N, Hite A, et al. Does Animation Improve Comprehension of Risk Information in Patients with Low Health Literacy? A Randomized Trial. Med Decis Making 2019 Dec 03;40(1):17-28. [doi: 10.1177/0272989x19890296]
16. Meppelink CS, van Weert JCM, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015 Jan 13;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
17. Brewer HR, Hirst Y, Sundar S, Chadeau-Hyam M, Flanagan JM. Cancer Loyalty Card Study (CLOCS): protocol for an observational case-control study focusing on the patient interval in ovarian cancer diagnosis. BMJ Open 2020 Sep 08;10(9):e037459 [FREE Full text] [doi: 10.1136/bmjopen-2020-037459] [Medline: 32900761]
18. Flanagan JM, Skrobanski H, Shi X, Hirst Y. Self-Care Behaviors of Ovarian Cancer Patients Before Their Diagnosis: Proof-of-Concept Study. JMIR Cancer 2019 Jan 17;5(1):e10447. [doi: 10.2196/10447] [Medline: 30664464]
19. The Cancer Loyalty Card Study (CLOCS). YouTube. 2020 Jan 15. URL: https://youtu.be/vr5mCxl7WIw [accessed 2022-08-11]
20. von Wagner C, Hirst Y, Waller J, Ghanouni A, McGregor LM, Kerrison RS, et al. The impact of descriptive norms on motivation to participate in cancer screening - Evidence from online experiments. Patient Educ Couns 2019 Sep;102(9):1621-1628 [FREE Full text] [doi: 10.1016/j.pec.2019.04.001] [Medline: 30975450]
21. Stoffel S, Hirst Y, Ghanouni A, McGregor LM, Kerrison R, Verstraete W, et al. Testing active choice for screening practitioner's gender in endoscopy among disinclined women: An online experiment. J Med Screen 2019 Jun;26(2):98-103 [FREE Full text] [doi: 10.1177/0969141318806322] [Medline: 30428750]
22. Stoffel S, Yang J, Vlaev I, von Wagner C. Testing the decoy effect to increase interest in colorectal cancer screening. PLoS ONE 2019 Mar 26;14(3):e0213668 [FREE Full text] [doi: 10.1371/journal.pone.0213668]
23. Bartlett G, Macgibbon B, Rubinowicz A, Nease C, Dawes M, Tamblyn R. The Importance of Relevance: Willingness to Share eHealth Data for Family Medicine Research. Front Public Health 2018 Sep 4;6:255 [FREE Full text] [doi: 10.3389/fpubh.2018.00255] [Medline: 30234095]

24. 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]
25. Park H, Lee E. Self-reported eHealth literacy among undergraduate nursing students in South Korea: a pilot study. Nurse Educ Today 2015 Feb;35(2):408-413. [doi: 10.1016/j.nedt.2014.10.022] [Medline: 25466791]
26. Park H, Moon M, Baeg JH. Association of eHealth literacy with cancer information seeking and prior experience with cancer screening. Comput Inform Nurs 2014 Sep;32(9):458-463. [doi: 10.1097/CIN.0000000000000077] [Medline: 25105588]
27. Average household income, UK: financial year ending 2020 (provisional). Office for National Statistics. 2020. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending2020provisional [accessed 2020-08-01]
28. Greszki R, Meyer M, Schoen H. The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. In: Online Panel Research: A Data Quality Perspective. Chichester, UK: Wiley; 2014:238-262.
29. Miller DP, Spangler JG, Case LD, Goff DC, Singh S, Pignone MP. Effectiveness of a web-based colorectal cancer screening patient decision aid: a randomized controlled trial in a mixed-literacy population. Am J Prev Med 2011 Jun;40(6):608-615 [FREE Full text] [doi: 10.1016/j.amepre.2011.02.019] [Medline: 21565651]
30. Jerant A, Kravitz RL, Sohler N, Fiscella K, Romero RL, Parnes B, et al. Sociopsychological tailoring to address colorectal cancer screening disparities: a randomized controlled trial. Ann Fam Med 2014 May 12;12(3):204-214 [FREE Full text] [doi: 10.1370/afm.1623] [Medline: 24821891]
31. Reynolds WW, Nelson RM. Risk perception and decision processes underlying informed consent to research participation. Soc Sci Med 2007 Nov;65(10):2105-2115. [doi: 10.1016/j.socscimed.2007.06.021] [Medline: 17689846]
32. Smith S, Raine R, Obichere A, Wolf MS, Wardle J, von Wagner C. The effect of a supplementary ('Gist-based') information leaflet on colorectal cancer knowledge and screening intention: a randomized controlled trial. J Behav Med 2015 Apr;38(2):261-272 [FREE Full text] [doi: 10.1007/s10865-014-9596-z] [Medline: 25253443]
33. Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. JAMA 2004 Oct 06;292(13):1593-1601. [doi: 10.1001/jama.292.13.1593] [Medline: 15467062]
34. Shaffer VA, Owens J, Zikmund-Fisher BJ. The effect of patient narratives on information search in a web-based breast cancer decision aid: an eye-tracking study. J Med Internet Res 2013 Dec 17;15(12):e273 [FREE Full text] [doi: 10.2196/jmir.2784] [Medline: 24345424]
35. UK government coronavirus lockdowns. Institute for Government. 2022. URL: https://www.instituteforgovernment.org.uk/charts/uk-government-coronavirus-lockdowns> [accessed 2022-05-10]
36. Hirst Y, Stoffel S, Brewer H, Timotijevic L, Raats M. Flanagan JM Understanding Public Attitudes and Willingness to Share Commercial Data for Health Research: A Survey Study. JMIR Preprints 2022 Jul 07:Preprint.
37. Patel MX, Doku V, Tennakoon L. Challenges in recruitment of research participants. Adv. psychiatr. treat 2018 Jan 02;9(3):229-238. [doi: 10.1192/apt.9.3.229]
38. Sully BGO, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials 2013 Jun 09;14(1):166 [FREE Full text] [doi: 10.1186/1745-6215-14-166] [Medline: 23758961]

Abbreviations

AOR: adjusted odds ratio
CLOCS: Cancer Loyalty Card Study
OR: odds ratio

Edited by A Mavragani; submitted 01.06.22; peer-reviewed by YK Lin, R Sampieri-Cabrera; comments to author 13.07.22; revised version received 21.07.22; accepted 31.07.22; published 26.08.22.

Please cite as:
Stoffel ST, Law JH, Kerrison R, Brewer HR, Flanagan JM, Hirst Y
Testing the Effectiveness of an Animated Decision Aid to Improve Recruitment of Control Participants in a Case-Control Study: Web-Based Experiment
J Med Internet Res 2022;24(8):e40015
URL: https://www.jmir.org/2022/8/e40015
doi:10.2196/40015
PMID:36018628

Original Paper

Evaluation of Medical Information on Male Sexual Dysfunction on Baidu Encyclopedia and Wikipedia: Comparative Study

Ming Ma[1], MD; Saifu Yin[1], MA; Mengli Zhu[2], MA; Yu Fan[1], MD; Xi Wen[3], BA; Tao Lin[1], MD; Turun Song[1], MD

[1]Department of Urology and Organ Transplantation Center, West China Hospital, Sichuan University, Chengdu, Sichuan Province, China
[2]Core Facilities, West China Hospital, Sichuan University, Chengdu, Sichuan Province, China
[3]Department of Urology, West China Hospital, Sichuan University, Chengdu, Sichuan Province, China

Corresponding Author:
Turun Song, MD
Department of Urology and Organ Transplantation Center
West China Hospital
Sichuan University
#37 Guoxue Alley
Chengdu, Sichuan Province, 610041
China
Phone: 86 028 854 229 25
Email: songturun1986@scu.edu.cn

Abstract

Background: Sexual dysfunction is a private set of disorders that may cause stigma for patients when discussing their private problems with doctors. They might also feel reluctant to initiate a face-to-face consultation. Internet searches are gradually becoming the first choice for people with sexual dysfunction to obtain health information. Globally, Wikipedia is the most popular and consulted validated encyclopedia website in the English-speaking world. Baidu Encyclopedia is becoming the dominant source in Chinese-speaking regions; however, the objectivity and readability of the content are yet to be evaluated.

Objective: Hence, we aimed to evaluate the reliability, readability, and objectivity of male sexual dysfunction content on Wikipedia and Baidu Encyclopedia.

Methods: The Chinese Baidu Encyclopedia and English Wikipedia were investigated. All possible synonymous and derivative keywords for the most common male sexual dysfunction, erectile dysfunction, premature ejaculation, and their most common complication, chronic prostatitis/chronic pelvic pain syndrome, were screened. Two doctors evaluated the articles on Chinese Baidu Encyclopedia and English Wikipedia. The Journal of the American Medical Association (JAMA) scoring system, DISCERN instrument, and Global Quality Score (GQS) were used to assess the quality of disease-related articles.

Results: The total DISCERN scores (P=.002) and JAMA scores (P=.001) for Wikipedia were significantly higher than those of Baidu Encyclopedia; there was no statistical difference between the GQS scores (P=.31) for these websites. Specifically, the DISCERN Section 1 score (P<.001) for Wikipedia was significantly higher than that of Baidu Encyclopedia, while the differences between the DISCERN Section 2 and 3 scores (P=.14 and P=.17, respectively) were minor. Furthermore, Wikipedia had a higher proportion of high total DISCERN scores (P<.001) and DISCERN Section 1 scores (P<.001) than Baidu Encyclopedia. Baidu Encyclopedia and Wikipedia both had low DISCERN Section 2 and 3 scores (P=.49 and P=.99, respectively), and most of these scores were low quality.

Conclusions: Wikipedia provides more reliable, higher quality, and more objective information than Baidu Encyclopedia. Yet, there are opportunities for both platforms to vastly improve their content quality. Moreover, both sites had similar poor quality content on treatment options. Joint efforts of physicians, physician associations, medical institutions, and internet platforms are needed to provide reliable, readable, and objective knowledge about diseases.

(J Med Internet Res 2022;24(8):e37339) doi:10.2196/37339

KEYWORDS

sexual dysfunction; digital health; Baidu Encyclopedia; Wikipedia; internet; health information; DISCERN instrument

Introduction

Knowledge regarding health and well-being is cobbled together from health care professionals, family, friends, books, newspapers, magazines, educational pamphlets, radio, television, and pharmaceutical advertisements [1]. However, we are increasingly heading online for answers rather than pursuing information through these other avenues [2]. Approximately 6% of all internet searches in the United States are health-related [3], and it is believed that internet searches have become people's first choice of method to seek information regarding health issues [4]. In addition, the population of netizens in mainland China reached 1011 million in 2021, and the number of online medical users in China had reached 239 million by June 2021, accounting for 23.7% of total internet users [5]. Information quality, emotional support, and source credibility have significant and positive impact on the likelihood of health care information adoption, and among these factors, information quality has the biggest impact on patients' adoption decisions [6]. Given the large amount of inaccurate information online, users are very easily misinformed [1]. Previous studies showed that the quality of online health information is problematic [7,8]. Thus, the assessment of source reputability and the veracity of information is a crucial and urgent task.

As the most common male sexual dysfunctions, erectile dysfunction (ED; the persistent inability to attain and maintain an erection sufficient to permit satisfactory sexual performance) and premature ejaculation (PE; poorly controlled and rapid ejaculation) greatly affect the quality of life of patients [9,10]. Furthermore, sexual dysfunction is closely associated with chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS; urologic pain or discomfort in the pelvic region associated with lower urinary tract symptoms) and is the most common complication [11,12]. The prevalence of CP/CPPS in men is about 8.2%, and men with CP/CPPS are more prone to ED and PE than the general population [13]. A previous study found that nearly half of patients with a self-reported diagnosis of CP/CPPS reported mild to severe ED [14]. A meta-analysis of 24 studies suggested that the overall prevalence of sexual dysfunction in patients with CP/CPPS was 0.62 [15]. In particular, our previous study found that "prostate" and "prostatitis" were the most queried terms by Chinese users with PE [16], which highlighted the stigma and preferences of these patients [17]. In addition, the complex and unclear etiology of CP/CPPS and sexual dysfunction not only challenges clinicians in the choice of treatment but also seriously affects the quality of life of patients. Previously, public interest and the change over time in the search volume for sexual dysfunctions and lower urinary tract symptoms were analyzed [16,18,19]. People tended to consult Dr. Internet in a combined manner on these issues for treatment decision-making. Therefore, the issue of sexual dysfunction is commonly investigated with CP/CPPS.

Wikipedia, the most popular and consulted encyclopedia website in English, is a web-based encyclopedia that provides valuable web-based health information [20]. Previous studies have shown that Wikipedia is a reasonably reliable medical resource and it was ranked higher on search engines than other general websites [21,22]. Unfortunately, on May 19, 2015, "Chinese Wikipedia" announced that mainland Chinese servers would be shut down because of violation of mainland China's laws due to the attack and destruction of the internet. As the equivalent Wikipedia for Chinese internet users, the Baidu platform and its Encyclopedia service is the most popular and frequently consulted encyclopedia site in mainland China [23,24]. In mainland China, with 766 million users actively using the Baidu search service, its usage in relation to health inquiries and symptom confirmation accounts for 66.83% of use, and health and medical topics ranked first among science topics [24]. Our previous research on the Baidu search index showed that the search demands by its users for sexual dysfunction and lower urinary tract symptoms are huge. However, users often get irrelevant online medical information, and there is little evaluation of the quality of Baidu-related content [16,19]. The purpose of this paper was to assess the reliability, readability, and objectivity of Wikipedia and Baidu Encyclopedia content on ED, PE, and CPPS/CP for the advancement of internet medicine.

Methods

Data Sources

The contents analyzed in this study are available on Chinese Baidu Encyclopedia and English Wikipedia. The Chinese Baidu Encyclopedia and English Wikipedia were investigated for articles on ICD-10 version 2016 codes. All possible synonymous and derivative keywords for each term were screened. Two doctors evaluated the articles on Chinese Baidu Encyclopedia and English Wikipedia. Any disagreement was reviewed by and arbitrated by a third reviewer who was an expert on sexual dysfunction. All authors have many years of experience in andrology and urology and are competent in the diagnosis and treatment of male sexual dysfunction and urinary disorders. These reviewers have professional knowledge of male sexual dysfunction and urinary disorders and can make professional evaluations.

Assessment of the Quality of the Research Articles

The Journal of the American Medical Association (JAMA) scoring system [25], DISCERN instrument [26], and Global Quality Score (GQS) [27] were used to assess the quality of disease-related articles. The contents of these scoring tables are shown in Multimedia Appendix 1. The JAMA scoring system is a well-known tool for evaluating the quality of information obtained from health-related websites. It includes 4 evaluation dimensions: author, attribution, disclosure, and currency. If it meets the requirements of each dimension, it will get 1 point, and the deimension with the highest quality will get 4 points. The DISCERN instrument has been developed to judge the quality of written health information [26]. To more comprehensively determine the quality of information in the article, the DISCERN tool consists of 15 questions plus an overall quality rating, and each is scored on a scale from 1 to 5. The first section of the DISCERN instrument is commonly used to evaluate the quality of published information, and the second section focuses on the quality of treatment choices offered to patients. The total score can range from 16 to 80, where a score of 63 to 80 suggests excellent quality, 51 to 62 indicates good quality, 39 to 50 indicates fair quality, and 16 to 38 indicates poor quality [26]. Experienced health information users and providers can use the DISCERN instrument to distinguish between high-quality and low-quality publications, so as to promote the generation of high-quality, evidence-based patient information. The GQS is a 5-point Likert scale that can subjectively rate the overall quality of each reviewed website. In addition to evaluating the overall quality of the website, GQS also considers the flow and ease of use of each website [28].

Statistical Analysis

All databases were constructed with Excel 2019 (Microsoft Corporation, Redmond, WA). The Shapiro-Wilk test was used to test the normality of the data. Descriptive analyses are reported as means and SDs for normally distributed variables. Medians and IQRs are reported for non-normally distributed variables. To ensure the quality of these scores, the intraclass correlation coefficient (ICC) was used to evaluate interobserver reliability. ICC values range from 0 (untrusted) to 1 (fully trusted), and any concordance values less than 0.75 were discussed by the research team to clarify the discrepancy. For nonparametric tests, the Mann-Whitney *U* test was conducted to test the significance of different ranks by using SPSS, version 22.0 (IBM Corp, Armonk, NY). The Fisher exact test was used to test the difference in the frequency distribution of DISCERN scores. We used Prism 8 for macOS, version 8.4.0 (455; GraphPad Software Inc, San Diego, CA) to conduct statistical analyses and create figures. For the statistical analysis, $P<.05$ was considered significant.

Results

Content Characteristics

We searched for “erectile dysfunction,” “premature ejaculation,” “chronic prostatitis/chronic pelvic pain syndrome,” and similar keywords on English Wikipedia and Chinese Baidu Encyclopedia. The search results are shown in Table 1. Wikipedia has only 1 entry for a disease, corresponding to a specific article. In Baidu Encyclopedia, a disease may have multiple entries and multiple articles. The information sources of these articles are different, and the number of views varies greatly. In Wikipedia, an article about a disease is constantly supplemented by different registered individuals. However, Baidu Encyclopedia's content providers are official organizations or unregistered individuals. Moreover, some of the recently updated articles in Baidu Encyclopedia show that the information is more often provided by organizations or institutions and is certified by experts. In addition, both Baidu and Wikipedia provide links to external information, including videos, articles, and images, while some links are unrelated advertisements. The latter especially appear in Baidu Encyclopedia. Furthermore, Wikipedia provides its own features for assessing the quality of articles, and all Wikipedia articles included in this study were rated as grade C.

Table 1. Characteristics of the search results from 2 online platforms.

Themes	Wikipedia	Baidu Encyclopedia	*P* value[a]
Available entries, n			
CPPS/CP[b]	2	3	N/A[c]
ED[d]	1	2	
PE[e]	1	3	
Real-time updates, n			
Yes	4	8	.99
No	0	0	
External links, n			
Yes	4	6	.52
No	0	2	
Advertisement, n			
Yes	0	4	.21
No	4	4	
Author type, n			
Organization	4	5	.49
Individuals	0	3	
Page views (x1000), median (IQR)	1673.2 (240.0-3878.9)	4119.7 (775.3-22029.8)	.37
Number of references, median (IQR)	53.5 (19.3-84.0)	0 (0.0-0.8)	.002

[a] A Mann-Whitney *U* test was conducted to test the significance of different ranks.

[b] CP/CPPS: chronic prostatitis/chronic pelvic pain syndrome.

[c] N/A: not applicable.

[d] ED: erectile dysfunction.

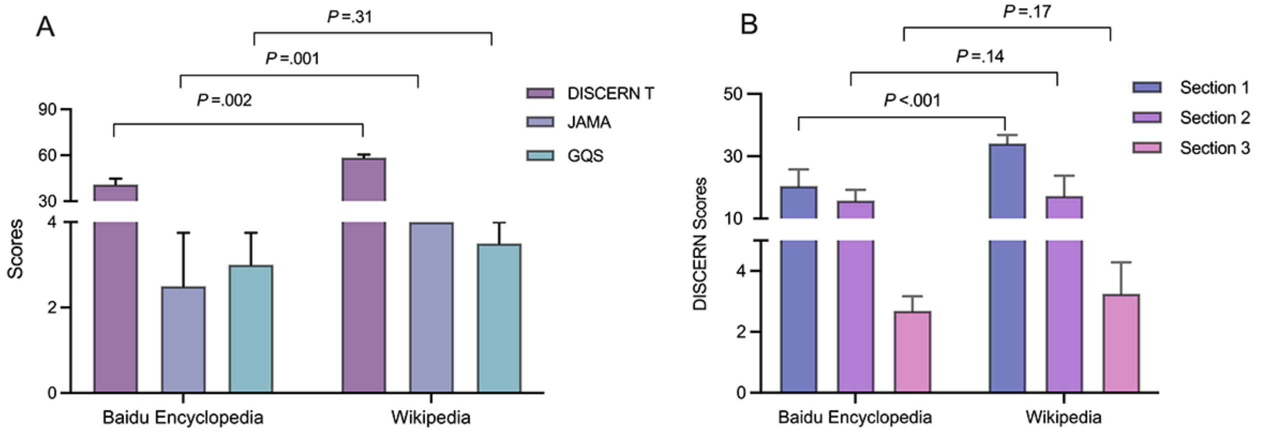
[e] PE, premature ejaculation.

Overall Scores for Baidu Encyclopedia and Wikipedia

A 2-way mixed/random effects model was used to analyze the consistency of the ratings by the 2 independent reviewers. The ICC results showed good consistency between the 2 reviewers for the GQS scores (ICC=0.87), JAMA scores (ICC=0.91), and DISCERN scores (ICC=0.82).

Comprehensively, the scores for Wikipedia were higher than those for Baidu Encyclopedia (Figure 1A). The contents in Wikipedia were significantly higher rated by the DISCERN tool and JAMA tool than those in Baidu Encyclopedia, suggesting that Wikipedia provides higher quality information. Although there was no statistical difference between the GQS scores for these websites, a numerically higher score on Wikipedia indicates that Wikipedia may provide better reading fluency and ease of use. In order to distinguish the differences between the 2 websites in more detail, we compared the DISCERN section scores for Baidu Encyclopedia and Wikipedia (Figure 1B). The DISCERN Section 1 score for Wikipedia was significantly higher than that for Baidu Encyclopedia, suggesting that Wikipedia provides more reliable and more objective information. The DISCERN Section 2 evaluates “How good is the quality of information regarding treatment choices?” There was no statistical difference between the DISCERN Section 2 scores for these websites, suggesting that they may have a similar impact on patients’ choice of treatment options. Section 3 is the overall rating of the publication, and the lack of statistical difference revealed that the overall quality of the publication as a source of information about treatment choices was similar for these websites.

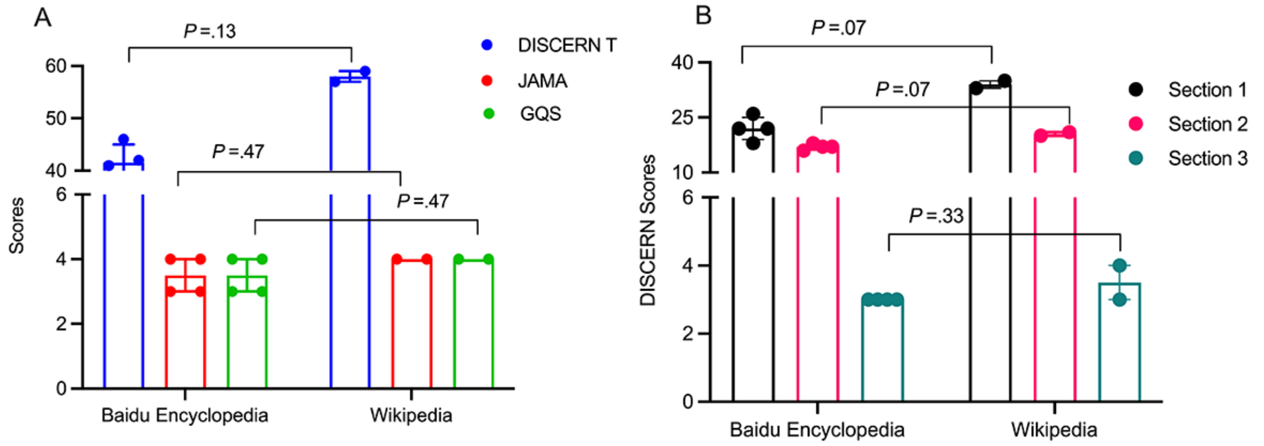
Figure 1. Overall comparison between Baidu Encyclopedia and Wikipedia: (A) median and IQR for DISCERN total scores, Journal of the American Medical Association (JAMA) scoring system scores, and Global Quality Score (GQS) scores; (B) median and IQR for the 3 DISCERN sections.



Overall Quality Comparison Between Wikipedia and Baidu Encyclopedia for the Theme of ED

ED is one of the most common male sexual dysfunctions. By comparing the content scores for ED articles on Baidu Encyclopedia and Wikipedia, Wikipedia appeared to have numerically higher total DISCERN scores, JAMA scores, and GQS scores, but there were no statistically significant differences (Figure 2A). Furthermore, the 3 DISCERN section scores for Baidu and Wikipedia were also compared separately (Figure 2B). Wikipedia appeared to have numerically higher DISCERN Section 1 and 2 scores. In addition, they had similar DISCERN Section 3 scores. These results suggest that there is no statistically significant difference between Wikipedia and Baidu Encyclopedia scores for ED content.

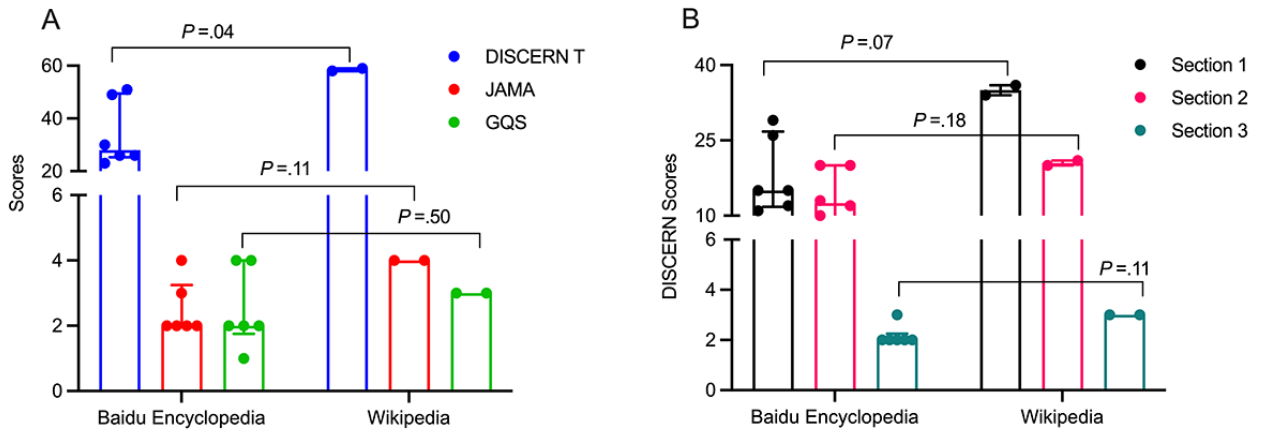
Figure 2. Comparison of erectile dysfunction (ED) scores between Baidu Encyclopedia and Wikipedia: (A) median and IQR for total DISCERN scores, Journal of the American Medical Association (JAMA) scoring system scores, and Global Quality Score (GQS); (B) median and IQR for the 3 DISCERN sections.



Overall Quality Comparison Between Wikipedia and Baidu Encyclopedia for the Theme of PE

A comparison of the scores for PE, the other most common sexual dysfunction disorder, showed that Wikipedia had a significantly higher total DISCERN score than Baidu Encyclopedia (Figure 3A). Although Wikipedia seemed to have higher JAMA and GQS scores than Baidu Encyclopedia (Figure 3A), this difference was not statistically significant, and all DISCERN section scores showed a similar trend (Figure 3B), which may be related to the great intragroup variability of Baidu Encyclopedia.

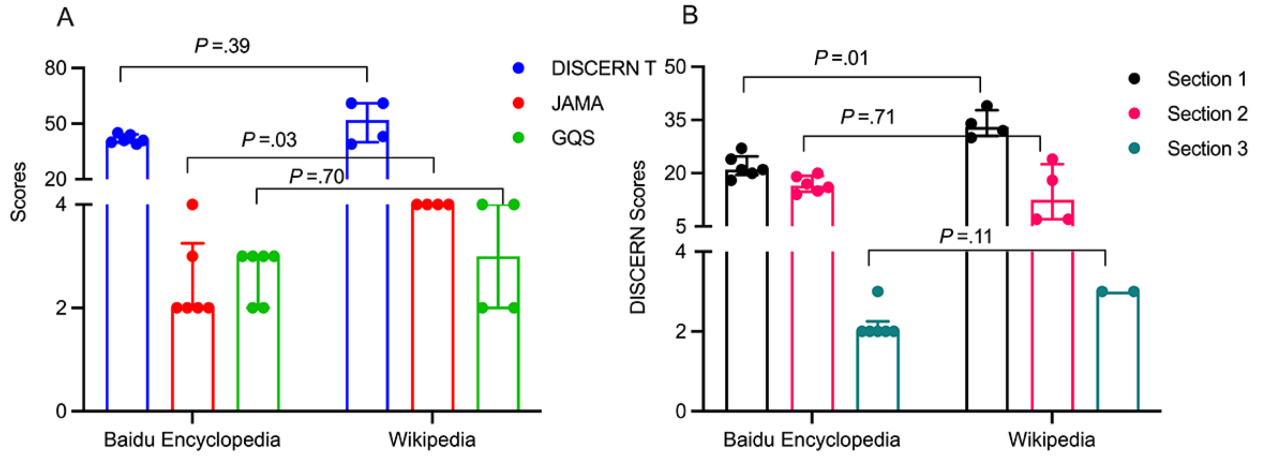
Figure 3. Comparison of premature ejaculation (PE) scores between Baidu Encyclopedia and Wikipedia: (A) median and IQR for total DISCERN scores, Journal of the American Medical Association (JAMA) scoring system scores, and Global Quality Score (GQS); (B) median and IQR for the 3 DISCERN sections.



Overall Quality Comparison Between Wikipedia and Baidu Encyclopedia for the Theme of CP/CPPS

CP/CPPS, as one of the most common concomitant diseases of sexual dysfunction, seriously affects the quality of life of male patients. By comparing the overall scores for Wikipedia and Baidu encyclopedia on CP/CPPS, we found that the scores of Baidu Encyclopedia were mostly fair quality, while the scores of Wikipedia ranged from fair quality to good quality (Figure 4A). Meanwhile, Wikipedia showed statistically higher JAMA scores, but there were no statistical differences between total DISCERN scores and GQS scores (Figure 4A). Furthermore, the DISCERN Section 1 score for Wikipedia was statistically significantly higher than that of Baidu Encyclopedia, while the DISCERN Section 2 and 3 scores for both sites were not significantly different from each other (Figure 4B).

Figure 4. Comparison of chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS) scores between Baidu Encyclopedia and Wikipedia: (A) median and IQR for total DISCERN scores, Journal of the American Medical Association (JAMA) scoring system scores, and Global Quality Score (GQS); (B) median and IQR for the 3 DISCERN sections.



Distribution of the DISCERN Scores

After comparing the overall quality of the information for different diseases on Baidu encyclopedia and Wikipedia, the overall scores for Wikipedia seemed to be higher than those of Baidu encyclopedia, but some scores only showed numerical differences without statistical significance. Nevertheless, the differences in the distribution of scores that had numerical differences were seemingly obvious. Therefore, we performed further statistical analyses of the score distributions for Wikipedia and Baidu Encyclopedia. As aforementioned, according to the DISCERN standard, a total DISCERN score <50 (near 60%) is fair or poor quality, while a score >50 is good or excellent quality [26]. Based on this rule, we took a score of 3 for each question as the cutoff value; that is, a score higher than 3 points was defined as good quality.

The score distributions for each disease are shown in Table 2. Wikipedia had a higher proportion of total DISCERN and Section 1 scores distributed above 3 points, whether compared with the overall score or the score for each disease, and was significantly better than Baidu Encyclopedia. However, Baidu Encyclopedia and Wikipedia had low Section 2 and 3 scores, and most of these scores were ≤3, which are defined as low quality.

Table 2. Distribution of the DISCERN scores for each disease and comparisons via the Fisher exact test.

DISCERN	Overall			CP/CPPS[a]			ED[b]			PE[c]		
	Wikipedia, n (%)	Baidu Encyclopedia, n (%)	*P* value	Wikipedia, n (%)	Baidu Encyclopedia, n (%)	*P* value	Wikipedia, n (%)	Baidu Encyclopedia, n (%)	*P* value	Wikipedia, n (%)	Baidu Encyclopedia, n (%)	*P* value
Total												
>3	38 (59.4)[d]	25 (19.5)[e]	<.001	17 (53.1)[f]	10 (20.8)[g]	.004	10 (62.5)[h]	7 (21.9)[f]	<.001	11 (68.8)[h]	8 (16.7)[g]	<.001
≤3	26 (40.6)[d]	103 (80.5)[e]		15 (46.9)[f]	38 (79.2)[g]		6 (37.5)[h]	25 (78.1)[f]		5 (31.2)[h]	40 (83.3)[g]	
Section 1												
>3	31 (96.9)[f]	19 (29.7)[d]	<.001	15 (93.8)[h]	7 (29.2)[i]	<.001	8 (100)[j]	6 (37.5)[h]	.006	8 (100)[j]	6 (25.0)[i]	<.001
≤3	1 (3.1)[f]	45 (70.3)[d]		1 (6.2)[h]	17 (70.8)[i]		0 (0)[j]	10 (62.5)[h]		0 (0)[j]	18 (75.0)[i]	
Section 2												
>3	5 (17.9)[k]	6 (10.7)[l]	.49	1 (7.1)[m]	3 (14.3)[n]	.64	1 (14.3)[o]	1 (7.1)[m]	.99	3 (42.9)[o]	2 (9.5)[n]	.08
≤3	23 (82.1)[k]	50 (89.3)[l]		13 (92.9)[m]	18 (85.7)[n]		6 (85.7)[o]	13 (92.9)[m]		4 (57.1)[o]	19 (90.5)[n]	
Section 3												
>3	2/4 (50.0)[p]	3 (37.5)[j]	.99	1 (50.0)[q]	0 (0)[r]	.40	1 (100)[s]	0 (0)[q]	.33	0 (0)[s]	0 (0)[r]	.99
≤3	2 (50.0)[p]	5 (62.5)[j]		1 (50.0)[q]	3 (100)[r]		0 (0)[s]	2 (100)[q]		1 (100)[s]	3 (100)[r]	

[a] CP/CPPS: chronic prostatitis/chronic pelvic pain syndrome.

[b] ED: erectile dysfunction.

[c] PE: premature ejaculation.

[d] n=64.

[e] n=128.

[f] n=32.

[g] n=48.

[h] n=16.

[i] n=24.

[j] n=8.

[k] n=28.

[l] n=56.

[m] n=14.

[n] n=21.

[o] n=7.

[p] n=4.

[q] n=2.

[R] n=3.

[S] n=1.

Discussion

Principal Findings

Internet-based information is playing an increasingly important role in the diagnosis and treatment of patients, especially for privacy-sensitive conditions such as sexual dysfunction and related concomitant diseases. Comprehensive and objective information can help patients understand their condition, choose the right time to visit a doctor, and then improve their prognosis. However, incorrect or incomplete information may leave patients vulnerable to misdiagnosis, leading to delays in treatment and considerable health risks [1]. As a consequence, at a time when internet health care is booming, there is an urgent need to evaluate the credibility, readability, and accuracy of online resources. This study evaluated the reliability, readability, and objectivity of Baidu Encyclopedia and Wikipedia in terms of ED, PE, and CP/CPPS content. Overall, the total DISCERN scores and DISCERN Section 1 scores for the content provided by Wikipedia were significantly higher than those of Baidu Encyclopedia. Also, Wikipedia had a higher proportion of total DISCERN and Section 1 scores distributed within the high-quality range than Baidu Encyclopedia. Combined with higher JAMA scores, the results suggest that Wikipedia provided more reliable, higher quality, and more objective information than Baidu Encyclopedia. Baidu Encyclopedia and Wikipedia had low DISCERN Section 2 and 3 scores, and most of these scores were low quality. Similar DISCERN Section 2 and 3 scores for Wikipedia and Baidu Encyclopedia indicated that they had an analogic and mediocre impact on patients' choice of treatment options. Although not statistically different, Wikipedia had numerically higher GQS scores, suggesting that Wikipedia might provide relatively better flow and be easier to use.

By June 2021, the number of online medical users in China was 239.33 million, and the utilization rate of the internet was 23.7%, an increase of 11.4% over December 2020 [29]. In an analysis of internet search trends in China, some scholars found that only 43.74% of the search results for PE were related to PE [16]. In another study on lower urinary tract symptoms, 1.13%-93.92% of the retrieved content was found to be irrelevant to lower urinary tract symptoms [19]. The study also found similar problems in the contents about these diseases in Wikipedia and Baidu Encyclopedia. Wikipedia provides more standardized and unified content, with standard templates for almost every disease, which allows readers to find the information they need quickly and accurately [30]. In contrast, the quality of content provided by Baidu Encyclopedia varies widely, with some recently updated articles providing more comprehensive content than Wikipedia, but the overall trend is a lack of standardization and formality. In Baidu Encyclopedia, the same disease may correspond to multiple entries and corresponding articles, which compare poorly with each other, and different articles may provide users with contradictory information, which can cause great confusion to users. The diversity of the content formats presented by Baidu Encyclopedia is consistent with the great variability of its overall score. The total DISCERN scores and JAMA scores for Wikipedia were significantly higher than those for Baidu Encyclopedia, and the proportion of Wikipedia scores within the high-quality distribution was also higher than those for Baidu Encyclopedia. These results suggest that Wikipedia provides higher quality information than Baidu Encyclopedia. In addition to the lack of a standard content presentation format, the low quality of Baidu Encyclopedia is also related to other features of its website, such as information sources and references. The contents of Baidu Encyclopedia are mostly sourced from official organizations or unregistered individuals, while information on Wikipedia is provided by registered users. The comparison shows that the quality of contents provided by unregistered individuals is always rated as "poor quality." Accurate citation of high-quality references is an important guarantee for the reliability of a paper [31]. The contents provided by these unregistered individual users are almost always without references and extended information. By contrast, the quality of contents provided by registered users or official organizations are almost rated as "good quality," with accurate references. These characteristics of the website are closely related to DISCERN Section 1 scores, and significantly higher DISCERN Section 1 scores for Wikipedia indicate that its publications are more reliable than those of Baidu Encyclopedia. The other 2 main focuses of the quality assessment are "How good is the quality of information regarding treatment choices?" and "the overall quality of the publication as a source of information about treatment choices." Similar scores on DISCERN Sections 2 and 3 for Wikipedia and Baidu Encyclopedia indicated that they had an analogic and mediocre impact on patients' choice of treatment options. Recent updates to Baidu Encyclopedia also show an increasing number of medical professionals involved in reviewing or writing the content, also significantly improving the DISCERN and JAMA scores. This comparison suggests that the inconsistency of disease presentation formats and differences in information sources may account for the lower Baidu scores.

CP/CPPS is characterized by localized pain or discomfort in the abdomen, pelvis, and genitals, usually with lower urinary tract symptoms, psychosocial disorders, and sexual dysfunction [11,12]. The relationship between sexual dysfunctions and CP/CPPS has been studied more extensively [32]. Previous studies have shown a good correlation between the severity of symptom scores between the 2 clinical conditions, CP and PE, and that approximately 49% of male patients with CP have concomitant sexual dysfunction [33]. It addition, "prostate" and "prostatitis" were the most queried terms by Chinese users with PE [16]. The complex and heterogeneous pathophysiology of CP/CPPS makes the management of this troublesome situation very challenging both for clinicians and patients, and approximately 50% of older patients experience recurrence [34]. The UPOINT System classifies CP/CPPS patients into 7 different subgroups based on symptoms: urologic, psychosocial, organ-specific, infectious, neurologic, tenderness (pelvic floor tenderness), and sexual dysfunction; then, it proposes specific treatment plans based on the different subgroups [35]. There is growing evidence that the addition of second-line therapies, such as 5-phosphodiesterase inhibitors, antidepressants and muscle relaxants, according to the UPOINT System approach, can significantly improve patients' CP/CPPS symptoms [36]. These results showed that CP/CPPS and sexual dysfunctions can directly or indirectly increase the economic burden of health care and seriously affect patients' quality of life. Patients with CP/CPPS or sexual dysfunction may feel too embarrassed to discuss their problems with doctors due to the influence of the Chinese culture, and they are likely more willing to look for disease-related information, such as symptoms, diagnosis, treatment methods, prognosis, and hospital rankings, on the internet first. There is no doubt that the information these patients access from the internet affects their perception of their health status, which in turn affects treatment choices and disease prognosis.

By comparing the contents for ED, PE, and CP/CPPS on Baidu Encyclopedia and Wikipedia, we found that the consistency of Wikipedia is better, with almost all content rated as "good quality," while the scores for Baidu Encyclopedia were mostly "fair quality." Take PE-related articles in Baidu Encyclopedia as examples. Both reviewers rated "早发性射精" (early-onset ejaculation) as “poor quality.” After analyzing the content on the web page for “early-onset ejaculation,” we found there was no introduction to “examination, diagnosis, and treatment,” and the content in the article was not objective and scientific. Contrary to the lack of effective information, there are more than 25 irrelevant advertising links and only one reference on this web page. The content on the “早泄” (premature ejaculation) page on Baidu Encyclopedia was rated as “good quality,” and the information was more comprehensive and objective than that for “early-onset ejaculation.” Corresponding to the quality grades for “early-onset ejaculation” and “premature ejaculation,” there was a huge difference in page views (early-onset ejaculation/premature ejaculation: 33,506/25,747,398). The discrepancy may be related to the inconsistent identity of content providers. The irrelevant advertising links or misleading information obtained by users using Baidu Encyclopedia may be related to the fee-based editing service. There are many third-party underground industries that charge fees to write Baidu Encyclopedia entries on their behalf, so as to insert advertisements and achieve the purpose of attracting patients. In order to improve the quality of the health information, Baidu Encyclopedia announced the “rainbow plan” on December 9, 2012, wherein all medical entries could only be edited and revised by certified medical experts [37]. This is consistent with the findings of this study that an increasing number of medical professionals are involved in reviewing or writing content for Baidu Encyclopedia. Consequently, attracting, encouraging, and even recruiting more medical professionals to draft or proofread the content about disease presentation provided on these websites may ensure the content is objective and comprehensive. At the same time, the Baidu Encyclopedia platform should strengthen content regulation and establish a review mechanism to remove interest-related content.

In contrast, Wikipedia has its own content quality evaluation system, such as the “Wiki-Project article quality grading scheme” and the “Wiki-Project priority assessments” [38,39]. In this study, all included Wikipedia articles were rated as grade C, which means “Useful to a casual reader, but would not provide a complete picture for even a moderately detailed study” and “Considerable editing is needed to close gaps in content and solve cleanup problems.” The “Wiki grading” for these Wikipedia articles is similar to the grading by the 3 grading tools applied in this paper. That is, the quality of these Wikipedia articles is almost “good quality” but far from “excellent quality,” and all articles needed further improvement. Despite this fact, the formality and drafting on Wikipedia are better because of the clear attribution and disclosure it provides. As mentioned earlier, there is a lack of uniform standards for writing Baidu Encyclopedia content, many of the information sources are not supported by academic references, and external links are mostly related to advertisements. Hence, though the content on both sites leaves much to be desired, as a source to popularize science, the content on Wikipedia could at least guide interested individuals to the right source of informations, while Baidu Encyclopedia is more likely to provide misleading information.

In the era of rapid internet development, more patients have started to try online consultations [40]. This change in mode of treatment has presented new opportunities and challenges for doctors, medical institutions, physician associations, internet platforms, and patients. In this study, we evaluated the objectivity, reliability, and readability of the content on sexual dysfunction and CP/CPPS on Baidu and Wikipedia and found that the quality of the content provided by both sites was not “excellent quality” and needed to be improved. This study is only a microcosm of the vast amount of information available in internet-based health care. Considering the increasing coverage of the internet, more users will be influenced by internet-based information, and incorrect or incomplete information will have a negative impact on users’ decision-making. Therefore, we believe that, in the era of the internet information explosion, physicians, physician associations, and medical institutions should make full use of their expertise and become more involved in the construction of internet-based health care by providing objective and comprehensive content. Internet platforms, on the other hand, should strengthen the regulation and review of medical-related content and remove false or irrelevant content. Wikipedia already has a relatively complete self-censorship system and self-evaluation system, but Baidu Encyclopedia has almost no achievements in this regard. In China, the country with the world’s largest population, the importance of popular science education for the whole society and the world is self-evident. Baidu Encyclopedia, as the largest platform for online science education in China, still needs to be greatly enhanced to take up the corresponding social responsibility. Through the joint efforts of physicians and the platform, we hope to achieve the goal of providing users with timely access to correct, objective, comprehensive, and valid information when seeking medical advice or searching for health science content on the internet.

Limitations

Some limitations must be addressed in this study. This study only presents the results of medical professionals’ evaluations of health-related science content on the internet, and further research is needed on the specific impact of this information on the audience and readers. Since information on the internet is updated quickly, there may be some bias between the study results and the actual situation, and the data need to be updated in real time to ensure that the findings are true and valid. In addition, the difference in the number of Chinese and English entries indicates the information received by users will be significantly different because of the entries they choose to click. Therefore, our “combined” evaluation cannot fully represent the quality of the information they really receive. Fortunately, with the availability of infodemiology research, academics can combine content analysis and infodemiology search trends to better elucidate the impact of health-related information on the internet on users, society, and the health care industry.

Conclusions

Internet medicine, as a new medical model in the new era, provides strong support for users to understand disease information and choose the timing of treatment in a timely manner. Although it is more formally composited, Wikipedia also provides more reliable, higher quality, and more objective information than Baidu Encyclopedia. They also have a similar impact on patients' choice of treatment options, and the websites are similar in terms of flow and ease of use. To promote the healthy and sustainable development of internet health care, the joint efforts of physicians, physician associations, medical institutions, and internet platforms are needed to provide more reliable, accessible, and comprehensible disease knowledge to the public.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant numbers 81870513, .81600584, 81470980), 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (grant number ZY2016104).

Authors' Contributions

This study was conceptualized by MM and TL. The methodology was designed by MM, YF, and TS. The investigation was carried out by MM and YF. MM and SY carried out the data statistics and interpretation. MM and MZ wrote the original draft. MM, YF, TS, and TL reviewed and edited the draft. Funding was acquired by TL. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Scoring tables for the Journal of the American Medical Association (JAMA) scoring system, DISCERN instrument, and Global Quality Score (GQS).
[PDF File (Adobe PDF File), 137 KB - jmir_v24i8e37339_app1.pdf]

References

1. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41(1):433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
2. Johnson S, Park H, Gross C, Yu J. Use of alternative medicine for cancer and its impact on survival. J Natl Cancer Inst 2018 Jan 01;110(1):1. [doi: 10.1093/jnci/djx145] [Medline: 28922780]
3. Asch JM, Asch DA, Klinger EV, Marks J, Sadek N, Merchant RM. Google search histories of patients presenting to an emergency department: an observational study. BMJ Open 2019 Feb 20;9(2):e024791 [FREE Full text] [doi: 10.1136/bmjopen-2018-024791] [Medline: 30787088]
4. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
5. The 48th "Statistical Report on Internet Development in China". China Internet Network Information Center. 2021 Sep 15. URL: https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202109/t20210915_71543.htm [accessed 2022-02-11]
6. Jin J, Yan X, Li Y, Li Y. How users adopt healthcare information: An empirical study of an online Q&A community. Int J Med Inform 2016 Feb;86:91-103. [doi: 10.1016/j.ijmedinf.2015.11.002] [Medline: 26616406]
7. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
8. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
9. Salonia A, Bettocchi C, Boeri L, Capogrosso P, Carvalho J, Cilesiz N, EAU Working Group on Male Sexual and Reproductive Health. European Association of Urology Guidelines on Sexual and Reproductive Health-2021 Update: Male Sexual Dysfunction. Eur Urol 2021 Sep;80(3):333-357. [doi: 10.1016/j.eururo.2021.06.007] [Medline: 34183196]
10. Wei S, Wu C, Yu B, Ma M, Qin F, Yuan J. Advantages and limitations of current premature ejaculation assessment and diagnostic methods: a review. Transl Androl Urol 2020 Apr;9(2):743-757 [FREE Full text] [doi: 10.21037/tau.2019.12.08] [Medline: 32420180]
11. Krieger JN, Nyberg L, Nickel JC. NIH consensus definition and classification of prostatitis. JAMA 1999 Jul 21;282(3):236-237. [doi: 10.1001/jama.282.3.236] [Medline: 10422990]
12. Magistro G, Wagenlehner FM, Grabe M, Weidner W, Stief CG, Nickel JC. Contemporary management of chronic prostatitis/chronic pelvic pain syndrome. Eur Urol 2016 Feb;69(2):286-297. [doi: 10.1016/j.eururo.2015.08.061] [Medline: 26411805]
13. Tran CN, Shoskes DA. Sexual dysfunction in chronic prostatitis/chronic pelvic pain syndrome. World J Urol 2013 Aug;31(4):741-746. [doi: 10.1007/s00345-013-1076-5] [Medline: 23579441]

14. Matta R, Doiron C, Leveridge MJ. The dramatic increase in social media in urology. J Urol 2014 Aug;192(2):494-498. [doi: 10.1016/j.juro.2014.02.043] [Medline: 24576656]
15. Li H, Kang D. Prevalence of sexual dysfunction in men with chronic prostatitis/chronic pelvic pain syndrome: a meta-analysis. World J Urol 2016 Jul 6;34(7):1009-1017 [FREE Full text] [doi: 10.1007/s00345-015-1720-3] [Medline: 26546073]
16. Wei S, Ma M, Wen X, Wu C, Zhu G, Zhou X. Online public attention toward premature ejaculation in mainland China: infodemiology study using the Baidu Index. J Med Internet Res 2021 Aug 26;23(8):e30271 [FREE Full text] [doi: 10.2196/30271] [Medline: 34435970]
17. Russo GI, di Mauro M, Cocci A, Cacciamani G, Cimino S, Serefoglu EC, EAU-YAU Men's Health Working Group. Consulting "Dr Google" for sexual dysfunction: a contemporary worldwide trend analysis. Int J Impot Res 2020 Jul 07;32(4):455-461. [doi: 10.1038/s41443-019-0203-2] [Medline: 31591474]
18. Hui J, He S, Liu R, Zeng Q, Zhang H, Wei A. Trends in erectile dysfunction research from 2008 to 2018: a bibliometric analysis. Int J Impot Res 2020 Jul;32(4):409-419 [FREE Full text] [doi: 10.1038/s41443-019-0161-8] [Medline: 31235897]
19. Wei S, Ma M, Wu C, Yu B, Jiang L, Wen X, et al. Using search trends to analyze web-based interest in lower urinary tract symptoms-related inquiries, diagnoses, and treatments in mainland China: infodemiology study of Baidu Index data. J Med Internet Res 2021 Jul 06;23(7):e27029 [FREE Full text] [doi: 10.2196/27029] [Medline: 34255683]
20. Weiner SS, Horbacewicz J, Rasberry L, Bensinger-Brody Y. Improving the quality of consumer health information on Wikipedia: case series. J Med Internet Res 2019 Mar 18;21(3):e12450 [FREE Full text] [doi: 10.2196/12450] [Medline: 30882357]
21. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]
22. Thomas G, Eng L, de Wolff JF, Grover S. An evaluation of Wikipedia as a resource for patient education in nephrology. Semin Dial 2013;26(2):159-163. [doi: 10.1111/sdi.12059] [Medline: 23432369]
23. Sun F, Yang F, Zheng S. Evaluation of the liver disease information in Baidu Encyclopedia and Wikipedia: longitudinal study. J Med Internet Res 2021 Jan 18;23(1):e17680 [FREE Full text] [doi: 10.2196/17680] [Medline: 33459597]
24. Report on Chinese Netizens' Search Behavior for Popular Science Demand (Q1 2019). China Association of Science and Technology. 2019 Apr 26. URL: https://www.cast.org.cn/art/2019/4/26/art_1281_94546.html [accessed 2022-02-11]
25. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [doi: 10.1001/jama.1997.03540390074039] [Medline: 9103351]
26. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb 01;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
27. Langille M, Bernard A, Rodgers C, Hughes S, Leddin D, van Zanten SV. Systematic review of the quality of patient information on the internet regarding inflammatory bowel disease treatments. Clin Gastroenterol Hepatol 2010 Apr;8(4):322-328. [doi: 10.1016/j.cgh.2009.12.024] [Medline: 20060070]
28. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007 Sep;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
29. Yacob M, Lotfi S, Tang S, Jetty P. Wikipedia in vascular surgery medical education: comparative study. JMIR Med Educ 2020 Jun 19;6(1):e18076 [FREE Full text] [doi: 10.2196/18076] [Medline: 32417754]
30. Good BM, Clarke EL, Loguercio S, Su AI. Building a biomedical semantic network in Wikipedia with Semantic Wiki Links. Database (Oxford) 2012 Mar 20;2012:bar060-bar060 [FREE Full text] [doi: 10.1093/database/bar060] [Medline: 22434829]
31. Heilman JM, West AG. Wikipedia and medicine: quantifying readership, editors, and the significance of natural language. J Med Internet Res 2015 Mar 04;17(3):e62 [FREE Full text] [doi: 10.2196/jmir.4069] [Medline: 25739399]
32. McMahon CG, Jannini EA, Serefoglu EC, Hellstrom WJG. The pathophysiology of acquired premature ejaculation. Transl Androl Urol 2016 Aug;5(4):434-449 [FREE Full text] [doi: 10.21037/tau.2016.07.06] [Medline: 27652216]
33. Liang C, Zhang X, Hao Z, Shi H, Wang K. Prevalence of sexual dysfunction in Chinese men with chronic prostatitis. BJU Int 2004 Mar;93(4):568-570. [doi: 10.1111/j.1464-410x.2003.04662.x] [Medline: 15008731]
34. Clemens JQ, Meenan RT, O'Keeffe Rosetti MC, Kimes T, Calhoun EA. Prevalence of and risk factors for prostatitis: population based assessment using physician assigned diagnoses. J Urol 2007 Oct;178(4 Pt 1):1333-1337. [doi: 10.1016/j.juro.2007.05.140] [Medline: 17706722]
35. Magri V, Boltri M, Cai T, Colombo R, Cuzzocrea S, De Visschere P, et al. Multidisciplinary approach to prostatitis. Arch Ital Urol Androl 2019 Jan 18;90(4):227-248 [FREE Full text] [doi: 10.4081/aiua.2018.4.227] [Medline: 30655633]
36. Magri V, Marras E, Restelli A, Wagenlehner F, Perletti G. Multimodal therapy for category III chronic prostatitis/chronic pelvic pain syndrome in UPOINTS phenotyped patients. Exp Ther Med 2015 Mar;9(3):658-666 [FREE Full text] [doi: 10.3892/etm.2014.2152] [Medline: 25667610]
37. Baidu: Medical entries will not be edited by ordinary people. BJNEWS. 2012 Dec 10. URL: https://www.bjnews.com.cn/finance/2012/12/10/238299.html [accessed 2022-02-11]

38. WikiProject article quality grading scheme. Wikipedia. URL: https://en.wikipedia.org/wiki/Wikipedia:Content_assessment [accessed 2022-02-11]
39. WikiProject priority assessments. Wikipedia. URL: https://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Release_Version_Criteria [accessed 2022-02-11]
40. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec 17;34(4):268-283 [FREE Full text] [doi: 10.1111/hir.12192] [Medline: 29045011]

Abbreviations

CP/CPPS: chronic prostatitis/chronic pelvic pain syndrome
ED: erectile dysfunction
GQS: Global Quality Score
ICC: intraclass correlation coefficient
JAMA: Journal of the American Medical Association
PE: premature ejaculation

Edited by JMIRPE Office; submitted 16.02.22; peer-reviewed by A Montilha, S Wei, B Yu, X Zhou; comments to author 03.04.22; revised version received 17.05.22; accepted 05.07.22; published 09.08.22.

Please cite as:
Ma M, Yin S, Zhu M, Fan Y, Wen X, Lin T, Song T
Evaluation of Medical Information on Male Sexual Dysfunction on Baidu Encyclopedia and Wikipedia: Comparative Study
J Med Internet Res 2022;24(8):e37339
URL: https://www.jmir.org/2022/8/e37339
doi:10.2196/37339
PMID:35943768

Original Paper

Patients' Willingness to Provide Their Clinical Data for Research Purposes and Acceptance of Different Consent Models: Findings From a Representative Survey of Patients With Cancer

Anja Köngeter[1], MA; Christoph Schickhardt[2], DPhil; Martin Jungkunz[2], DPhil; Susanne Bergbold[3], Dipl; Katja Mehlis[1*], DPhil; Eva C Winkler[1*], DPhil, Prof Dr

[1]Section for Translational Medical Ethics, Department of Medical Oncology, National Center for Tumor Diseases, Heidelberg University Hospital, Heidelberg, Germany

[2]Section for Translational Medical Ethics, National Center for Tumor Diseases, German Cancer Research Center, Heidelberg, Germany

[3]Epidemiological Cancer Registry Baden-Württemberg, German Cancer Research Center, Heidelberg, Germany

[*]these authors contributed equally

Corresponding Author:
Eva C Winkler, DPhil, Prof Dr
Section for Translational Medical Ethics
Department of Medical Oncology
National Center for Tumor Diseases, Heidelberg University Hospital
Im Neuenheimer Feld 280
Heidelberg, 69120
Germany
Phone: 49 6221 5636049
Email: eva.winkler@nct-heidelberg.de

Abstract

Background: Secondary use of clinical data for biomedical research purposes holds great potential for various types of noninterventional, data-driven studies. Patients' willingness to support research with their clinical data is a crucial prerequisite for research progress.

Objective: The aim of the study was to learn about patients' attitudes and expectations regarding secondary use of their clinical data. In a next step, our results can inform the development of an appropriate governance framework for secondary use of clinical data for research purposes.

Methods: A questionnaire was developed to assess the willingness of patients with cancer to provide their clinical data for biomedical research purposes, considering different conditions of data sharing and consent models. The Cancer Registry of the German federal state of Baden-Württemberg recruited a proportionally stratified random sample of patients with cancer and survivors of cancer based on a full census.

Results: In total, 838 participants completed the survey. Approximately all participants (810/838, 96.7%) showed general willingness to make clinical data available for biomedical research purposes; however, they expected certain requirements to be met, such as comparable data protection standards for data use abroad and the possibility to renew consent at regular time intervals. Most participants (620/838, 73.9%) supported data use also by researchers in commercial companies. More than half of the participants (503/838, 60%) were willing to give up control over clinical data in favor of research benefits. Most participants expressed acceptance of the broad consent model (494/838, 58.9%), followed by data use by default (with the option to opt out at any time; 419/838, 50%); specific consent for every study showed the lowest acceptance rate (327/838, 39%). Patients expected physicians to share their data (763/838, 91.1%) and their fellow patients to support secondary use with their clinical data (679/838, 81%).

Conclusions: Although patients' general willingness to make their clinical data available for biomedical research purposes is very high, the willingness of a substantial proportion of patients depends on additional requirements. Taking these perspectives into account is essential for designing trustworthy governance of clinical data reuse and sharing. The willingness to accept the loss of control over clinical data to enhance the benefits of research should be given special consideration.

(J Med Internet Res 2022;24(8):e37665) doi:10.2196/37665

KEYWORDS

secondary use; consent; data sharing; data access; research benefit and control of data; health data; clinical data; private sector; international data sharing; patient perspective

Introduction

Background

Secondary use of clinical data for biomedical research purposes has great potential for various types of noninterventional, data-driven studies. We define secondary use of clinical data as the collection and reuse of clinical data in data gathering, noninterventional biomedical research, or learning activities; clinical data are collected during and for the purpose of patient care [1]. Research using clinical data has the ethical and efficiency advantages of not requiring additional physical interventions or collection of additional data. Although secondary use aims at improving biomedical knowledge and, in turn, medical care, it does not imply a direct benefit for the patient who has released their data.

The blurring of the boundaries between research and care, as envisaged in concepts of learning health care systems, is currently visible only in few areas [1,2]. The endeavor to merge these different system logics is faced with emerging challenges such as limited utility of specific consent models for research or false expectations regarding their benefits on the part of patients [3]. The goal of this paper was to contribute the patients' perspective to the debate and potential solutions to the current challenges of secondary use of clinical data in the context of learning health care systems.

Previous studies with citizens and patients have already shown that certain aspects seem to be crucial for supporting secondary use, such as who conducts the research (eg, academic or commercial), whether data are transferred to other countries, and what consent model is applied [2-9]. However, owing to varying research designs, for example, by examining different study units, applying different survey instruments, and being conducted in diverse health care systems, these studies, taken together, have heterogeneous results.

Consent is a crucial component of respecting patient autonomy and building trust in health research. However, the specific consent paradigm of clinical trials cannot easily be applied to the secondary use of clinical data because most scientific questions are unknown at the time consent is obtained, that is, when the patient receives care. Newly applied models for secondary use of clinical data, such as broad consent or data use by default (with the option to opt out any time), facilitate research with clinical data, but are criticized from an informational self-determination perspective for offering patients insufficient control over their clinical data. However, previous studies have identified patients' and citizens' openness toward these new models [5,10-12]. Other empirical studies show that, to increase research benefits, participants seem willing to accept the loss of control over their data [13-16]. However, no studies have yet been conducted to assess the acceptance of consent models in light of the trade-off between the control of clinical data and research utility.

Aim

The objective of this study was to assess (1) patients' general willingness and relevant requirements to share pseudonymized clinical data for research purposes, (2) acceptance of different consent models including characteristics of data control and research utility, (3) preferences regarding the setting to provide consent, and (4) general expectations toward data use and other stakeholders.

Methods

Survey Development

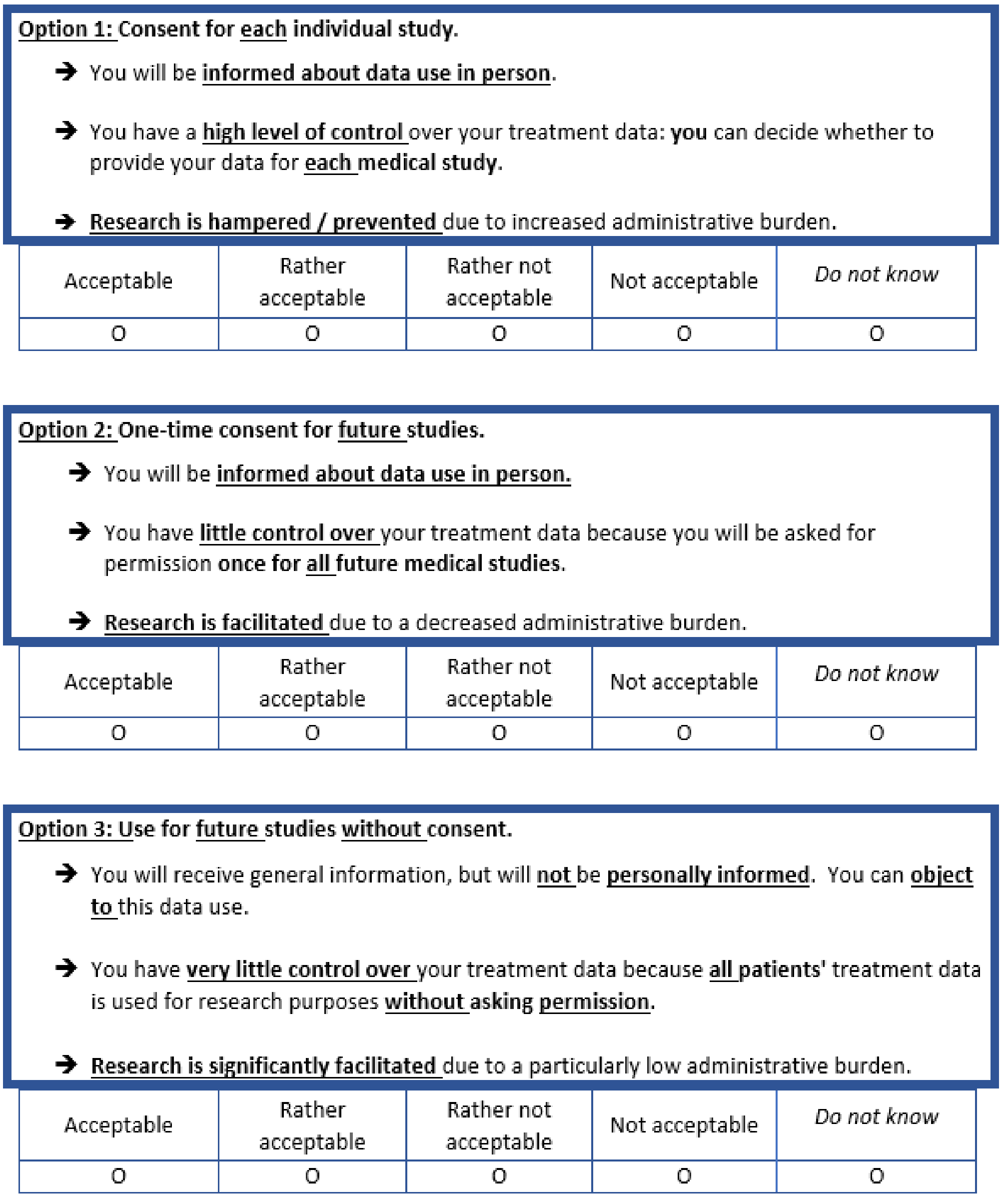
The questionnaire (Multimedia Appendix 1) was based on a review of the relevant scientific literature and a preparatory expert interview study among stakeholder groups engaged or affected by the planned secondary use of clinical data in Germany [17]. In total, 2 representatives for patient interests were included in the expert sample. The questionnaire was developed through several discussion and feedback rounds by the international and interdisciplinary project team, consisting of social scientists; ethicists; legal scholars; and clinicians with expertise in social, ethical, legal, or practical aspects of secondary use of clinical data. To ensure comprehensibility and technical functionality of the questionnaire, cognitive interviews (n=5) with patients with cancer and survivors of cancer who had provided consent were conducted in the pretest phase, resulting in minor adaptions.

To allow participants to develop an informed opinion, the survey included background information about risks and benefits associated with the secondary use of clinical data. The survey consisted of 33 items on the following topics: sociodemographic and disease-related information, expectations and risk perception toward secondary use, willingness to provide clinical data under certain requirements, and acceptability of consent models and procedures. Attitudinal questions were designed as 5-point Likert scale. The survey was approved by the data protection officer of the Heidelberg University Hospital.

Operationalization of Consent Scenarios

In total, 3 vignettes were developed to measure the acceptability of 3 consent scenarios: *specific consent*, *broad consent*, and *data use by default* (with the option to opt out at any time). Acceptance was measured using a 4-point Likert scale (Figure 1).

Figure 1. Display of the 3 consent scenarios in the questionnaire (English translation).



Option 1: Consent for each individual study.

➔ You will be **informed about data use in person**.

➔ You have a **high level of control** over your treatment data: **you** can decide whether to provide your data for **each medical study**.

➔ **Research is hampered / prevented** due to increased administrative burden.

Acceptable	Rather acceptable	Rather not acceptable	Not acceptable	*Do not know*
○	○	○	○	○

Option 2: One-time consent for future studies.

➔ You will be **informed about data use in person.**

➔ You have **little control over** your treatment data because you will be asked for permission **once for all future medical studies**.

➔ **Research is facilitated** due to a decreased administrative burden.

Acceptable	Rather acceptable	Rather not acceptable	Not acceptable	*Do not know*
○	○	○	○	○

Option 3: Use for future studies without consent.

➔ You will receive general information, but will **not** be **personally informed**. You can **object to** this data use.

➔ You have **very little control over** your treatment data because **all patients'** treatment data is used for research purposes **without asking permission**.

➔ **Research is significantly facilitated** due to a particularly low administrative burden.

Acceptable	Rather acceptable	Rather not acceptable	Not acceptable	*Do not know*
○	○	○	○	○

Previous studies have reported that participants made a trade-off between research utility and data control [13]. Hence, 3 consent scenarios were designed with information about research utility and control over data. In the process of operationalization, we further reduced the complexity of the theoretical concept to ensure good comprehensibility of the survey material:

1. By *specific consent*, we understand that consent is provided for each individual study (option 1 in Figure 1), as currently performed in clinical trials. Consistent with our preliminary studies [1,17], we inform about high degree of control over the secondary use of clinical data and low research benefit owing to the administrative burden on researchers.
2. In the case of the *broad consent* scenario, 1-time consent is provided for future medical studies with clinical data; moderate control and research utility are presumed (option 2 in Figure 1).
   This vignette refers to the implementation of a broad consent process for the German Medical Informatics Initiative; the development of a unified template for broad consent was accompanied by the German Working Group of Research Ethics Committees [18]. In practice, this model involves safeguards such as the review of each individual research project by a research ethics committee and data access committees, organizational measures to protect patient data, and comprehensive information for patients [18]. To ensure comprehensibility, the details of these safeguards are not provided to the participants of this study.
3. *Data use by default* is use of data for secondary research by default (comparable with Denmark or Estonia) without individual informed consent process, but with the possibility to opt out at any time. This scenario is associated with low degree of data control for patients and facilitation of research as no individual consent needs to be obtained (option 3 in Figure 1).
   Regarding law, the European Union (EU) data protection regulation provides some scope for this scenario of data use based on a legal basis other than informed consent if

> the potential research benefit clearly outweighs the right to informational self-determination (Art. 9, Paragraph 2, lit. j [19]). Compensating efforts such as ambitious security and privacy measures and extensive general public education about data use and data governance are likely to be ethically and legally necessary. To ensure comprehensibility, the details of these safeguards are not provided to the participants of this study.

Sampling and Recruitment

The Cancer Registry of the German federal state of Baden-Württemberg sent postal invitations to a random sample of patients with cancer and survivors of cancer, proportionally stratified by age and gender, requesting study participation (n=4219). The sample frame consisted of all registered patients in Baden-Württemberg, Germany, with a diagnosed tumor disease who were aged ≥18 years. Participants had the option of either completing an anonymous and self-administered web-based survey (the hyperlink was provided in the cover letter) or returning an envelope by mail, consenting that their address may be forwarded to the research group to subsequently receive a paper-and-pencil questionnaire. Survey instruments were adapted to the requirements of a mixed-mode survey [20].

Individuals who completed the survey were not compensated.

Data collection occurred from May 2021 to July 2021.

Analysis

Descriptive statistics were used to express the categorical variables as counts and percentages. Differences in proportions were assessed for statistical significance ($P<.05$) using chi-square tests. The 2-tailed Pearson correlation coefficients were computed. All analyses were performed using SPSS (version 28; IBM Corp).

Ethics Approval

The study obtained ethics approval from the University of Heidelberg's research ethics committee (reference number S-361/2018). Informed consent was obtained from the individuals who participated in the study pretest measurement and the written survey.

Results

Demographics of Participants

Of the 4155 patients with cancer approached by the Cancer Registry Baden-Württemberg, 838 (20.17%) participants completed the survey. Approximately half of the participants who answered the respective question were women (389/820, 47.4%; Table 1). Of 832 participants, 390 (46.9%) participants were aged between 60 and 74 years, and of 826 participants, 541 (65.5%) participants were retired. In total, 29.8% (247/830) of the participants had a university degree. The most common types of cancer were breast cancer (204/826, 24.7%), prostate cancer (187/826, 22.6%), and gastrointestinal cancer (79/826, 9.6%). The distribution of age, gender, and cancer entity mirrored that of the general distribution of patients with cancer in the Cancer Registry Baden-Württemberg, with minor deviation.

Table 1. Demographics of participants.

Characteristics	Values, n (%)
Gender (n=820)	
Women	389 (47.4)
Men	431 (52.6)
Age groups (years; n=832)	
18-59	186 (22.4)
60-74	390 (46.9)
≥75	256 (30.8)
Highest educational degree (n=830)	
Elementary school diploma	84 (10.1)
Secondary school diploma	398 (47.9)
Qualification for university entrance	97 (11.7)
University degree	247 (29.8)
No school diploma	4 (0.5)
Employment status (n=826)	
Employed or self-employed	219 (26.5)
Not employed owing to health reasons	45 (5.4)
Retired	541 (65.5)
Not employed owing to other reasons	21 (2.5)
Type of cancer (n=826)	
Breast	204 (24.7)
Prostate	187 (22.6)
Gastrointestinal	79 (9.6)
Skin cancer	63 (7.6)
Non-Hodgkin lymphoma	39 (4.7)
Lung	31 (3.8)
Leukemia	22 (2.7)
Kidney	22 (2.7)
Head and neck	22 (2.7)
Uterine or endometrial	21 (2.5)
Urinary bladder	18 (2.2)
Stomach	16 (1.9)
Pancreas	9 (1.1)
Other	93 (11.3)

General Willingness to Provide Clinical Data for Biomedical Research Purposes and Requirements for Data Provision

Most participants indicated that they are generally willing to make their clinical data available either without restrictions (527/838, 62.9%) or under certain conditions (283/838, 33.8%). Only 0.7% (6/838) of the participants generally refused to provide clinical data.

Then, the participants who indicated general willingness were asked about certain requirements under which they would provide their clinical data. When asked about the general requirements they deemed relevant, most participants stated the highest possible data security standards (482/838, 57.5%), followed by use of their data for as many research projects as possible (254/838, 30.3%), and being informed about the most important research results (208/838, 24.8%; Multimedia Appendix 2).

Most participants (591/832, 70.5%) stated that they would support research with their data in countries with high level of data protection comparable with German standards; 17.9% (149/832) of the participants stated that they would restrict data use to domestic research projects; and 8.8% (73/832) of the participants agreed to support international projects, independent of the level of data protection (Multimedia Appendix 3).

When asked how long their initial consent should be valid, 38.5% (320/832) of the participants set no time limit and approximately half of the participants demanded to renew consent either after 3 years (181/832, 21.8%), 10 years (227/832, 27.3%), or 30 years (10/832, 1.2%), respectively. In total, 10.2% (85/832) of the participants favored renewal of consent each time their data are used for specific research projects (Multimedia Appendix 4).

A large proportion of participants (532/832, 63.4%) said that they would grant access to researchers, independent of their affiliation; however, 22.7% (189/832) of them did not want to share their data with researchers at for-profit companies that conduct medical research (Multimedia Appendix 5). Only a small proportion opposed the secondary use of their clinical data by their physicians (56/832, 6.7%) or researchers at universities and university hospitals (48/832, 5.8%).

Acceptance of Consent Models

The questionnaire provided information about 3 consent models that correspond to specific consent, broad consent, and data use by default (with the option to opt out at any time), including the trade-offs of each model between control over clinical data and the facilitation of medical research (Table 2). For each consent model, the participants rated the level of acceptance on a 4-point Likert scale. Each of the 3 consent models showed a medium degree of acceptance with significant mean differences. Of the 838 participants, 491 (58.6%) accepted the broad consent model, 421 (50.2%) accepted data use by default (with the option to opt out at any time), and 323 (38.5%) accepted the specific consent model. Of the 323 participants accepting the specific consent model, 102 (31.6%) did not accept any other model (102/838, 12.2% of the total sample). Sociodemographic characteristics were not significant, except for older participants being more likely to accept data use by default (Pearson coefficient, 2-tailed: r=0.138; P<.001).

Table 2. Acceptance rates of 3 consent models: broad consent, data use by default, and specific consent (N=838)[a].

Model	Description	Accepted, n (%)	Not accepted, n (%)	Do not know or not answered, n (%)
Broad consent	One-time consent for future studies, informed in person, low level of control, and research is facilitated	491 (58.6)	230 (27.4)	117 (13.9)
Data use by default	Use for future studies without consent process, not personally informed, very low level of control, and research is significantly facilitated	421 (50.2)	347 (41.4)	70 (8.4)
Specific consent	Consent for each study, informed in person, high level of control, and research is hampered	323 (38.5)	372 (44.4)	143 (17.1)

[a] Acceptance was measured using a 4-point scale; results were collapsed into 2 groups (*not acceptable*: not acceptable and rather not acceptable; *acceptable*: acceptable and rather acceptable).

Preferences Regarding the Setting for Providing Consent

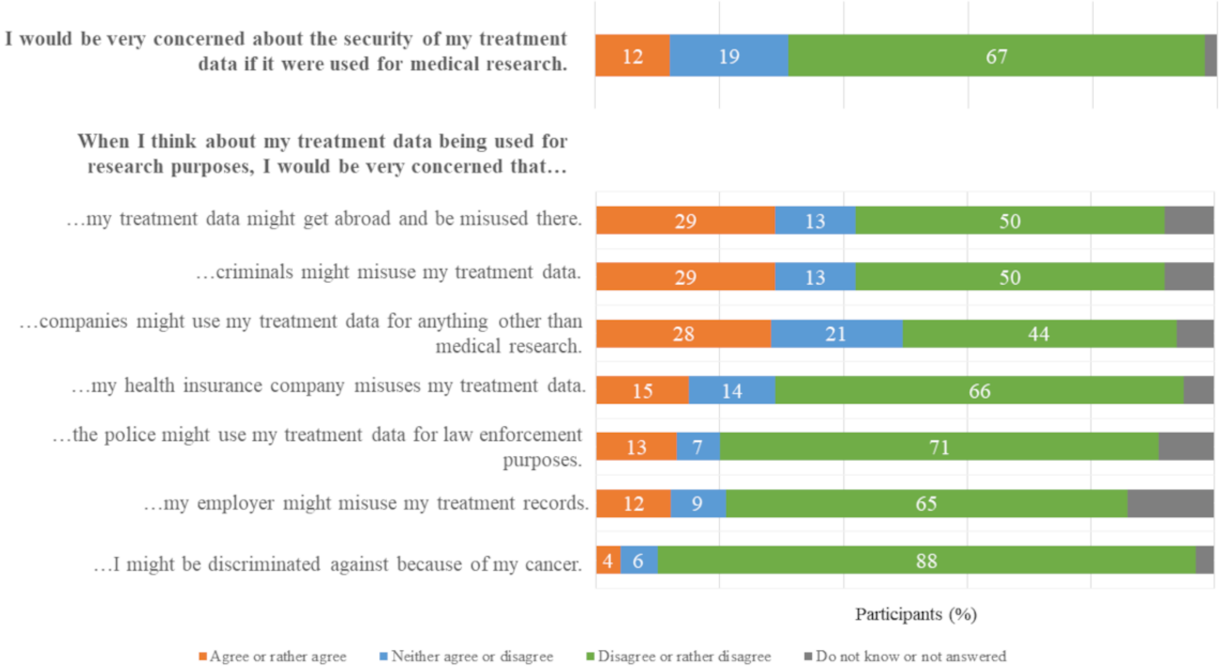
Participants were asked about the most appropriate setting for providing consent for the secondary use of their clinical data for research purposes. Most of them preferred to decide at their general practitioner's practice (528/838, 63%), and a small proportion of participants preferred to decide during the admission to a hospital (174/838, 20.8%; Multimedia Appendix 6).

When asked about preferred information formats, most participants selected a brief written summary of key points in easy-to-understand language to learn more about secondary use (616/838, 73.5%), followed by face-to-face consultation with physicians (347/838, 41.4%; Multimedia Appendix 7). Participants were asked about who should decide about data access and use by individual research projects: most participants (393/838, 46.9%) favored committees with experts in which the opinion of patients is represented, for example, by patient representatives, whereas a small proportion of participants preferred to leave the decision to an expert committee (without patient representation; 185/838, 22.1%) or to decide for themselves (200/838, 23.9%; Multimedia Appendix 8).

Concerns in the Event of Data Use

A small proportion of the participants (99/838, 11.8%) showed major general concerns regarding their clinical data being used for research purposes (Figure 2). Then, all participants were asked about more specific concerns: the largest proportion of participants were worried about the data being misused in countries other than Germany (246/838, 29.4%), data being misused by criminals (244/838, 29.1%), and data being used by companies for something other than medical research (235/838, 28%). Concerns about participants being discriminated against because of cancer were very low (32/838, 3.8%).

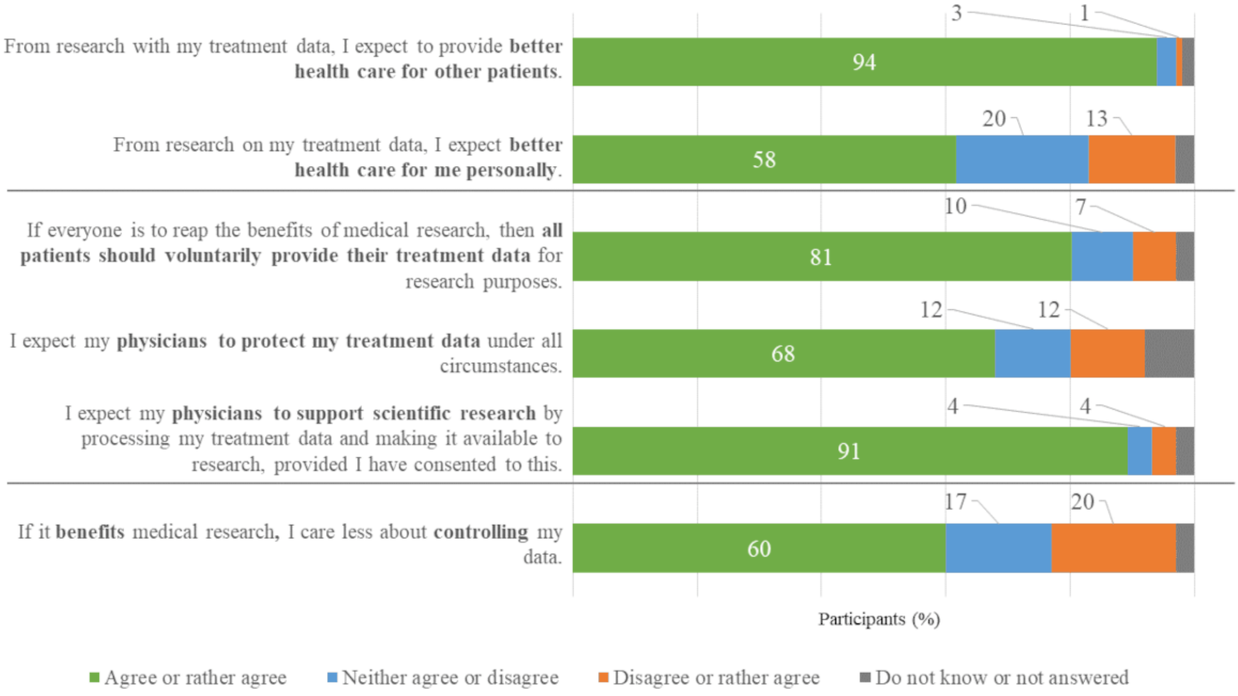
Figure 2. Concerns in the event of data use (N=838).



Expectations Toward Benefits, Other Patients, and Physicians

Approximately all participants (788/838, 94%) expected a benefit for other patients from making their clinical data available for research purposes (Figure 3). More than half of the participants (482/838, 57.5%) mistakenly expected a personal benefit, even though the explanatory text explicitly stated the opposite. Of the 838 participants, 676 (80.7%) participants supported the claim that all patients should voluntarily make their clinical data available for research purposes. In total, 68.3% (572/838) of the participants expected their physicians to protect the participants' clinical data in all circumstances, and approximately all participants (758/838, 90.5%) expected their physicians to support research, if consent was provided, by making their patients' clinical data available for research.

Figure 3. Expectations toward physicians and other patients (N=838).



Discussion

Principal Findings

Information about the requirements under which patients would make their treatment data available for research is important for any form of policy that regulates the secondary use of such data. This paper provides the results of a representative sample of German patients with cancer on general willingness and decisive requirements for sharing their data for research purposes and on the acceptance of consent models and expectations toward relevant stakeholders. The following are the main findings. First, we found an unprecedentedly high general willingness (810/838, 96.7%) to make clinical data available even after being informed about the potential risks of secondary use; however, relevant requirements included the following: ensuring a high level of data security, comparable data protection standards for data use abroad, and renewed consent at regular time intervals. Second, in contrast to previous studies, three-fourths of respondents (620/838, 73.9%) supported data use also by researchers in commercial companies. Third, the highest acceptance rate was found for a broad consent model (494/838, 58.9%), followed by data use by default (419/838, 50%); and specific consent for every study (327/838, 39%). Fourth, high expectations for physicians and fellow patients to support data sharing for research purposes were found.

To the best of our knowledge, this is the first representative study on attitudes toward the secondary use of clinical data and acceptance of consent models in combination with characteristics of data control and research utility.

High General Willingness to Provide Clinical Data

Overview

An important finding of our study was the high willingness of patients with cancer to make their clinical data available for research purposes (810/838, 96.7%), either without any restrictions (527/838, 62.9%) or under certain conditions (283/838, 33.8%). Only 0.7% (6/838) of the participants generally refused to provide clinical data. A population-representative study in the United States found low proportions of general willingness (76%) [21], similar to representative studies in Germany in the contexts of the COVID-19 pandemic (65%) [22] and medical data including genetic data (56%) [13]. These different results suggest that patients with cancer are more willing to provide clinical data for medical research because they may either have benefited or hope to benefit from research. As potential beneficiaries of past studies, they may also feel greater responsibility than citizens and other patient groups to support research to help future generations of patients [11,23-25]. Although patients with cancer are not representative of all patients, we assume that they can hint well at the attitude of other patient groups with severe or rare diseases, such as leukodystrophies [26]. A study conducted in the United States shows slightly lower willingness among patients with cancer and survivors of cancer (71%) [27] than among the general population (76%) [21], which may point toward country-specific factors in the context of health systems and trust in institutions.

General Requirements: Data Security, Maximizing Data Use, and Transparency

The most relevant general requirements for supporting the secondary use of clinical data for research were high data security (486/838, 57.9%), maximizing data use (251/838, 29.9%), and information about research results that made use of patients’ clinical data (210/838, 25.1%). These findings indicate the relevance of the ability of data governance to protect clinical data, maximize accessibility (and usability) of data for research, and report transparently on the results of data use. These findings are largely consistent with previous literature that describes secure data use, public benefits through effective use by researchers, and transparency as important requirements for data sharing [3,8,14]. It may well be that participants value the reporting of results as an act of recognition and reciprocity. Suggestions for future set up of governance for secondary use of data to respond to the abovementioned requirements include appropriate safeguards to protect patient data; high degree of transparency regarding data use and benefits to society; and technical, organizational, and legal data infrastructure that enables researchers to maximize research benefits. Involving patients to better understand their concrete needs in designing these requirements for secondary use seems advisable [28].

Data Transfer Only to Countries With Comparable Data Protection Standard

Most participants stated that they would restrict their data to research in countries with data protection standards comparable with those in Germany (737/838, 87.9%), and a small minority of the participants was willing to provide data to other countries (75/838, 8.9%). This resonates with another German study with outpatients who generally support data donation in favor of public research institutions in EU countries with similar data protection standards (92%); only a minority of the participants approved data access to countries outside the EU (24%), which is a large share compared with our findings [6]. The high relevance of this aspect is consistent with studies of Canadian citizens [8,9]. However, further studies are needed to explore the exact kinds of misuse that make people fearful about international data transfers. Our study suggests that comparable data protection standards are a decisive requirement for patients. A suggestion to address this need is that policy makers and data initiatives explain well to patients what the additional benefit of multinational research is, what the specific risks are (eg, foreign government access and less ability to enforce rights), and how risks to data protection in these countries are mitigated. They are well advised to give patients the choice of whether to consent to data transfer to countries with low data protection standards.

Most Participants Support Data Use by Corporate Researchers

Low willingness of citizens to share data with the private industry has been reported in several studies [3,4]. This finding poses challenges to the biomedical research landscape, as many studies are conducted by companies or in cooperation with companies. In contrast, our results show that approximately three-fourths of the participants (620/838, 73.9%) were willing to make their clinical data available to company researchers.

This is a much higher acceptance than in studies with German citizens [5] and outpatients [6], which reported that only a minority of those participants who agreed to data donation were willing to provide data to the industry (17% and 29%, respectively). A cross-country study found particularly low support for medical and genetic data sharing with for-profit researchers among German participants (22% compared with 32% on average across all countries) [29]. We hypothesize that willingness to share data with company researchers may change owing to experiences with a severe illness: patients with cancer may develop strong awareness of contributions by corporate researchers, possibly based on their experiences during their therapy. In addition, our questionnaire item included a brief explanation of the relevant contribution of industry to medical research and of industry as an important collaborator with public research institutions. We suppose the explanation increased the participants' understanding and willingness to provide clinical data to the industry, which is consistent with a study examining public attitudes toward commercial data access, in which provision of information and deliberative methods increased willingness to share data [15]. In addition, our findings indicate that low willingness to share data with corporate researchers can be addressed through collaboration with public research institutions in public-private partnerships.

Renewed Consent Within Certain Time Intervals

The participants' stance was divided on the duration of data use after initial consent is provided. Most participants (408/838, 48.7%) preferred to renew consent for broad research use after a period of 3 or 10 years. Only approximately one-third of the participants (243/838, 28.9%) preferred 1-time consent with unlimited duration of consent validity. In contrast, in a representative study of German citizens, more than half of the participants favored unlimited validity of consent (56%), and a minority favored consent validity of 5 years (17%) [5]. Our reported relatively high proportion of participants preferring renewed consent may have resulted owing to the following reasons. First, patients with cancer experience changing health conditions, leading to a subjective sensitivity to release clinical data without time limit. Second, our questionnaire explicitly mentioned risks of data release, possibly reducing the approval of unlimited data use. Third, the abovementioned study among German citizens asked for unlimited use for "data donation," which can be understood as irrevocable by definition. To address this potential need for patients to renew consent, further studies should investigate the preferences using neutral wording.

Broad Consent and Data Use by Default Was More Accepted Than Specific Consent—Research Benefits Partially Outweigh Loss of Control

Overview

Participants were presented with general information about 3 consent models (specific consent for every study; broad consent; and data use by default, with the option to opt out at any time). Specific consent is related to maximum informational control for patients, but less utility for research projects, whereas data use by default is associated with less informational control, but maximum utility for research projects. The broad consent model features moderate control and research utility (Table 2). The opportunity of being personally given information by health personnel is not available in the case of data use by default. Participants rated the level of acceptance for each consent model. The broad consent model received the highest acceptance rate (491/838, 58.6%), followed by data use by default (421/838, 50.2%) and the specific consent model with only a moderate acceptance rate (323/838, 38.5%). The relatively high acceptance rate for the broad consent model is consistent with the results of previous studies. Different study designs and minor deviations regarding the definition of consent procedures apply; therefore, comparisons should be considered cautiously. In total, 2 studies with a German patient sample and a large sample of Dutch patients found even higher acceptance rates in the context of health care–embedded biobanking and data donation (92%-93%) [5,10]. An earlier study of German patients (87%) [11] and a study of a smaller sample of US citizens (96%) [12] showed similar results. Our acceptance rates for each of the presented consent models were lower than those in other studies. This may be a consequence of the choice among 3 different models, rather than only 1, as presented in other studies. The low acceptance rates may also result from a trade-off decision between support for research and control over one's clinical data. Previous studies have described this trade-off between control and research benefits as a relevant influencing factor in decision-making [12-16]. Accordingly, in our study, most participants (520/838, 62.1%) agreed to give up control if it increased the benefits of research. This finding is significant because most participants (804/838, 95.9%) believe in the benefits of secondary use for other patients. Evidence from other studies [11,23-25] and our findings not only suggest that research benefits partially outweigh the loss of control but also that they are a critical motivational aspect of making data available for research.

As none of the models achieved wide-ranging acceptance in our study, it is worth discussing whether a meta-consent model that allows participants to choose their preferred consent variants [30] accounts best for individual ways of balancing control and research benefits regarding consent models.

Preferred Framework Conditions for Providing Consent and Data Release

When asked for consent, participants expected brief and understandable written information (616/838, 73.5%) about data use and preferred their primary care physician as a venue for informed consent (528/838, 63%) over providing consent upon hospital admission (174/838, 20.8%). This finding is underpinned by a qualitative study in which support by health care professionals was seen as an important facilitator [8].

Our findings indicate that, regarding place and time (ie, where and when patients are informed and asked for consent), consent in the clinical context is preferred over consent before becoming a patient. This is consistent with the finding that patients prefer providing consent at hospitals (64%-76%) over providing consent outside the clinic [6]. However, another study concluded that the decision about making data available for research should be separated from the clinical context and anchored in everyday life [31]. Owing to possible age and disease effects, further studies should investigate the differences between the general population's and patients' acceptance.

When asked who should decide on data release when individual research projects apply for using participants' clinical data after having personally released their treatment data for research purposes, approximately half of the participants (394/838, 47%) preferred a committee with experts and patient representatives over a committee with experts only or deciding for themselves. A suggestion to address this need is to involve patients in data access committees.

Low Concerns and High Expectations

Low Level of Concern in General and About Discrimination

In our study, the proportion of participants who were concerned about the use of clinical data (101/838, 12.1%) was considerably lower than the findings of 2 surveys conducted in Australia among citizens (24%-25%) [32] and patients (24%) [33]. A study conducted in the United States found that privacy concerns had the strongest influence on individuals' intentions to provide clinical data [34]. This discrepancy may be attributable to country-specific differences regarding trust in health care and government institutions [3] and the lack of experience with extensive data leaks or the misuse of clinical data in Germany. Participants' concerns about discrimination owing to their cancer were very low (34/838, 4.1%).

False Expectations of Personal Benefit

Most participants (486/838, 57.9%) incorrectly expected personal benefits from making their clinical data available for research purposes—even though the wording of the questionnaire had been adjusted during the pretest phase. Another study found that more than one-fourth of German patients hoped for personal benefit (28%) after being asked for consent for secondary use of clinical data and biomaterial collected during routine care [11]. Owing to the severity of the disease, patients with cancer may be particularly prone to this false expectation of personal benefit from research with their health data, which is comparable with therapeutic misconception [35] in clinical trials. The study showed that the proportion of those holding false expectation decreased considerably after the modification of consent information material (12%). To reduce the risk of false expectations, particularly in vulnerable groups such as patients with severe illnesses, careful education about the unlikelihood of direct benefits from making their clinical data available for research purposes is needed.

High Expectations of Other Patients and Physicians

Our results indicate a clear expectation toward fellow patients (696/838, 83.1%) to support medical research with clinical data, which is consistent with a study conducted in Germany among outpatients (80%-90%) [6]. Interestingly, more participants expected their physicians to share clinical data for research (754/838, 89.9%) than to protect their clinical data under all circumstances (570/838, 68%). This is the first study to investigate the expectations toward physicians.

Limitations

The recruited sample is largely representative of the population of patients with cancer in the federal state of Baden-Württemberg in age, gender, and cancer entity. However, we found that the educational level in our sample was higher than that of the corresponding age cohorts of the German population [36]. The educational level of the German population presumably applies to the group of patients with cancer and survivors of cancer. Owing to the topic of the survey, we suspect a self-selection bias correlating with high educational level. According to a study in the context of genetic research and biobanking, high educational level positively correlates with willingness to provide data [24]; consequently, our results may overestimate willingness to provide clinical data. In addition, false expectations of personal benefits from providing data for secondary use may have increased the participants' willingness to share clinical data.

A considerable proportion of participants who had previously agreed to hypothetically make their clinical data available *without restrictions*, favored restricted use of their clinical data when asked about specific requirements such as data user, duration, and data use in other countries (Multimedia Appendix 9). We assume that the participants have not yet formed a strong opinion about sharing their clinical data. Hence, the general willingness to provide clinical data seems to measure an overall attitude toward secondary use, rather than the actual willingness to provide clinical data without restrictions for research purposes.

Conclusions

Our study shows very high general willingness of patients with cancer to make their clinical data available for biomedical research purposes. However, the willingness to provide clinical data may be overstated owing to the above-average educational level of the respondents. For a considerable proportion of patients with cancer, willingness depends on certain requirements. In addition to the basic prerequisite of high level of data security and transparency in the use of the data, most patients shared the view that the data must not be used in countries with low data protection standards and that they should have the possibility to renew consent. In contrast to previous studies, the exclusion of use of data for private sector studies is not a requirement for most participants.

High willingness on the part of patients to accept loss of control over clinical data in favor of research benefits and request to maximize accessibility (and usability) of data for research were found. This is consistent with the acceptance of more research-friendly and low-control models, namely the broad consent model, followed by data use by default (with the option to opt out at any time). The striving for maximizing data use is also reflected by patients' expectations toward physicians and other patients to support secondary use.

Policy makers are well advised to account for patients' views when designing and implementing secondary use, with the aim to contribute to a socially legitimized culture of data sharing.

Acknowledgments

This paper was developed within the framework of the project, *Learning from Clinical Data*, funded by the Deutsche Forschungsgemeinschaft (German Research Foundation)—406103282. The project funding was awarded to ECW. The funders did not engage in the study design; collection, analysis, and interpretation of data; writing of the report; or decision to submit the paper for publication. No additional external funding was received for this study. The authors would like to thank the Cancer Registry of Baden-Württemberg for supporting data collection. The authors also thank their project partners, Kai Cornelius, Markus Spitz (University of Heidelberg, Germany), and Adrian Thorogood (University of Luxemburg), for conceptual counseling. The authors acknowledge Hellen Kachler and Niamh Sulzbach (both from the National Center for Tumor Diseases, Heidelberg University Hospital, Germany) for their assistance in researching the literature, data entry, and linguistic correction of the manuscript. The authors would also like to express their gratitude to Anne Müller, who assisted them in the recruitment of patients for the pretest phase, and to all the patients with cancer and survivors of cancer who participated in the study.

Data Availability

The data sets generated and analyzed in this study are available on the web [37].

Authors' Contributions

All the authors contributed to the conception and design of the study. AK and KM were involved in material preparation and data collection and analysis. ECW and KM supervised the study and supported data interpretation. AK wrote the first draft of the manuscript, and all the authors commented on the previous versions of the manuscript. All the authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[PDF File (Adobe PDF File), 167 KB - jmir_v24i8e37665_app1.pdf]

Multimedia Appendix 2
Participants' requirements that are to be met to provide clinical data.
[DOCX File , 14 KB - jmir_v24i8e37665_app2.docx]

Multimedia Appendix 3
The acceptance of use of clinical data for biomedical research purposes in other countries.
[DOCX File , 14 KB - jmir_v24i8e37665_app3.docx]

Multimedia Appendix 4
Preferred duration of use of clinical data for biomedical research purposes.
[DOCX File , 13 KB - jmir_v24i8e37665_app4.docx]

Multimedia Appendix 5
Participants' restrictions for the provision of their clinical data regarding different researcher groups.
[DOCX File , 13 KB - jmir_v24i8e37665_app5.docx]

Multimedia Appendix 6
Preferred context of providing informed consent.
[DOCX File , 13 KB - jmir_v24i8e37665_app6.docx]

Multimedia Appendix 7
Ways of obtaining information.
[DOCX File , 13 KB - jmir_v24i8e37665_app7.docx]

Multimedia Appendix 8
Decision on data release for individual research projects.
[DOCX File , 13 KB - jmir_v24i8e37665_app8.docx]

Multimedia Appendix 9

Conditional data release of participants initially stating to provide data without restriction.

[DOCX File , 13 KB - jmir_v24i8e37665_app9.docx]

References

1. Jungkunz M, Köngeter A, Mehlis K, Winkler EC, Schickhardt C. Secondary use of clinical data in data-gathering, non-interventional research or learning activities: definition, types, and a framework for risk assessment. J Med Internet Res 2021 Jun 08;23(6):e26631 [FREE Full text] [doi: 10.2196/26631] [Medline: 34100760]
2. Braunack-Mayer A, Fabrianesi B, Street J, O'Shaughnessy P, Carter SM, Engelen L, et al. Sharing government health data with the private sector: community attitudes survey. J Med Internet Res 2021 Oct 01;23(10):e24200 [FREE Full text] [doi: 10.2196/24200] [Medline: 34596573]
3. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2022 Jan;48(1):3-13 [FREE Full text] [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
4. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
5. Richter G, Borzikowsky C, Lesch W, Semler SC, Bunnik EM, Buyx A, et al. Secondary research use of personal medical data: attitudes from patient and population surveys in The Netherlands and Germany. Eur J Hum Genet 2021 Mar;29(3):495-502 [FREE Full text] [doi: 10.1038/s41431-020-00735-3] [Medline: 33005018]
6. Richter G, Borzikowsky C, Hoyer BF, Laudes M, Krawczak M. Secondary research use of personal medical data: patient attitudes towards data donation. BMC Med Ethics 2021 Dec 15;22(1):164 [FREE Full text] [doi: 10.1186/s12910-021-00728-x] [Medline: 34911502]
7. Shah N, Coathup V, Teare H, Forgie I, Giordano GN, Hansen TH, et al. Motivations for data sharing-views of research participants from four European countries: a DIRECT study. Eur J Hum Genet 2019 May;27(5):721-729 [FREE Full text] [doi: 10.1038/s41431-019-0344-2] [Medline: 30700834]
8. Cumyn A, Barton A, Dault R, Safa N, Cloutier A, Ethier J. Meta-consent for the secondary use of health data within a learning health system: a qualitative study of the public's perspective. BMC Med Ethics 2021 Jun 29;22(1):81 [FREE Full text] [doi: 10.1186/s12910-021-00647-x] [Medline: 34187453]
9. McCormick N, Hamilton CB, Koehn CL, English K, Stordy A, Li LC. Canadians' views on the use of routinely collected data in health research: a patient-oriented cross-sectional survey. CMAJ Open 2019;7(2):E203-E209 [FREE Full text] [doi: 10.9778/cmajo.20180105] [Medline: 30948649]
10. Richter G, Borzikowsky C, Lieb W, Schreiber S, Krawczak M, Buyx A. Patient views on research use of clinical data without consent: legal, but also acceptable? Eur J Hum Genet 2019 Jun;27(6):841-847 [FREE Full text] [doi: 10.1038/s41431-019-0340-6] [Medline: 30683927]
11. Richter G, Krawczak M, Lieb W, Wolff L, Schreiber S, Buyx A. Broad consent for health care-embedded biobanking: understanding and reasons to donate in a large patient sample. Genet Med 2018 Jan;20(1):76-82 [FREE Full text] [doi: 10.1038/gim.2017.82] [Medline: 28640237]
12. Hammack-Aviran CM, Brelsford KM, McKenna KC, Graham RD, Lampron ZM, Beskow LM. Research use of electronic health records: patients' views on alternative approaches to permission. AJOB Empir Bioeth 2020;11(3):172-186 [FREE Full text] [doi: 10.1080/23294515.2020.1755383] [Medline: 32338567]
13. Voigt TH, Holtz V, Niemiec E, Howard HC, Middleton A, Prainsack B. Willingness to donate genomic and other medical data: results from Germany. Eur J Hum Genet 2020 Aug;28(8):1000-1009 [FREE Full text] [doi: 10.1038/s41431-020-0611-2] [Medline: 32238912]
14. Tully MP, Bozentko K, Clement S, Hunn A, Hassan L, Norris R, et al. Investigating the extent to which patients should control access to patient records for research: a deliberative process using citizens' juries. J Med Internet Res 2018 Mar 28;20(3):e112 [FREE Full text] [doi: 10.2196/jmir.7763] [Medline: 29592847]
15. Ipsos M. The One-Way Mirror: public attitudes to commercial access to health data. Wellcome Trust. 2017. URL: https://wellcome.figshare.com/articles/journal_contribution/The_One-Way_Mirror_Public_attitudes_to_commercial_access_to_health_data/5616448/1 [accessed 2022-07-21]
16. Oliver JM, Slashinski MJ, Wang T, Kelly PA, Hilsenbeck SG, McGuire AL. Balancing the risks and benefits of genomic data sharing: genome research participants' perspectives. Public Health Genomics 2012;15(2):106-114 [FREE Full text] [doi: 10.1159/000334718] [Medline: 22213783]
17. Köngeter A, Jungkunz M, Winkler E, Schickhardt C, Mehlis K. Sekundärnutzung klinischer Daten aus der Patientenversorgung für Forschungszwecke – Eine qualitative Interviewstudie zu Nutzen- und Risikopotenzialen aus Sicht von Expertinnen und Experten für den deutschen Forschungskontext. In: Datenreiche Medizin und das Problem der Einwilligung. Berlin, Heidelberg: Springer; 2021.

18. Zenker S, Strech D, Ihrig K, Jahns R, Müller G, Schickhardt C, et al. Data protection-compliant broad consent for secondary use of health care data and human biosamples for (bio)medical research: towards a new German national standard. J Biomed Inform 2022 Jul;131:104096. [doi: 10.1016/j.jbi.2022.104096] [Medline: 35643273]
19. Regulation (EU) 2016/679 of the European parliament and of the council. Official Journal of the European Union. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN [accessed 2022-08-03]
20. Stadtmüller S, Beuthner C, Silber H. Mixed-Mode Surveys (Version 1.0). GESIS - Leibniz-Institut für Sozialwissenschaften. 2020. URL: https://www.ssoar.info/ssoar/bitstream/handle/document/71311/ssoar-2020-stadtmuller_et_al-Mixed-Mode_Surveys_Version_10.pdf?sequence=1&isAllowed=y&lnkname=ssoar-2020-stadtmuller_et_al-Mixed-Mode_Surveys_Version_10.pdf [accessed 2022-07-21]
21. Kim KK, Joseph JG, Ohno-Machado L. Comparison of consumers' views on electronic data sharing for healthcare and research. J Am Med Inform Assoc 2015 Jul;22(4):821-830 [FREE Full text] [doi: 10.1093/jamia/ocv014] [Medline: 25829461]
22. Pulscheck Deutschland zu COVID-19 und Datenspende. Data4Life. URL: https://www.data4life.care/de/corona/umfrage/ [accessed 2022-03-03]
23. Nobile H, Bergmann MM, Moldenhauer J, Borry P. Participants' accounts on their decision to join a cohort study with an attached biobank: a qualitative content analysis study within two German studies. J Empir Res Hum Res Ethics 2016 Jul;11(3):237-249. [doi: 10.1177/1556264616657463] [Medline: 27381010]
24. Kettis-Lindblad A, Ring L, Viberth E, Hansson MG. Genetic research and donation of tissue samples to biobanks. What do potential sample donors in the Swedish general public think? Eur J Public Health 2006 Aug;16(4):433-440. [doi: 10.1093/eurpub/cki198] [Medline: 16207726]
25. Aitken M, Porteous C, Creamer E, Cunningham-Burley S. Who benefits and how? Public expectations of public benefits from data-intensive health research. Big Data Soc 2018 Dec 06;5(2):205395171881672. [doi: 10.1177/2053951718816724]
26. Darquy S, Moutel G, Lapointe A, D'Audiffret D, Champagnat J, Guerroui S, et al. Patient/family views on data sharing in rare diseases: study in the European LeukoTreat project. Eur J Hum Genet 2016 Mar 17;24(3):338-343 [FREE Full text] [doi: 10.1038/ejhg.2015.115] [Medline: 26081642]
27. Franklin EF, Nichols HM, House L, Buzaglo J, Thiboldeaux K. Cancer patient perspectives on sharing of medical records and mobile device data for research purposes. J Patient Exp 2020 Dec;7(6):1115-1121 [FREE Full text] [doi: 10.1177/2374373520923837] [Medline: 33457553]
28. Muller SH, Kalkman S, van Thiel GJ, Mostert M, van Delden JJ. The social licence for data-intensive health research: towards co-creation, public value and trust. BMC Med Ethics 2021 Aug 10;22(1):110 [FREE Full text] [doi: 10.1186/s12910-021-00677-5] [Medline: 34376204]
29. Middleton A, Milne R, Almarri MA, Anwer S, Atutornu J, Baranova EE, et al. Global public perceptions of genomic data sharing: what shapes the willingness to donate DNA and health data? Am J Hum Genet 2020 Oct 01;107(4):743-752 [FREE Full text] [doi: 10.1016/j.ajhg.2020.08.023] [Medline: 32946764]
30. Ploug T, Holm S. Eliciting meta consent for future secondary research use of health data using a smartphone application - a proof of concept study in the Danish population. BMC Med Ethics 2017 Aug 15;18(1):51 [FREE Full text] [doi: 10.1186/s12910-017-0209-6] [Medline: 28810914]
31. Lesch W, Richter G, Semler S. Daten teilen für die Forschung: Einstellungen und Perspektiven zur Datenspende in Deutschland. In: Datenreiche Medizin und das Problem der Einwilligung. Berlin, Heidelberg: Springer; 2022.
32. King T, Brankovic L, Gillard P. Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. Int J Med Inform 2012 Apr;81(4):279-289. [doi: 10.1016/j.ijmedinf.2012.01.005] [Medline: 22306206]
33. Krahe M, Milligan E, Reilly S. Personal health information in research: perceived risk, trustworthiness and opinions from patients attending a tertiary healthcare facility. J Biomed Inform 2019 Jul;95:103222 [FREE Full text] [doi: 10.1016/j.jbi.2019.103222] [Medline: 31176040]
34. Abdelhamid M, Gaia J, Sanders GL. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
35. Appelbaum PS, Roth LH, Lidz CW, Benson P, Winslade W. False hopes and best data: consent to research and the therapeutic misconception. Hastings Cent Rep 1987 Apr;17(2):20-24. [Medline: 3294743]
36. Bildungsstand: Verteilung der Bevölkerung in Deutschland nach höchstem Schulabschluss. Statista. URL: https://de.statista.com/statistik/daten/studie/197269/umfrage/allgemeiner-bildungsstand-der-bevoelkerung-in-deutschland-nach-dem-alter/#professional [accessed 2022-03-03]
37. Köngeter A, Schickhardt C, Jungkunz M, Bergbold S, Mehlis K, Winkler EC. Patients’ willingness to provide their clinical data for research purposes and acceptance of different consent models. Findings from a representative survey of cancer patients in Germany. NCT - Section of Translational Medical Ethics. 2022. URL: https://heidata.uni-heidelberg.de/dataset.xhtml?persistentId=doi:10.11588/data/PJCNTN [accessed 2022-08-10]

Abbreviations

EU: European Union

Edited by R Kukafka; submitted 17.03.22; peer-reviewed by B Leach, J van Delden, K Fultz Hollis; comments to author 14.05.22; revised version received 27.06.22; accepted 11.07.22; published 25.08.22.

Please cite as:
Köngeter A, Schickhardt C, Jungkunz M, Bergbold S, Mehlis K, Winkler EC
Patients' Willingness to Provide Their Clinical Data for Research Purposes and Acceptance of Different Consent Models: Findings From a Representative Survey of Patients With Cancer
J Med Internet Res 2022;24(8):e37665
URL: https://www.jmir.org/2022/8/e37665
doi:10.2196/37665
PMID:36006690

Original Paper

Deciphering the Diversity of Mental Models in Neurodevelopmental Disorders: Knowledge Graph Representation of Public Data Using Natural Language Processing

Manpreet Kaur[1], BTech, MEng; Jeremy Costello[2], BEng; Elyse Willis[1]; Karen Kelm[1], BASoc; Marek Z Reformat[2,3], MSc (Hons), PhD; Francois V Bolduc[1,4,5,6], MD, PhD

[1]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada
[2]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada
[3]Information Technology Institute, University of Social Sciences, Łódź, Poland
[4]Department of Medical Genetics, University of Alberta, Edmonton, AB, Canada
[5]Women and Children Health Research Institute, University of Alberta, Edmonton, AB, Canada
[6]Neuroscience and Mental Health Research Institute, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Francois V Bolduc, MD, PhD
Department of Pediatrics
University of Alberta
3-020 Katz Building
11315 87 Avenue
Edmonton, AB, T6G 2E1
Canada
Phone: 1 780 492 9713
Email: fbolduc@ualberta.ca

Abstract

Background: Understanding how individuals think about a topic, known as the mental model, can significantly improve communication, especially in the medical domain where emotions and implications are high. Neurodevelopmental disorders (NDDs) represent a group of diagnoses, affecting up to 18% of the global population, involving differences in the development of cognitive or social functions. In this study, we focus on 2 NDDs, attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD), which involve multiple symptoms and interventions requiring interactions between 2 important stakeholders: parents and health professionals. There is a gap in our understanding of differences between mental models for each stakeholder, making communication between stakeholders more difficult than it could be.

Objective: We aim to build knowledge graphs (KGs) from web-based information relevant to each stakeholder as proxies of mental models. These KGs will accelerate the identification of shared and divergent concerns between stakeholders. The developed KGs can help improve knowledge mobilization, communication, and care for individuals with ADHD and ASD.

Methods: We created 2 data sets by collecting the posts from web-based forums and PubMed abstracts related to ADHD and ASD. We utilized the Unified Medical Language System (UMLS) to detect biomedical concepts and applied Positive Pointwise Mutual Information followed by truncated Singular Value Decomposition to obtain corpus-based concept embeddings for each data set. Each data set is represented as a KG using a property graph model. Semantic relatedness between concepts is calculated to rank the relation strength of concepts and stored in the KG as relation weights. UMLS disorder-relevant semantic types are used to provide additional categorical information about each concept's domain.

Results: The developed KGs contain concepts from both data sets, with node sizes representing the co-occurrence frequency of concepts and edge sizes representing relevance between concepts. ADHD- and ASD-related concepts from different semantic types shows diverse areas of concerns and complex needs of the conditions. KG identifies converging and diverging concepts between health professionals literature (PubMed) and parental concerns (web-based forums), which may correspond to the differences between mental models for each stakeholder.

Conclusions: We show for the first time that generating KGs from web-based data can capture the complex needs of families dealing with ADHD or ASD. Moreover, we showed points of convergence between families and health professionals' KGs.

Natural language processing–based KG provides access to a large sample size, which is often a limiting factor for traditional in-person mental model mapping. Our work offers a high throughput access to mental model maps, which could be used for further in-person validation, knowledge mobilization projects, and basis for communication about potential blind spots from stakeholders in interactions about NDDs. Future research will be needed to identify how concepts could interact together differently for each stakeholder.

(J Med Internet Res 2022;24(8):e39888) doi:10.2196/39888

KEYWORDS

concept map; neurodevelopmental disorder; knowledge graph; text analysis; semantic relatedness; PubMed; forums; mental model

Introduction

Neurodevelopmental disorders (NDDs) are common and represent a group of diagnoses consisting of differences in the development of cognitive, motor, or social skills [1]. Attention deficit hyperactivity disorder (ADHD) is the most common cause of NDDs and affects the ability of children and adults to focus their attention and regulate their motor activity. Another condition is autism spectrum disorder (ASD), which is associated with differences in social interaction, language, and behavior. The prevalence of NDDs is up to 18% worldwide when considering its most common conditions (ADHD) [2,3], while some conditions like ASD will have prevalence closer to 1% [4]. Individuals with ASD and ADHD frequently experience, in addition to their core disorders symptoms, a variety of associated issues, including sleep difficulties, challenging behaviors, and mental health concerns, with repercussions not only on health but also on education and social needs. This creates a level of complexity for parents and a need for large care teams and challenges in communication for health professionals involved with families with NDDs.

Research in medical complexity has shown how communication and care can be improved by establishing each stakeholder's representation of a condition known as the mental model. Mental models are dynamic and are constantly evolving sets of beliefs and knowledge, which dictate parents' and professionals' decisions and behaviors [5,6]. When collaborating with others, having contradictory mental models can lead to conflicting expectations and impede communication [7,8]. Representing mental models visually as a map increases communication and collaboration in education [9] and health care [10]. Mental models have been mapped using various in-person techniques such as cognitive task analysis and concept mapping [11]. Nonetheless, those require trained professionals and access to stakeholders, thereby limiting their scalability.

Knowledge graphs (KGs), as a graph-based information representation format, have been widely applied in artificial intelligence and structural representation of information [12]. KG represents knowledge in a structured way—concepts are nodes connected to each other with edges denoting relationships similar to concept maps. Web-based information has been increasingly used to identify themes of interest to patients. For instance, analysis of web-based information for individuals with cancer has been used to compare patients' and family members' concerns [13], patients' concern and research questionnaires [14], or clinical trial topics [15]. In addition, natural language processing (NLP) techniques have been used to identify and compare the language used to describe different mental health disorders [16]. The word co-occurrence analysis has been used extensively to extract the meanings from text, including health [17], cancer [18], and COVID-19 information, from Twitter [19]. Semantic relatedness tasks play an important role in many NLP applications such as word sense disambiguation [20,21], aspect-based sentiment analysis [22], query expansion [23], and information retrieval from electronic health records [24]. Our study is the first, to our knowledge, to leverage KG building tools to represent mental models from different stakeholders. Moreover, it remains unclear how medical professional literature addresses the topics of most interest to families. Therefore, we propose an approach for comparing ASD-related or ADHD-related concepts that are important and frequently occurring in family forums and in the PubMed literature related to these conditions. Our proposed approach is different from that in the prior mentioned work as it utilizes the vector space model (VSM)–based semantic relatedness technique to construct the KG representation of ASD-related and ADHD-related unified medical language system (UMLS) concepts.

The developed KGs depict concept maps of information from 2 sources: online communities and PubMed abstracts. They help identify concepts with similar and dissimilar relevancy or priority and their frequency of occurrence for the case of both stakeholders. Such a methodology is essential, as obtaining such information directly from stakeholders requires extensive effort involving recruitment and conducting interviews or distributing surveys (with often limited response rate).

Methods

Data Collection

PubMed Abstracts

Search queries "neurodevelopmental disorders [MeSH]," "autism," "autism spectrum disorder [MeSH]," "autistic disorder," "attention deficit and disruptive behavior disorders [MeSH]," "attention deficit disorder with hyperactivity [MeSH]," and "ADHD" were performed in PubMed using Entrez Programming Utilities application programming interface by the National Center for Biotechnology Information. A unique list of 226,660 article identifiers was created, and abstracts were retrieved by making another PubMed application programming interface call, which returned 118,153 nonempty abstracts.

Forum Posts

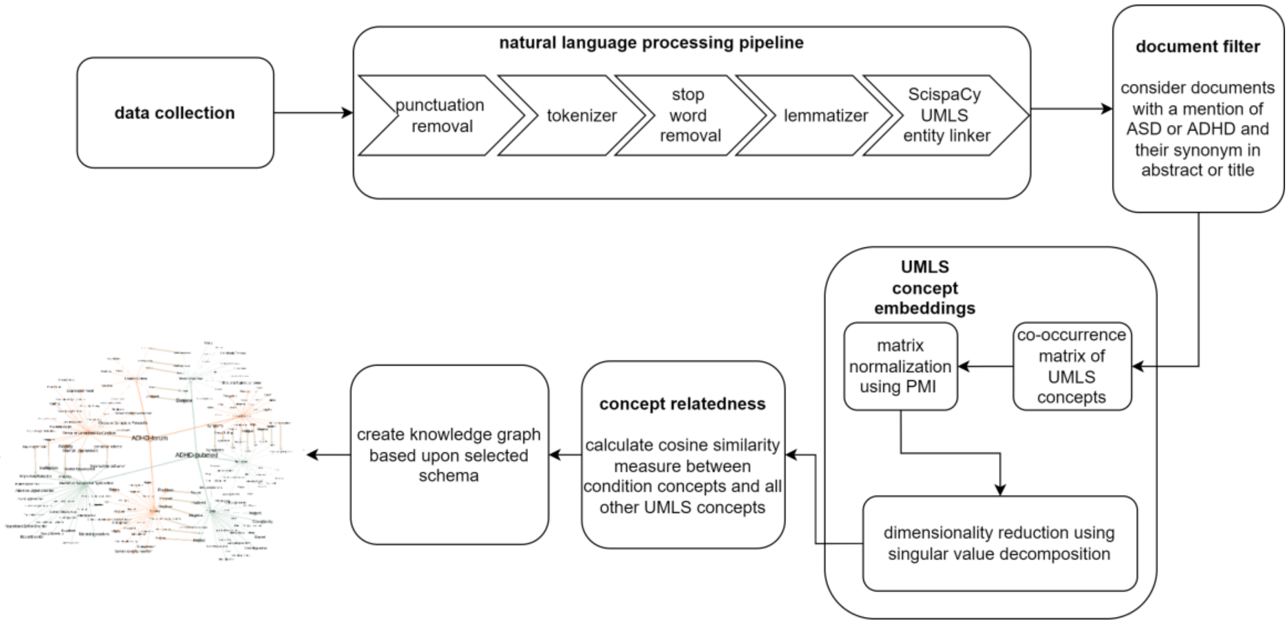
We manually googled publicly available web-based forums or communities and subreddits around the NDD topics to gather social media data and reviewed their privacy policies and terms of use. We selected 3 sources: healthboards.com [25], psychforums.com [26], and reddit [27], for which ethics approval for data collection and analysis was obtained from the University of Alberta. No HTML element containing identifiable personal information such as username was scraped, and only the one containing post was retrieved and stored locally. We did not contact any users for this research. As these online communities are not exclusively focused on NDD topics, we selected subforums about ASD and ADHD such as autism, Asperger syndrome, ADHD, and attention deficit disorder. We found various subreddits around ASD and ADHD, including askAutism, AutismBlogs, TeenAspies, ADD, adhd_anxiety, ADHD, and ParentingADHD. Python Scrapy framework [28] was used to scrape the posts from healthboards.com and psychforums.com, while Reddit application programming interface wrapper [29] was used to collect data from different subreddits. We only considered the main post of the thread and did not collect the list of replies to the thread. We assumed that the main thread consists of the concern posted by the forum user (which was our primary goal in building the KG). We did not include the replies as they would consist of the mention of the same concepts and would falsely boost the frequency of co-occurrence unless the text analysis pipeline has the ability to understand the complete sentence context such as relation extraction task. We did not filter the posts that were posted by parents only; therefore, these could be from any family member, caregiver, a friend of an individual with ASD or ADHD, or an individual with a condition itself.

NLP Pipeline

Data Preprocessing

All PubMed abstracts and forum posts (henceforth referred to as documents) were preprocessed using the Natural Language Toolkit Python library in order to remove punctuation, tokenize sentences into words, remove stop words, and lemmatize the words [30]. This process is illustrated in Figure 1. Stop words refer to the words that are not informative but occur a number of times such as is, am, are, and have. The default list of stop words provided by the Natural Language Toolkit was used as is.

Figure 1. Text processing and knowledge graph generation methodology. Data collection consists of forum scraping using Scrapy, Reddit wrapper application programming interface call, and PubMed application programming interface call. Collected documents are processed through the natural language processing pipeline. The knowledge graph is developed from top 25 concepts related to the condition concepts (autism spectrum disorder or attention deficit hyperactivity disorder) under selected semantic types. ADHD: attention deficit hyperactivity disorder; ASD: autism spectrum disorder; PMI: pointwise mutual information; UMLS: unified medical language system.



UMLS Entity Linker

The UMLS is a collection of over 100 controlled vocabularies, including but not limited to the International Classification of Diseases-10th classification, medical subject headings, and SNOMED Clinical Terms and contains over 4 million concepts [31]. UMLS facilitates biomedical entity detection by combining synonyms from different source vocabularies into canonical terms called concepts. UMLS also classifies all of its concepts into broader categories called semantic types; for instance, the ASD concept is classified as a mental or a behavioral dysfunction and the training programs concept as an educational activity. Semantic types provide the additional categorical information about the concept and are utilized in this project. An existing open-source Python library scispaCy is used to detect the UMLS concepts from documents [32]. The scispaCy UMLS entity linker provides the score for each detected concept, which ranges from 0 to 1. Low-scored terms would have higher chances of false positives, and we set the probability cutoff of 0.7 to reduce the chances of false positives. Therefore, only the concepts with scores greater than 0.7 along with their semantic type were considered in the final annotation.

In total, 124 UMLS semantic types from PubMed and 122 semantic types from the forum were detected, which could be applicable to all subfields of the medical domain. Peng et al [33] found that the precision of the UMLS entity linker tools could be low if the entities are not specific to ASD, and they used 13 semantic types in their analysis. Our preliminary analysis of all the semantic types was performed by comparing the frequencies of occurrence of each semantic type, which were calculated using all detected concepts from the documents corpus in each source. It showed that the most frequent semantic types such as qualitative concept, functional concept, and idea or concept in the database were not related to ASD and ADHD. Multimedia Appendix 1 shows the top frequent semantic types in each source. Considering the absence of established NDD-related semantic types, we prioritized a set of 26 types by reviewing associated concepts in collaboration with the NDD expert. The selected 26 semantic types are "activity," "age group," "behavior," "congenital abnormality," "diagnostic procedure," "daily or recreational activity," "disease or syndrome," "educational activity," "family group," "finding," "health care–related organization," "health care activity," "individual behavior," "injury or poisoning," "mental process," "mental or behavioral dysfunction," "occupational activity," "occupation or discipline," "organization," "patient or disabled group," "professional or occupational group," "professional society," "self-help or relief organization," "social behavior," "sign or symptom," and "therapeutic or preventive procedure." We excluded the frequent semantic types such as qualitative concept, functional concept, and idea or concept from the KG developed for this analysis. However, we are aiming to use those in future works.

If a concept is associated with more than one semantic type, then the scispaCy entity linker returns the list of all semantic types and does not consider the context of the sentences to select the semantic type being discussed. As it returns a list of all semantic types, we considered only the first returned semantic type. Concepts that occur in at least 10 documents in the corpus were considered for further analysis. Thus, we had 4494 unique concepts in PubMed documents and 3627 unique concepts in the forum.

Document Filter

All documents annotated with UMLS concepts passed through a filter that removed documents without mentioning ASD-related and ADHD-related concepts in the text. In UMLS, ASD, Asperger syndrome, and autistic disorder are different concepts; all the documents that mention any of these in either the abstract or the title are considered under ASD. Further, Asperger syndrome and autistic disorder concepts were replaced with ASD. As a result, we obtained a final data set of 55,461 PubMed abstracts in which 37,728 mentioned ASD, 20,805 mentioned ADHD, and 3072 mentioned both conditions. For the forum, the final data set contained 153,098 posts, in which 72,669 posts were about ASD, 90,372 were about ADHD, and 9943 had statements related to both conditions. Table 1 lists the number of posts collected from 3 web-based forums.

Table 1. Number of documents collected from different data sources.

Source	Autism spectrum disorder documents	Attention deficit hyperactivity disorder documents	Both autism spectrum disorder and attention deficit hyperactivity disorder documents
Reddit	66,552	87,022	9302
Psych forums	5029	1966	395
Health boards	1088	1384	246
Total documents from the 3 forums	72,669	90,372	9943
PubMed	37,728	20,805	3072

UMLS Concept Embeddings

Corpus-based numerical representation of concepts in the VSM represents the meaning of a concept based upon its context. It assumes that concepts that occur together in an environment (either document level, sentence level, or a neighborhood window of a particular size) would be related or similar to each other. The size of the context frames affects the representation of the concepts in the VSM, and many of the word embedding models such as the Skip-gram model and Continuous-bag-of-words model use window-context–based approaches called a local context. Document-level co-occurrence, referred to as a global context, provides more topical information around the concept, as many topic modeling approaches use the global context to detect the latent topics from a document [34]. As we want to detect topically most related concepts to ASD and ADHD, a global context-based co-occurrence matrix of size n × n is created where n refers to the total number of unique UMLS concepts in a source. The co-occurrence matrix is computed separately for PubMed and forum, as contextual information around a concept could be different depending upon the text corpus, which will eventually affect the relatedness scores.

Positive Pointwise Mutual information

Positive pointwise mutual information (PPMI) followed by truncated singular value decomposition (SVD) is used to embed the concepts, which provide comparative performance to neural network–based embedding models such as Word2Vec [35]. SVD PPMI usually produces consistent/stable results, where stability refers to the change in a word's neighborhood in the VSM, whereas neural network–based approaches (Word2vec, Glove) could lead to different results in different runs, as the weight of the hidden layers representing the word embeddings differs in multiple runs. SVD-based embeddings are not affected by this problem [36,37]. Pointwise mutual information (PMI) is a probabilistic approach to quantify the likelihood of co-occurrence and tells whether the co-occurrence is informative or by chance. It is defined as follows:

$$\text{PMI}(c_i, c_j) = \log [p(c_i, c_j) / (p(c_i) \times p(c_j))] \quad \textbf{(1)}$$

where c_i = i^{th} concept or the row

c_j = j^{th} context concept or the column

$p(c_i)$ = marginal probability of c_i

$p(c_j)$ = marginal probability of c_j

$p(c_i, c_j)$ = marginal probability of c_i and c_j

PMI varies from –1 to 1. If PMI is 0, co-occurrence of 2 concepts does not provide any information and is just by chance. When the joint probability is much higher than marginal probabilities, the co-occurrence is not by chance. If PMI is less than 0, then the independent occurrences of the concepts c_i and c_j are more informative as compared to co-occurrence. PPMI sets the PMI to 0 if it is less than 0.

$$\text{PPMI}(c_i, c_j) = \max(\text{PMI}(c_i, c_j), 0) \quad \textbf{(2)}$$

PPMI provides a square matrix M of size n × n. For the PubMed, n=4494 and for the forum, n=3627, which leads to high dimensionality of the VSM.

Truncated SVD

SVD is a dimensionality reduction technique used to obtain a low-rank approximation of a dense matrix M. SVD factorizes the matrix M as a product of 3 matrices:

$$M = USV^T \textbf{(3)}$$

where U and V are orthogonal matrices of size $n \times n$ and S is a $n \times n$ diagonal matrix with diagonal values sorted from high to low. The rank k (k<n) approximation of matrix M can be obtained from equation (3) as follows:

$$M_k = U_k S_k V_k^T \textbf{(4)}$$

Where U_k is a $n \times k$ matrix, S_k is a $k \times k$ diagonal matrix and V_k^T is a $k \times n$ matrix. U_kS_k is the matrix of size $n \times k$, which represents the n concepts in k dimensions. We set k=300 and used Python scikit-learn library to implement truncated SVD and obtain the 300D concept embedding [38]. Different low embedding sizes (usually 300-500) are shown to be used without specific mention of its effect on the final results and 300 dimensions of one of the commonly used sizes [39-41]. PPMI followed by SVD, once applied on forum and PubMed corpus separately, provides 2 VSMs, which represent the concepts depending upon their contextual information in each source.

Concept Relatedness

Semantic relatedness approaches detect the most related concepts for a given concept based upon the context in which it is used. Semantic similarity and relatedness tasks appear the same, but similarity refers to the concepts that are synonymous and can be used interchangeably, and relatedness refers to concepts that are related because of their usage in the same context. For example, ASD and aggressive behavior are related but not similar. The concept relatedness between 2 concepts c_i and c_j is measured using cosine similarity as the normalized dot product of the context vectors C_i and C_j:

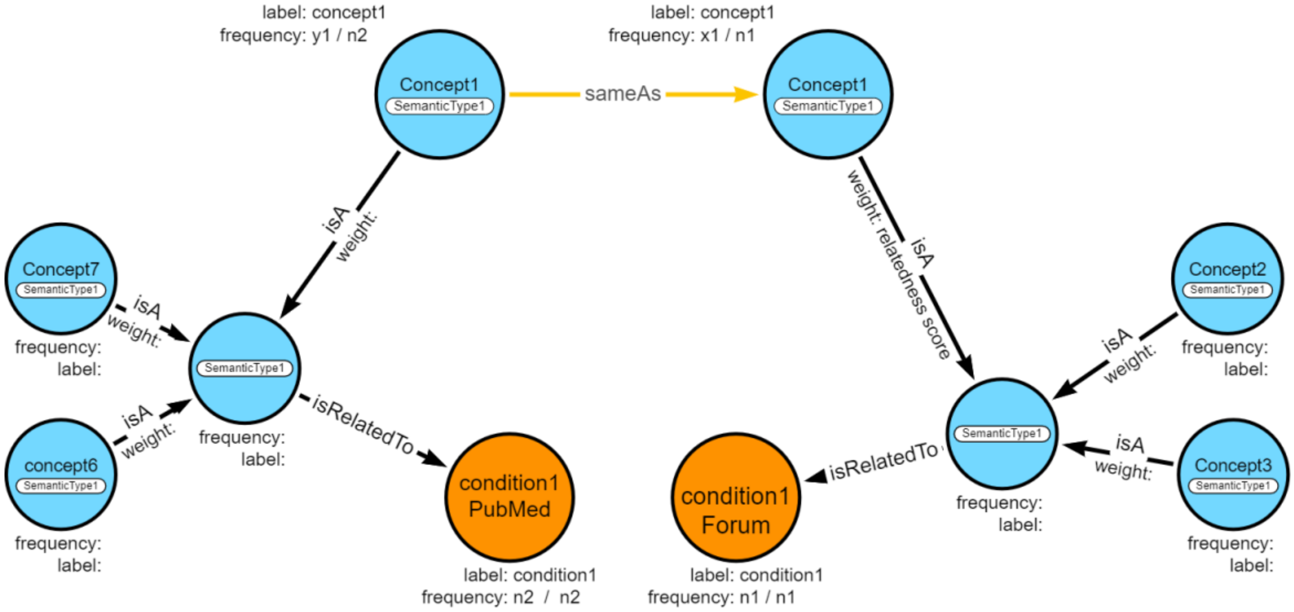
$$\text{relatedness}_{ij} = \text{cosineSim}(C_i, C_j) = C_i \cdot C_j \, / \, \| C_i \| \, \| C_j \|$$

relatedness_{ij} varies from (–1,1), where a value close to 1 means c_i and c_j are closely related to each other and both vectors have the same orientation in the VSM; a value close to 0 means c_i and c_j are dissimilar and both vectors are orthogonal in the VSM; and relatedness_{ij} of –1 indicates that c_i and c_j are in the opposite direction in multidimensional space.

KG Representation

The property graph schema, Figure 2, represents concepts associated with different UMLS semantic terms. There are nodes representing the condition (ASD or ADHD), related UMLS semantic types, and related concepts. Based upon relatedness scores between the condition and the concepts, the top 25 related concepts associated with each UMLS semantic type are used for creating the graph. An edge “isRelatedTo” links a semantic type node to a condition node, and each related concept is connected to its semantic type using the “isA” relationship. A set of property value pairs are stored on nodes as well as edges. All nodes have a label, which refers to the concept name, and the frequency, which is the proportion of documents in which a given concept co-occurred with the condition (ASD or ADHD), in each source data set. The frequency of a semantic type node refers to the average frequency of its top 25 concepts. The weight of the “isA” relationship indicates the relatedness score between the concept and the condition in a source data set, and no weight is assigned to “sameAs” and “isRelatedTo” relations. The Neo4j graph database is used to store the constructed KG [42].

Figure 2. Knowledge graph schema. The co-occurrence frequency of the concept (blue circle) and the condition (orange circle) is stored as the frequency of concept. The relatedness score of the concept to the condition is stored as the weight of isA relationship between the concept and the semantic type (also blue circle). The direction of the sameAs relationship can be either way.



Results

Diverse Areas of Concerns Around ADHD and ASD

The developed KG representation of PubMed and forums depict the mental models of both the stakeholders. We found a number of UMLS concepts associated with different semantic types in ADHD-related and ASD-related PubMed and forum data sets. All the detected concepts along with their semantic relatedness score are listed in Multimedia Appendix 2. In order to analyze the different areas of concerns, we assessed health care (PubMed abstracts) and family (forum posts) concepts associated with ADHD by visualizing the KGs from PubMed abstracts and family forums by using the Gephi network visualization tool [43]. In the KG visualizations, the thickness and color darkness of the relationship is proportional to cosine-based relatedness score of the concept to the condition (ADHD or ASD), and the size of the node/label is proportional to the co-occurrence frequency. We detected a few insignificant concepts in some of the semantic type groups. These concepts were then checked against the original text in the PubMed and forum documents, which showed that these concepts were false positives and therefore were removed from all of the analyses. Multimedia Appendix 3 shows the removed concepts along with the frequency of words linked to these concepts. Table 2 summarizes some of the most relevant terms for PubMed and forum documents on ADHD under different UMLS semantic types, which shows the different areas of concern for ADHD.

ADHD KGs generated from PubMed abstracts (see Multimedia Appendix 4) and forums (see Multimedia Appendix 5) show other areas of concerns such as “diagnostic procedure,” “individual behavior,” “health care activity,” and “professional or occupational group.” Similar to ADHD, ASD was found to be linked to diverse concepts in different domains represented by UMLS semantic types as shown in Table 3.

The KG representation of ASD PubMed abstracts (see Multimedia Appendix 6) and forums (see Multimedia Appendix 7) shows concepts under other semantic types, indicating other areas of concerns around ASD.

Table 2. Attention deficit hyperactivity disorder–related concepts in PubMed and forums for specific unified medical language system semantic types.

Unified medical language system semantic type	PubMed	Forum
Mental or behavioral dysfunction	• Inattention • Impulsive behavior • Hyperactive behavior • Attention deficit disorder • Substance abuse problem • Conduct disorder	• Executive dysfunction • Psychiatric problem • Anxiety • Hyperactive behavior • Inattention • Mental depression
Age group	• Adolescent • Adult • Young adult	• Adult • Adolescent • Child
Daily or recreational activity	• Sports • Youth sports • Recreational activity	• Reading activity • Speaking activity • Exercise
Educational activity	• Psychoeducation • Training programs • Socialization	• Homework • Home schooling • Training programs
Social behavior	• Parenting behavior • Social skills • Parent-child relationship	• Lifestyle • Conversation • Social behavior

Table 3. Autism spectrum disorder–related concepts in PubMed and forums under specific unified medical language system semantic types.

Unified medical language system semantic type	PubMed	Forum
Mental or behavioral dysfunction	• Developmental disabilities • Social communication disorder • Schizophrenia • Mental retardation	• Bullying • Aphasia • Social anxiety • Stereotypic movement disorder
Age group	• Child • Adult • Infant	• Child • Adult • Adolescent
Social behavior	• Communication • Social skills • Social cognition	• Social skills • Social situation • Eye contact
Mental process	• Perception • Cognition	• Stereotyping • Intelligence
Daily or recreational activity	• Physical activity • Youth sports • Speaking and reading activity	• Sports • Game • Speaking and reading activity
Educational activity	• Socialization • Training programs • Computer-assisted instruction • Special education • Parent training	• Socialization • Training programs • Special education • Toilet training • Home schooling

Comparing PubMed and Forum KG

KG helps identify concepts of similar and different relevance/priority between families and health professionals. Knowing that shared understanding (shared mental model) has been shown as a key factor in effective collaboration and quality communication in health care [44], we aimed at identifying potential concepts of similar and different relevance between forums and medical literature. For comparing concepts, we considered the top 25 concepts under selected UMLS semantic types, which were the most related to each condition (ASD and ADHD) based upon the relatedness scores, and visualized them using Gephi. As shown in Figure 3, KGs—one for PubMed and one for forum—are connected via the concepts that are of concern for both health professionals and online communities using "*sameAs*" relationship (orange arrow). The direction of this relationship can be either way. For the *“isA”* relationship (purple arrow), its thickness refers to the relatedness score of the concept to the condition (ADHD), which indicates the level of relevance or priority. Different node sizes of concepts connected with *“sameAs”* relationships show differences between the frequency of the concept in respective sources, such as mental depression and anxiety being more commonly discussed in ADHD forums as compared to ADHD PubMed abstracts, while hyperactive behavior, inattention, and impulsive behavior are more discussed in PubMed comparatively.

To summarize the concepts of similar and dissimilar relevance/priority, we compared the relatedness score of all the concepts in forum (FR) and PubMed (PR) and computed the score difference (score difference = FR – PR). The concept is of similar priority if its relatedness score is similar to both stakeholders and score difference of the concept is within $\mu \pm 2\sigma$, where μ is mean and σ is standard deviation of score difference. If score difference $> \mu + 2\sigma$, then the concept is more relevant for families (forum) and considered as a priority for them because of the substantial score difference. If score difference $< \mu - 2\sigma$, then the concept is considered as more relevant or as a priority for health professionals (PubMed). Interestingly, as shown in Table 4, we found several concepts of similar and dissimilar relevance to ADHD between PubMed and forum (see Multimedia Appendix 8 for KG visualization). The detailed relevance scores of all these concepts can be found in Tables S1-S3 in Multimedia Appendix 9.

Similarly, comparing the ASD-related concepts in both sources using relatedness score difference and KG representation provided various concepts of similar and dissimilar relevance, as shown in Table 5 (Multimedia Appendix 10 for KG visualization). Detailed relevance scores for all these concepts are listed in Tables S4-S6 in Multimedia Appendix 9.

Figure 3. Knowledge graph representation of forums and PubMed around attention deficit hyperactivity disorder for mental or behavioral dysfunction semantic type (green arrow). Concept label font size is proportional to the frequency of the concept in the source. The “sameAs” relationship (orange arrow) connects the overlapping concepts. The thickness of the edge “isA” (purple arrow) refers to the relatedness score of the concept to the condition (attention deficit hyperactivity disorder). ADHD: attention deficit hyperactivity disorder.

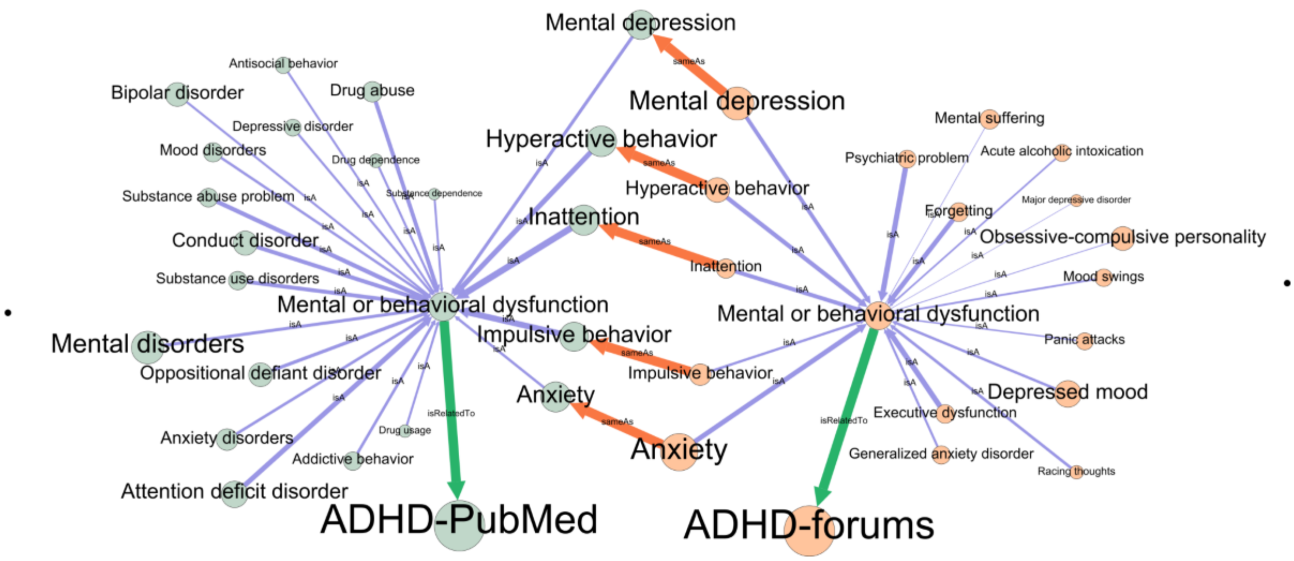


Table 4. Converging and diverging priority level for concepts in attention deficit hyperactivity disorders in PubMed and forum data.

	List of concepts
Concepts with similar relevance for both attention deficit hyperactivity disorder sources	• Impulsive behavior • Inattention • Anxiety • Mental depression • Hyperactive behavior • Sleeplessness • Emotional regulation • Attention • Training programs • Socialization
Concepts with high relevance to attention deficit hyperactivity disorder forums	• Executive dysfunction • Forgetting • Racing thoughts • Psychiatric problem • Nervousness • Exhaustion • Oversleep • Sluggishness • Study habits • Procrastination
Concepts with high relevance to attention deficit hyperactivity disorder in PubMed	• Substance abuse problem • Substance dependence • Conduct disorder • Antisocial behavior • Addictive behavior • Sleep phase delay • Amotivation • Anxiety symptoms • Oppositional behavior • Regulation of behavior

Table 5. Converging and diverging priority levels for concepts in autism spectrum disorder in PubMed and forum data sets.

	List of concepts
Concepts with similar relevance for both autism spectrum disorder sources	• Social communication disorder • Developmental disabilities • Aphasia • Autistic behavior • Intellectual disability • Mental disorders • Pervasive development disorder • Cerebral palsy • Seizures • Repetitive behavior • Social interaction • Nonverbal • Communication • Social behavior • Eye contact • Social skills • Aggressive behavior • Self-injurious behavior • Stereotyped behavior • Behavioral tic
Concepts with high relevance to autism spectrum disorder forums	• Bullying • Obsessions • Social phobia • Social anxiety • Temper tantrum • Mutism • Nervousness • Social problems • Introvert • Social life • Crowding
Concepts with high relevance to autism spectrum disorder in PubMed	• Immune dysregulation • Social cognition • Behavior adaptive

Discussion

Principal Findings

Understanding the needs and concerns of patients and their families is recently being recognized as a key factor for better communication between health professionals and families. This has led to emerging research into the role of mental models in medical practice [45-48] and their mapping [49]. Current approaches include interviews with patients, families, or experts and the identification of main concepts. Crandall et al [6] identified cognitive task analysis as one approach to building mental models. These rich interviews take place over a period of 60-90 minutes with approximately 10 participants. Although the information is rich and in depth, the process is both time-consuming and limited in participant numbers and diversity potentially.

From a theoretical perspective, our work shows how KG building techniques and NLP could help create mental models by using large-scale data sets and avoid bottlenecks such as limited access to experts and privacy/availability for families. Although the NLP methods used are well-established, the use of NLP to generate KGs to derive mental models and to compare them between families and health care professionals' perspective is completely novel to our knowledge. We show that web-based data from forums capture the diversity of concerns of parents of individuals with 2 important NDDs: ASD and ADHD. Publicly available web-based data could reflect the data obtained from more traditional approaches such as consultations or surveys as published in the literature. We show how using web-based data allows us to identify information about not only diagnostic criteria, medication, symptoms, or comorbidities of a condition but also other areas of concerns such as educational activities, recreational activities, and social issues around a condition, which were usually thought to be accessible mostly by interviews. We also show that the topics are not only related to controversies or unproven therapies, which has often been the rationale for not using web-based information in the medical domain. Similarly, interviews with medical experts are often a bottleneck in understanding concerns in the medical domain.

We also illustrate how web-based data can be used to identify points of convergence in priorities between the different stakeholders involved in complex medical conditions such as ADHD and ASD. Identification of converging points, that is, concepts of similar interest to health professionals and families could help clinicians and extension policy makers to identify "conversation starters" or shared interest. Identifying the diverging concepts or even blind spots for each stakeholder plays an important role for both clinicians and families. For instance, concepts that are highly relevant to families could be used by clinicians to frame continual medication education or training enhancement. For families, they could be the focus of knowledge mobilization, public education campaigns, or further studies aimed at enhancing literacy about their disorder and related conditions.

From a practical point of view, we present a framework that allows us to identify and rank relevant concepts for different sources by using corpus-based embeddings and semantic relatedness approaches as compared to simple co-occurrence frequency to rank related concepts. Developing a KG of the related concepts to represent the mental model visualizations could further assist in comparing converging and diverging concepts between both sources. To our knowledge, as there is no gold standard data set to evaluate the relatedness of concepts in NDDs, our framework proposes to use graph analysis tools such as Gephi to analyze and explore the KG visualizations manually, which could help validate the results by experts. Involving experts (expert in the middle) to review results of NLP approaches facilitates detection of incorrect concepts, which are the result of wrong mapping of abbreviations to concepts. Together, our research provides a proof-of-principle that will generate awareness about KGs as mental model maps and be of use to multidisciplinary researchers in a wide range of medical domains.

Comparison of KG-Based and Traditional Sources of Information

We compared our findings with previous literature or reports, which are the result of studies using traditional approaches such as interviews or surveys and involving participants (parents or health professionals) from the ASD and ADHD community. For ADHD, for instance, we found that priorities for individuals using the forum (parents, friends, caregivers) were related to prescription of medication and physician types. This reflected what has been discussed in the literature where participating parents were concerned about medication and nonpharmacological interventions (preferred behavior interventions) [50,51]. Another aspect of the topic of health professionals is around the source of information, which was noted previously as a major source of knowledge along with the internet [52]. Focus groups–based study, with caregivers included, showed that the major concern for the parents is about their child becoming a successful adult and improving school behavior [53,54] as well as improving their social situation and emotional state [55], which were identified as a priority before. We found the "behavioral habits" concept with relevancy score of 0.51 as the second most related to ADHD forums in UMLS semantic type "individual behavior." However, our current approach is based upon UMLS concept recognition and lacks the ability to understand the location as well as age context from the sentences that whether the "behavioral habits" is being discussed for school or home and child or teen. The NLP forum analysis also did not pick on an important trend for parents (and health professionals) to use multimodal interventions [56]. Similarly, our analysis of PubMed papers on ADHD identified topics previously identified by health experts as priorities. We found that the highest ranking topics were discussion of core symptoms of ADHD as well as comorbidities, conduct disorders, and substance use. This mirrors the health experts' consensus reports highlighting the importance of treatment efficacy for symptoms and raising the point of emotional aspects, academic performance, and work performance [57] as well as comorbidities such as mental illness and substance abuse [58-60]. Overall, we found that the perspectives in family ADHD forums and PubMed papers ranked at similar priority to the core symptoms of ADHD, comorbid conditions such as anxiety and depression, and the educational concerns of training programs and socialization.

With regard to ASD, our other NDD use case in this study, we found that the most overlapping topics had a similar priority level for the different stakeholders reflected by PubMed abstracts and ASD forums. These topics included classification of the condition, symptoms and behaviors that accompany ASD, and topics related to social interaction. Indeed, we found that priorities for people using ASD forums included concerns about social interaction such as social skills, communication, and friendship, as well as daily activities like speaking. This is similar to the findings of a survey distributed by Lai and Weiss [61] investigating service needs for ASD, which found that caregivers prioritized social skills and life-skills programs. Another study also found that the parents' main concern was social interaction [62], but that study found that the next most prevalent concerns were problem behavior and academics, which we did not see in our analysis of forums. A Serbian study similarly supported communication, social interaction, and daily activities as being caregiver priorities [63]. In addition, our analysis of PubMed abstracts revealed frequent discussion of classification of ASD and its relation to fetal alcohol spectrum and NDDs, concerns about social interaction and communication, and a focus on children with ASD. These priorities are supported by physicians' approach to ASD, which takes advantage of a diverse team of professionals to focus on improving social interaction and communication [64,65]. This is not to say that parents and research priorities are always aligned as shown in a recent survey in the United Kingdom, illustrating how research tends to be focused on biomedical aspects rather than services and supports [66].

We show that the KG derived from PubMed papers recapitulated the findings of position papers on the topic of ADHD and ASD as mentioned above. However, some of the differences in our findings and the participant-based study results could result from the differences in sample sizes or selection bias (age of caregivers and thus, children could be younger than school or adulthood ages). The collected web-based forum data are considerably larger than the number of participants in interview-based studies and therefore could include points of views not identified before. Alternatively, we could speculate that families may be more inclined to share personal concerns online than in an interview, although we did not find published studies looking into this topic. Further, we have included all the PubMed papers and web-based forums regardless of their publication or posting time (PubMed may include older concepts, which are no longer contemporary concerns), as opposed to the abovementioned expert opinions that were from the last 5 years or less.

Advantages of Our Approach

Although representing priorities and conceptions of individuals involved in a relation has already been shown to be beneficial to communication and efficacy, using web-based data offers the ability to include a larger number of individuals as shown here from the forum. This would allow for better coverage of the diverse opinion and reflect differences in experience. We also found that forum posts and PubMed papers presented with equivalent density of coverage for all domains examined, suggesting that they present a richness in perspectives and not only trends for instance. Moreover, in the future, our approach could be used to compare concerns of individuals in different countries, in city versus rural settings, or for newcomers to a country, for example. Obtaining the related concepts from the corpus-based VSM and representing those as connected nodes in a property graph model–based KG helps identify convergent and divergent concepts by using different dimensions of interpretability. Node size, which is the frequency of concepts in documents about a condition, tells how widely the concept is discussed in a source. Edge thickness, which is proportional to semantic relatedness score, tells how related a concept is to the condition (ASD or ADHD) depending upon the context in which it is used. This is important as it can help focus attention for knowledge translation and medical education and policy and research development.

Limitations of Our Approach

Some of the limitations relate to the nature of the data used to construct the graph. Forum posts present some challenges. The forums do not precisely define if the users are parents, caregivers, or potentially family members of individuals with ASD and ADHD. This may influence the type of information requested. In addition, the users are by definition selected on the basis of them using technology to gather information. This could represent a bias based on access to technology, which would be influenced by social determinants of health and therefore could have an incomplete representation of the concerns of parents. In addition, owing to concerns about confidentiality, parents may not share all the concerns they have about their family member with ADHD or ASD. Another important point is that health care is represented by PubMed literature here. Although it is true that PubMed represents a high-quality corpus of medical literature, it may not reflect completely what would be discussed by health care providers, say using web-based forums if they were present. In addition, from a technical standpoint, our proposed semantic relatedness–based KG representation utilizes only the categorical information about the UMLS concepts, which is indicated by the "isA" relationship in KG. However, UMLS provides a semantic network, which shows several meaningful relationships between different semantic types in the form of triples, that is, type1, relation, type2, etc: for instance, ("Mental or Behavioral Dysfunction," "associated_with," "Daily or Recreational Activity") and ("Disease or Syndrome," "co-occurs_with," "Mental or behavioral dysfunction"). Utilizing this information could provide more meaningful and direct relations between the concepts of different semantic types. We aim to apply the distantly supervised relation extraction approach on each document corpus, which utilizes the UMLS semantic network to obtain diverse relations between different concepts [67,68]. The output of this approach can also be used as training data for deep learning algorithms to train relation extraction models, which would allow us to create KG by processing text corpus not only for the NDD domain but also for any other condition.

Conclusion

Our study shows the benefits of using KGs developed based on the results of NLP analysis of a text. The graphs representing the mental models of key concerns from parents of individuals with ASD and ADHD are compared to those built on medical expert knowledge in the same field. The comparison allows identifying points of overlapping and diverging interest. We showed that there are several points of convergence and an extensive list of concerns in both types of stakeholders. This is important, as obtaining such information directly from stakeholders requires extensive effort for recruitment and conducting of interviews or distribution of surveys (with often limited response rate). Furthermore, we found that published reports of polling or interviews with ADHD or ASD families or medical experts identified similar concerns to what we identified through NLP and the comparison of graphs. Future field work would complement our work, which could help understand how different concepts present with complex interactions or how specific populations may differ from one another based on different factors such as social determinants of health.

Acknowledgments

We would like to thank Dr Osmar Zaiane and colleagues at the University of Alberta for helpful discussions. This work is funded by the Canadian Institute of Health Research and Natural Science and Engineering Research Council operating grant to FVB. PubMed data are collected by courtesy of the US National Library of Medicine and are shared in Multimedia Appendix 11 under National Center for Biotechnology Information's terms and conditions [69]. This data set will not reflect the most current/accurate data available from the National Library of Medicine in future.

Authors' Contributions

FVB conceptualized the project. MZR and MK designed the methodology. MK implemented the text analysis pipeline and contributed to the analysis of the results. FVB and EW analyzed the results. MK, EW, JC, KK, MZR, and FVB wrote the manuscript. FVB and MZR supervised the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bar graph of frequent semantic types in PubMed and forum.
[PNG File , 173 KB - jmir_v24i8e39888_app1.png]

Multimedia Appendix 2
Autism spectrum disorder and attention deficit hyperactivity disorder concept rankings with relatedness scores (multiple sheets).
[XLSX File (Microsoft Excel File), 411 KB - jmir_v24i8e39888_app2.xlsx]

Multimedia Appendix 3
Mapping of unified medical language system canonical concepts to words in forum and PubMed text.
[PDF File (Adobe PDF File), 60 KB - jmir_v24i8e39888_app3.pdf]

Multimedia Appendix 4
Knowledge graph of attention deficit hyperactivity disorder PubMed-selected semantic types and concepts (split into subgraphs for clarity).
[PNG File , 4182 KB - jmir_v24i8e39888_app4.png]

Multimedia Appendix 5
Knowledge graph of attention deficit hyperactivity disorder forum-selected semantic types and concepts (split into subgraphs for clarity).
[PNG File , 3505 KB - jmir_v24i8e39888_app5.png]

Multimedia Appendix 6
Knowledge graph of autism spectrum disorder PubMed-selected semantic types and concepts (split into subgraphs for clarity).
[PNG File , 3775 KB - jmir_v24i8e39888_app6.png]

Multimedia Appendix 7
Knowledge graph of autism spectrum disorder forum-selected semantic types and concepts (split into subgraphs for clarity).
[PNG File , 3451 KB - jmir_v24i8e39888_app7.png]

Multimedia Appendix 8
Knowledge graph representing similarities and differences between the most relevant attention deficit hyperactivity disorder concepts.
[PNG File , 3589 KB - jmir_v24i8e39888_app8.png]

Multimedia Appendix 9
Comparison of concept relatedness scores in forum and PubMed.
[PDF File (Adobe PDF File), 185 KB - jmir_v24i8e39888_app9.pdf]

Multimedia Appendix 10
Knowledge graph representing similarities and differences between the most relevant autism spectrum disorder concepts.
[PNG File , 2995 KB - jmir_v24i8e39888_app10.png]

Multimedia Appendix 11
PubMed abstract data set.
[ZIP File (Zip Archive), 23963 KB - jmir_v24i8e39888_app11.zip]

References

1. Ismail FY, Shapiro BK. What are neurodevelopmental disorders? Curr Opin Neurol 2019 Aug;32(4):611 [FREE Full text] [doi: 10.1097/wco.0000000000000710]
2. Skounti M, Philalithis A, Galanakis E. Variations in prevalence of attention deficit hyperactivity disorder worldwide. Eur J Pediatr 2007 Feb;166(2):117-123. [doi: 10.1007/s00431-006-0299-5] [Medline: 17033803]

3. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry 2015 Mar;56(3):345-365. [doi: 10.1111/jcpp.12381] [Medline: 25649325]
4. Chiarotti F, Venerosi A. Epidemiology of Autism Spectrum Disorders: A Review of Worldwide Prevalence Estimates Since 2014. Brain Sci 2020 May 01;10(5):274 [FREE Full text] [doi: 10.3390/brainsci10050274] [Medline: 32370097]
5. Gentner D. Mental Models, Psychology of. In: International Encyclopedia of the Social & Behavioral Sciences. Oxford: Pergamon; Dec 2001:9683-9687.
6. Crandall B, Klein G, Hoffman R. Working Minds: A Practitioner's Guide to Cognitive Task Analysis. Cambridge: MIT Press; 2006.
7. Lewis KB, Stacey D, Matlock DD. Making decisions about implantable cardioverter-defibrillators from implantation to end of life: an integrative review of patients' perspectives. Patient 2014;7(3):243-260. [doi: 10.1007/s40271-014-0055-2] [Medline: 24668214]
8. Holtrop JS, Scherer LD, Matlock DD, Glasgow RE, Green LA. The Importance of Mental Models in Implementation Science. Front Public Health 2021;9:680316 [FREE Full text] [doi: 10.3389/fpubh.2021.680316] [Medline: 34295871]
9. Ebener S, Khan A, Shademani R, Compernolle L, Beltran M, Lansang M, et al. Knowledge mapping as a technique to support knowledge translation. Bull World Health Organ 2006 Aug;84(8):636-642 [FREE Full text] [doi: 10.2471/blt.06.029736] [Medline: 16917651]
10. Zhang C, Zhang J, Long C, Zheng J, Su C, Hu W, et al. Analyses of research on the health of college students based on a perspective of knowledge mapping. Public Health 2016 Aug;137:188-191. [doi: 10.1016/j.puhe.2015.11.002] [Medline: 26715319]
11. Adams S, Nicholas D, Mahant S, Weiser N, Kanani R, Boydell K, et al. Care maps and care plans for children with medical complexity. Child Care Health Dev 2019 Jan;45(1):104-110. [doi: 10.1111/cch.12632] [Medline: 30462842]
12. Ji S, Pan S, Cambria E, Marttinen P, Yu PS. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. IEEE Trans. Neural Netw. Learning Syst 2022 Feb;33(2):494-514. [doi: 10.1109/tnnls.2021.3070843]
13. Andy A, Andy U. Understanding Communication in an Online Cancer Forum: Content Analysis Study. JMIR Cancer 2021 Sep 07;7(3):e29555 [FREE Full text] [doi: 10.2196/29555] [Medline: 34491209]
14. Ping Q, Yang CC, Marshall SA, Avis NE, Ip EH. Breast Cancer Symptom Clusters Derived from Social Media and Research Study Data Using Improved K-Medoid Clustering. IEEE Trans Comput Soc Syst 2016 Jun;3(2):63-74 [FREE Full text] [doi: 10.1109/TCSS.2016.2615850] [Medline: 29152536]
15. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What Patients Can Tell Us: Topic Analysis for Social Media on Breast Cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]
16. Gemmell J, Isenegger K, Dong Y, Glaser E, Morain A. Comparing Automatically Extracted Topics from Online Mental Health Disorder Forums. 2019 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); December 5-7; Las Vegas. [doi: 10.1109/csci49370.2019.00252]
17. Jiang L, Yang CC. Using Co-occurrence Analysis to Expand Consumer Health Vocabularies from Social Media Data. 2013 Presented at: IEEE International Conference on Healthcare Informatics; September 9-11; Philadelphia. [doi: 10.1109/ICHI.2013.16]
18. Liu M, Zou X, Chen J, Ma S. Comparative Analysis of Social Support in Online Health Communities Using a Word Co-Occurrence Network Analysis Approach. Entropy (Basel) 2022 Jan 25;24(2):174 [FREE Full text] [doi: 10.3390/e24020174] [Medline: 35205469]
19. Dyer J, Kolic B. Public risk perception and emotion on Twitter during the Covid-19 pandemic. Appl Netw Sci 2020;5(1):99 [FREE Full text] [doi: 10.1007/s41109-020-00334-7] [Medline: 33344760]
20. Turdakov D, Velikhov P. Semantic Relatedness Metric for Wikipedia Concepts Based on Link Analysis and its Application to Word Sense Disambiguation. 2008 Presented at: Proceedings of the Fifth Spring Young Researchers Colloquium on Databases and Information Systems, SYRCoDIS'2008; May 29-30; St Petersburg, Russia URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.143.864
21. Mao Y, Fung KW. Use of Word and Graph Embedding to Measure Semantic Relatedness Between Unified Medical Language System Concepts. J Am Med Inform Assoc 2020 Oct 01;27(10):1538-1546 [FREE Full text] [doi: 10.1093/jamia/ocaa136] [Medline: 33029614]
22. Zhou T, Law KM. Semantic Relatedness Enhanced Graph Network for aspect category sentiment analysis. Expert Systems with Applications 2022 Jun;195:116560. [doi: 10.1016/j.eswa.2022.116560]
23. Aggarwal N, Buitelaar P. Query expansion using Wikipedia and Dbpedia. 2012 Presented at: CLEF (Online Working Notes/Labs/Workshop); September 17-20; Rome, Italy URL: http://ceur-ws.org/Vol-1178/CLEF2012wn-CHiC-AggarwalEt2012.pdf/>
24. Schulz C, Levy-Kramer J, Van AC, Kepes M, Hammerla N. Biomedical concept relatedness: a large EHR-based benchmark. 2020 Dec Presented at: Proceedings of the 28th International Conference on Computational Linguistics; December; Barcelona, Spain. [doi: 10.18653/v1/2020.coling-main.577]
25. HealthBoards: Find a Forum. URL: https://www.healthboards.com/boards/healthAZ.php [accessed 2022-01-31]

26. Psychology and Mental Health Forum. URL: https://www.psychforums.com/ [accessed 2022-01-31]
27. Reddit. URL: https://www.reddit.com/ [accessed 2022-01-31]
28. Scrapy 2.6. URL: https://docs.scrapy.org/en/latest/ [accessed 2022-01-15]
29. Podolak M. PMAW: A Multithread Pushshift.io API Wrapper for Reddit. URL: https://github.com/mattpodolak/pmaw [accessed 2022-01-20]
30. Bird S, Klein E, Loper E. Natural Language Processing With Python: Analyzing Text With the Natural Language Toolkit. Sebastopol, CA: O'Reilly Media, Inc; 2009.
31. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
32. Neumann M, King D, Beltagy I, Ammar W. ScispaCy: Fast and robust models for biomedical natural language processing. 2019 Presented at: Proceedings of the 18th BioNLP Workshop and Shared Task; August; Florence, Italy. [doi: 10.18653/v1/w19-5034]
33. Peng J, Zhao M, Havrilla J, Liu C, Weng C, Guthrie W, et al. Natural language processing (NLP) tools in extracting biomedical concepts from research articles: a case study on autism spectrum disorder. BMC Med Inform Decis Mak 2020 Dec 30;20(Suppl 11):322 [FREE Full text] [doi: 10.1186/s12911-020-01352-2] [Medline: 33380331]
34. Xun G, Li Y, Gao J, Zhang A. Collaboratively improving topic discovery and word embedding by coordinating global and local contexts. 2017 Aug Presented at: KDD'17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August; Halifax, NS, Canada. [doi: 10.1145/3097983.3098009]
35. Levy O, Goldberg Y. Neural word embedding as implicit matrix factorization. 2014 Presented at: Advances in Neural Information Processing Systems; December; Montreal, Canada URL: https://proceedings.neurips.cc/paper/2014/hash/feab05aa91085b7a8012516bc3533958-Abstract.html/>
36. Hellrich J, Hahn U. Exploring Diachronic Lexical Semantics with JeSemE. 2017 Presented at: Proceedings of ACL 2017, System Demonstrations; July; Vancouver, Canada URL: https://aclanthology.org/P17-4006/ [doi: 10.18653/v1/p17-4006]
37. Chugh M, Whigham P, Dick G. Stability of word embeddings Using Word2Vec. 2018 Presented at: AI 2018: Advances in Artificial Intelligence; December 11-14; Wellington, New Zealand p. 812-818. [doi: 10.1007/978-3-030-03991-2_73]
38. Pedregosa F, Varoquaux G, Gramfort A, Thirion B, Michel V, Thirion B, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825-2830 [FREE Full text]
39. Hamilton WL, Leskovec J, Jurafsky D. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. 2016 Aug Presented at: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); August; Berlin, Germany p. 1489-1501. [doi: 10.18653/v1/P16-1141]
40. Salle A, Villavicencio A, Idiart M. Matrix Factorization Using Window Sampling and Negative Sampling for Improved Word Representations. 2016 Presented at: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers); August; Berlin, Germany. [doi: 10.18653/v1/P16-2068]
41. Nguyen KA, Schulte IWS, Vu NT. Introducing two Vietnamese Datasets for Evaluating Semantic Models of (dis-)Similarity and Relatedness. 2018 Presented at: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers); June; New Orleans, Louisiana. [doi: 10.18653/v1/N18-2032]
42. Neo4j Graph Data Platform. URL: https://neo4j.com/ [accessed 2022-01-31]
43. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. ICWSM. 2009 Mar 19. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/13937 [accessed 2022-07-21]
44. Seo S, Kennedy-Metz LR, Zenati MA, Shah JA, Dias RD, Unhelkar VV. Towards an AI Coach to Infer Team Mental Model Alignment in Healthcare. IEEE CogSIMA (2021) 2021 May;2021:39-44 [FREE Full text] [doi: 10.1109/cogsima51574.2021.9475925] [Medline: 35253018]
45. Sturgiss E, Luig T, Campbell-Scherer DL, Lewanczuk R, Green LA. Using Concept Maps to compare obesity knowledge between policy makers and primary care researchers in Canada. BMC Res Notes 2019 Jan 14;12(1):23 [FREE Full text] [doi: 10.1186/s13104-018-4042-x] [Medline: 30642382]
46. McComb S, Simpson V. The concept of shared mental models in healthcare collaboration. J Adv Nurs 2014 Jul;70(7):1479-1488. [doi: 10.1111/jan.12307] [Medline: 24237202]
47. Mogford RH. Mental Models and Situation Awareness in Air Traffic Control. The International Journal of Aviation Psychology 1997 Oct;7(4):331-341. [doi: 10.1207/s15327108ijap0704_5]
48. Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers JA. The influence of shared mental models on team process and performance. J Appl Psychol 2000 Apr;85(2):273-283. [doi: 10.1037/0021-9010.85.2.273] [Medline: 10783543]
49. Davies M. Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? High Educ 2010 Nov 27;62(3):279-301. [doi: 10.1007/s10734-010-9387-6]
50. Johnston C, Hommersen P, Seipp C. Acceptability of behavioral and pharmacological treatments for attention-deficit/hyperactivity disorder: relations to child and parent characteristics. Behav Ther 2008 Mar;39(1):22-32. [doi: 10.1016/j.beth.2007.04.002] [Medline: 18328867]
51. Clarke JN, Lang L. Mothers whose children have ADD/ADHD discuss their children's medication use: an investigation of blogs. Soc Work Health Care 2012;51(5):402-416. [doi: 10.1080/00981389.2012.660567] [Medline: 22583027]

52. Bussing R, Zima BT, Mason DM, Meyer JM, White K, Garvan CW. ADHD knowledge, perceptions, and information sources: perspectives from a community sample of adolescents and their parents. J Adolesc Health 2012 Dec;51(6):593-600 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.03.004] [Medline: 23174470]
53. Ross M, Bridges JFP, Ng X, Wagner LD, Frosch E, Reeves G, et al. A best-worst scaling experiment to prioritize caregiver concerns about ADHD medication for children. Psychiatr Serv 2015 Feb 01;66(2):208-211 [FREE Full text] [doi: 10.1176/appi.ps.201300525] [Medline: 25642618]
54. Ross M, Nguyen V, Bridges JFP, Ng X, Reeves G, Frosch E, et al. Caregivers' Priorities and Observed Outcomes of Attention-Deficit Hyperactivity Disorder Medication for Their Children. J Dev Behav Pediatr 2018;39(2):93-100 [FREE Full text] [doi: 10.1097/dbp.0000000000000530]
55. Mühlbacher AC, Rudolph I, Lincke H, Nübling M. Preferences for treatment of Attention-Deficit/Hyperactivity Disorder (ADHD): a discrete choice experiment. BMC Health Serv Res 2009 Aug 13;9:149 [FREE Full text] [doi: 10.1186/1472-6963-9-149] [Medline: 19678946]
56. Lu SV, Leung BMY, Bruton AM, Millington E, Alexander E, Camden K, et al. Parents' priorities and preferences for treatment of children with ADHD: Qualitative inquiry in the MADDY study. Child Care Health Dev Preprint posted online on March 4, 2022. [doi: 10.1111/cch.12995] [Medline: 35244227]
57. Ferguson JH. National Institutes of Health Consensus Development Conference Statement: diagnosis and treatment of attention-deficit/hyperactivity disorder (ADHD). J Am Acad Child Adolesc Psychiatry 2000 Feb;39(2):182-193 [FREE Full text] [doi: 10.1097/00004583-200002000-00018] [Medline: 10673829]
58. Colvin MK, Stern TA. Diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder. J Clin Psychiatry 2015 Sep;76(9):e1148 [FREE Full text] [doi: 10.4088/JCP.12040vr1c] [Medline: 26455686]
59. Jangmo A, Stålhandske A, Chang Z, Chen Q, Almqvist C, Feldman I, et al. Attention-Deficit/Hyperactivity Disorder, School Performance, and Effect of Medication. J Am Acad Child Adolesc Psychiatry 2019 Apr;58(4):423-432 [FREE Full text] [doi: 10.1016/j.jaac.2018.11.014] [Medline: 30768391]
60. Lambez B, Harwood-Gross A, Golumbic EZ, Rassovsky Y. Non-pharmacological interventions for cognitive difficulties in ADHD: A systematic review and meta-analysis. J Psychiatr Res 2020 Jan;120:40-55. [doi: 10.1016/j.jpsychires.2019.10.007] [Medline: 31629998]
61. Lai JKY, Weiss JA. Priority service needs and receipt across the lifespan for individuals with autism spectrum disorder. Autism Res 2017 Aug;10(8):1436-1447 [FREE Full text] [doi: 10.1002/aur.1786] [Medline: 28383156]
62. Azad G, Mandell DS. Concerns of parents and teachers of children with autism in elementary school. Autism 2016 May;20(4):435-441 [FREE Full text] [doi: 10.1177/1362361315588199] [Medline: 26069200]
63. Pejovic-Milovancevic M, Stankovic M, Mitkovic-Voncina M, Rudic N, Grujicic R, Herrera AS, et al. Perceptions on Support, Challenges and Needs among Parents of Children with Autism: the Serbian Experience. Psychiatr Danub 2018 Sep;30(Suppl 6):354-364 [FREE Full text] [Medline: 30235173]
64. Lee PF, Thomas RE, Lee PA. Approach to autism spectrum disorder: Using the new DSM-V diagnostic criteria and the CanMEDS-FM framework. Can Fam Physician 2015 May;61(5):421-424 [FREE Full text] [Medline: 25971758]
65. Knott F, Dunlop A, Mackay T. Living with ASD: how do children and their parents assess their difficulties with social interaction and understanding? Autism 2006 Nov;10(6):609-617. [doi: 10.1177/1362361306068510] [Medline: 17088276]
66. Pellicano E, Dinsmore A, Charman T. What should autism research focus upon? Community views and priorities from the United Kingdom. Autism 2014 Oct;18(7):756-770 [FREE Full text] [doi: 10.1177/1362361314529627] [Medline: 24789871]
67. Zhu T, Qin Y, Xiang Y, Hu B, Chen Q, Peng W. Distantly supervised biomedical relation extraction using piecewise attentive convolutional neural network and reinforcement learning. J Am Med Inform Assoc 2021 Nov 25;28(12):2571-2581. [doi: 10.1093/jamia/ocab176] [Medline: 34524450]
68. Xing R, Luo J, Song T. BioRel: towards large-scale biomedical relation extraction. BMC Bioinformatics 2020 Dec 16;21(Suppl 16):543 [FREE Full text] [doi: 10.1186/s12859-020-03889-5] [Medline: 33323106]
69. National library of medicine terms and conditions. National Library of Medicine. URL: https://www.nlm.nih.gov/databases/download/terms_and_conditions.html [accessed 2022-01-10]

Abbreviations

ADHD: attention deficit hyperactivity disorder
ASD: autism spectrum disorder
KG: knowledge graph
NDD: neurodevelopmental disorder
NLP: natural language processing
PMI: pointwise mutual information
PPMI: positive pointwise mutual information
SVD: singular value decomposition
UMLS: unified medical language system
VSM: vector space model

Edited by G Eysenbach, T Leung; submitted 26.05.22; peer-reviewed by S Chen, KM Thaker, Y Liu; comments to author 16.06.22; revised version received 05.07.22; accepted 08.07.22; published 05.08.22.

Please cite as:
Kaur M, Costello J, Willis E, Kelm K, Reformat MZ, Bolduc FV
Deciphering the Diversity of Mental Models in Neurodevelopmental Disorders: Knowledge Graph Representation of Public Data Using Natural Language Processing
J Med Internet Res 2022;24(8):e39888
URL: https://www.jmir.org/2022/8/e39888
doi:10.2196/39888
PMID:35930346

Original Paper

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:

This is a corrected version. See correction statement: https://www.jmir.org/2022/10/e42716

Abstract

Background: During the COVID-19 pandemic, there have been increased reports of racial biases against Asian American and Native Hawaiian and Pacific Islander individuals. However, the extent to which different Asian American and Native Hawaiian and Pacific Islander groups perceive and experience (firsthand or as a witness to such experiences) how COVID-19 has negatively affected people of their race has not received much attention.

Objective: This study used data from the COVID-19 Effects on the Mental and Physical Health of Asian Americans and Pacific Islanders Survey Study (COMPASS), a nationwide, multilingual survey, to empirically examine COVID-19–related racial bias beliefs among Asian American and Native Hawaiian and Pacific Islander individuals and the factors associated with these beliefs.

Methods: COMPASS participants were Asian American and Native Hawaiian and Pacific Islander adults who were able to speak English, Chinese (Cantonese or Mandarin), Korean, Samoan, or Vietnamese and who resided in the United States during the time of the survey (October 2020 to May 2021). Participants completed the survey on the web, via phone, or in person. The Coronavirus Racial Bias Scale (CRBS) was used to assess COVID-19–related racial bias beliefs toward Asian American and Native Hawaiian and Pacific Islander individuals. Participants were asked to rate the degree to which they agreed with 9 statements on a 5-point Likert scale (ie, *1=strongly disagree* to *5=strongly agree*). Multivariable linear regression was used to examine the associations between demographic, health, and COVID-19–related characteristics and perceived racial bias.

Results: A total of 5068 participants completed the survey (mean age 45.4, SD 16.4 years; range 18-97 years). Overall, 73.97% (3749/5068) agreed or strongly agreed with ≥1 COVID-19–related racial bias belief in the past 6 months (during the COVID-19 pandemic). Across the 9 racial bias beliefs, participants scored an average of 2.59 (SD 0.96, range 1-5). Adjusted analyses revealed that compared with Asian Indians, those who were ethnic Chinese, Filipino, Hmong, Japanese, Korean, Vietnamese, and other or multicultural had significantly higher mean CRBS scores, whereas no significant differences were found among Native Hawaiian and Pacific Islander individuals. Nonheterosexual participants had statistically significant and higher mean CRBS scores than heterosexual participants. Compared with participants aged ≥60 years, those who were younger (aged <30, 30-39, 40-49, and 50-59 years) had significantly higher mean CRBS scores. US-born participants had significantly higher mean CRBS scores than foreign-born participants, whereas those with limited English proficiency (relative to those reporting no limitation) had lower mean CRBS scores.

Conclusions: Many COMPASS participants reported racial bias beliefs because of the COVID-19 pandemic. Relevant sociodemographic contexts and pre-existing and COVID-19–specific factors across individual, community, and society levels were associated with the perceived racial bias of *being Asian* during the pandemic. The findings underscore the importance of addressing the burden of racial bias on Asian American and Native Hawaiian and Pacific Islander communities among other COVID-19–related sequelae.

(J Med Internet Res 2022;24(8):e38443) doi:10.2196/38443

KEYWORDS

COVID-19; racial bias; Asian American; Native Hawaiian and Pacific Islander; mobile phone

Introduction

Background

Over the past 2 years, Asian American individuals have been facing a dual pandemic—COVID-19, as well as increased experiences and fear of racial bias, discrimination, and hate. This has been fueled by racist rhetoric (eg, linking the COVID-19 pandemic with China) rather than a neutral framing of the virus in public health messaging as part of the pandemic response [1,2]. A recent review highlighted the increased anxiety associated with being Asian American during the COVID-19 pandemic among Asian American individuals overall and individuals of Chinese or East Asian descent in particular [3].

Although it is not new, reports of discrimination and hate incidents have been on the rise since the start of the pandemic. An April 2020 survey of 1001 adults found that 32% reported witnessing someone blaming Asians for the COVID-19 pandemic, whereas 60% of the Asian respondents had witnessed the same [4]. In a later public poll, Asian American respondents were most likely (than White, Hispanic, and Black adults) to report that they had experienced people acting as if they were uncomfortable around them (39%), been subject to slurs or jokes (31%), and feared that someone might threaten or physically attack them (26%) because of their race since the COVID-19 outbreak [5]. According to Stop AAPI Hate, >10,000 reports of anti-Asian hate incidents were reported on this site between March 2020 and September 2021 [6]. Compared with reports of anti-Asian hate incidents made in 2020, of the total number of hate incidents, physical assaults rose from 10.8% to 16.6%, and web-based hate incidents increased from 6.1% to 10.6% in 2021; furthermore, these hate incidents occurred more often in public streets, schools, and places of worship [6]. Few studies collected data on discrimination experiences or incidents broken down by Asian American and Native Hawaiian and Pacific Islander ethnic groups. A large survey (N=4971) of Asian American and Pacific Islander participants found that 60.7% experienced discrimination during the COVID-19 pandemic, and reports of these experiences were particularly high for Hmong (80%), ethnic Chinese (64.7%), Korean (64.2%), Filipino (61.3%), Japanese (57.7%), Vietnamese (55.7%), Asian Indian (41.5%), and Native Hawaiian and Pacific Islander (40.5%) participants [7]. In another survey of >2400 Asian American and Pacific Islander women during January and February of 2022, Native Hawaiian and Pacific Islander women reported the highest percentage (80%) of reporting racism or discrimination incidents or harassment, followed by similar levels reported by East Asian (72%), South Asian (73%), and Southeast Asian (75%) women [8].

There has been substantial coverage of anti-Asian incidents and hate crimes via popular news media, reports from Stop AAPI Hate, and findings from multiple surveys [4,5,8,9]. However, there is limited understanding of the extent to which Asian American and Native Hawaiian and Pacific Islander individuals' perceptions and experiences (firsthand or as a witness) relate to how COVID-19 has negatively affected people of their race, particularly about perceived racial bias sentiments by diverse Asian American and Native Hawaiian and Pacific Islander populations. Racial bias is defined as a personal or unreasonable belief or judgment of a person based on their race and is rooted in stereotypes and prejudices [10]. Prior research has shown a multitude of impacts of racial bias on one's lived experiences in the form of legal decisions regarding disparities in education and economic opportunities [11], as well as health care and health status [12,13].

Objectives

Social determinants of health and contextual models suggest that race and ethnicity, socioeconomic status, and social environment are important factors related to experiences of discrimination and racial bias [14,15]. Research on discrimination and Asian American individuals has shown that perceived discrimination varies by age, gender, and other demographic characteristics [7,16,17]. Drawing on these frameworks and previous research, we used data from the COVID-19 Effects on the Mental and Physical Health of Asian Americans and Pacific Islanders Survey Study (COMPASS), a nationwide, multilingual, community-based survey conducted in the United States to empirically examine racial bias beliefs on Asian American and Native Hawaiian and Pacific Islander individuals, as related to the COVID-19 pandemic, and a cross-sectional analysis of the factors associated with these beliefs. Uniquely, we also examined COVID-19–related racial bias beliefs in different Asian American and Native Hawaiian and Pacific Islander ethnic subgroups. Data disaggregation is important and pivotal for identifying the distinct challenges and needs of diverse Asian American and Native Hawaiian and Pacific Islander communities.

Methods

Data Source

Between October 2020 and February 2021, a total of 5218 Asian American and Native Hawaiian and Pacific Islander adults completed the COMPASS survey. COMPASS is a multilingual community-based nationwide survey that assesses the effects of COVID-19 on Asian Americans and Native Hawaiian and Pacific Islanders. The eligibility criteria for participating in the COMPASS included (1) self-identifying as Asian American or Native Hawaiian and Pacific Islander alone or in combination with other races and ethnicities; (2) being aged at least 18 years; (3) residing in the United States; and (4) able to speak or read English, Chinese (traditional or simplified), Korean, Samoan, or Vietnamese. These languages were selected as they represent some of the commonly spoken languages among Asian American and Pacific Islander individuals with limited English proficiency (LEP) [18], and this was the language capacity supported by the parent award for COMPASS at the time of the study [19]. Participants completed the survey on the web via the study's website [19], via phone, or in person in the abovementioned languages. We used the World Health Organization's process of translating and adapting instruments [20] to guide translations of study materials not readily available in the targeted Asian American and Native Hawaiian and Pacific Islander languages. REDCap (Research Electronic Data Capture; Vanderbilt University) [21,22] was used to capture and store data securely.

Participants could have heard about COMPASS through community partners who serve Asian American and Native Hawaiian and Pacific Islander individuals, personal or professional networks, social media, emails or listservs, flyers, and ethnic media. COMPASS also recruited from the Collaborative Approach for Asian Americans and Pacific Islanders Research and Education (CARE) registry [19], which is the first and only research recruitment registry that purposively engages Asian American and Native Hawaiian and Pacific Islander participants in multiple languages via strong community partnerships. Of the 2600 CARE registry participants who received an email invitation to participate in COMPASS, 526 (20.23%) completed the COMPASS survey. Approximately 86.64% (4521/5218) of participants completed the survey by themselves, and 11.33% (591/5218) of participants received help from family, friends, or research staff. The participants had the option to receive a US $10 electronic gift card upon survey completion. All participants provided informed consent for inclusion before participating in the study.

Measures

Overview

The measurement framework for this study was guided by the previously discussed literature on social determinants of health and contextual models suggesting that sociodemographic contexts and social environments are important factors related to experiences of racial bias [14,15]. We posit that there are multiple influences on COVID-19–related racial bias beliefs that operate within the sociodemographic contexts of Asian American and Native Hawaiian and Pacific Islander individuals (eg, cultural group, age, sex, sexual orientation, education, income, employment, marital status, English proficiency, nativity, percentage of life spent in the United States, geographic regions, and survey completion month), which intersect with individuals' experiences of COVID-19–specific impacts. These included individuals' COVID-19 status; perceived severity of COVID-19 in one's neighborhood or community compared with others; length of shelter-in-place (SIP); and COVID-19 effect on family income or employment, social support, and medical and mental health care access.

COVID-19–Related Racial Bias Beliefs

The 9-item Coronavirus Racial Bias Scale (CRBS) [23], accessed via the PhenX Toolkit [24], was used to assess beliefs about how the COVID-19 pandemic affected public attitudes (eg, the country becoming more dangerous), racial or ethnic biases affecting employment and access to health services, and racially charged social media and cyberbullying toward Asian American and Native Hawaiian and Pacific Islander individuals in the United States. Participants were asked to rate the degree to which they agreed with 9 statements on a 5-point Likert scale: 1=strongly disagree, 2=somewhat disagree, 3=neutral, 4=somewhat agree, and 5=strongly agree. Cronbach α for this study sample was .90; the α of CRBS ranged from .85 to .92 across the 5 survey languages. For participants who responded to all the 9 items (5068/5218, 97.12%), the total score was computed by averaging all the items (scores ranged from 1 to 5). A higher score indicated a greater degree of agreement with COVID-19–related racial bias beliefs.

Sociodemographic Characteristics

Participants' demographic characteristics included cultural group (Asian Indian, Ethnic Chinese, Filipino, Hmong, Japanese, Korean, Native Hawaiian and Pacific Islander, Vietnamese, and other or multicultural), age (<30, 30-39, 40-49, 50-59, and ≥60 years), sex (male, female, and other or decline to state), sexual orientation (heterosexual, not heterosexual, and decline to state), education level (high school or less, some college or technical school, bachelor's degree, and master's degree or higher), annual household income (≤US $25,000, >US $25,000 to US $75,000, >US $75,000 to US $150,000, >150,000, and decline to state), employment status (full-time, part-time, homemaker, unemployed, retired, and other), and marital status (single, married or living with partner, separated or divorced, and widowed). Self-rated English proficiency was assessed using "How well can you speak, read, and/or write English?" with responses of *a little bit* or *not at all* categorized as having LEP. Nativity was assessed using yes or no or whether the participant was born in the United States. For US-born participants, the percentage of life spent in the United States was 100%, and for non–US-born participants, this was calculated by subtracting the age of entry into the United States from the current age and dividing the current age. The US region (Midwest, Northeast, South, and West) was determined by converting the zip code or IP address in cases of missing zip codes (143/5068, 2.82%) per the Census Bureau definition [25]. The survey completion date was classified by month and year.

COVID-19–Related Experiences and Impacts

Individuals' *COVID-19 status* was measured by asking participants, "Have you been diagnosed with COVID-19 by a doctor or other health care provider?" The responses were recorded as yes, no, or unsure. This item was taken from the questionnaire for assessing the impact of the COVID-19 pandemic and accompanying mitigation efforts on older adults [26].

Perceived severity of COVID-19 was a single item developed by the COMPASS study team. Participants were asked, "How would you rate the severity of COVID-19 outbreak at where you live in comparison to other locations in the US?" The response options were 1 (much less severe than most other places in the United States), 2 (somewhat less severe), 3 (about the same), 4 (somewhat more severe), and 5 (much more severe).

The length of the SIP item was also developed by COMPASS. Participants were asked, "How long was the SIP (or stay-at-home) order at where you live?" The response options were 0 (no order), 1 (<1 month), 2 (1-2 months), and 3 (≥2 months).

Four items from the Coronavirus Impact Scale [27] were used to measure the impact of COVID-19 on four areas: family income or employment, social support, and access to medical and mental health care. Participants were asked to rate the extent to which COVID-19 changed their lives in each of the four areas using response options: 0 (no change), 1 (mild change), 2 (moderate), or 3 (severe). For medical health care access and mental health treatment access, participants were provided *not applicable* as a response option in addition to the 0 to 3 options.

Statistical Analyses

Descriptive statistics were computed to describe the study sample. Given that this study was among the first to use the CRBS with a large diverse sample of Asian American and Native Hawaiian and Pacific Islander individuals in multiple languages, we reported the percentages endorsed for each CRBS item. For descriptive purposes only, to provide an overall description of participants' endorsement of any of the COVID-19–related racial bias beliefs measured by the CRBS, we created a dichotomous variable—*any racial bias belief*—from the CRBS items. Participants who responded *strongly agree* or *agree* to any of the 9 CRBS items were classified as having COVID-19–related racial bias beliefs, and those who did not respond *strongly agree* or *agree* to any of the statements were classified as not having these beliefs. Chi-square tests were used to examine associations between sample characteristics and the presence of racial bias beliefs.

To examine the associations among demographic, health, and COVID-19–related characteristics and COVID-19–related racial bias beliefs, we fit unadjusted and fully adjusted multivariable linear regression models using the CRBS score as a continuous outcome. To ensure that all the important variables were included, this study selected a P value of <.10 a priori to include candidate variables in the final model. To avoid potential collinearity in the final model, as well as to capture the effect of both language proficiency and nativity, these variables were included in the final model, whereas the percentage of life in the United States was not. The model with LEP and nativity and the model with the percentage of life in the United States both performed similarly with respect to the model R^2 (0.33 vs 0.34, respectively). In addition, because of the rapidly evolving landscape of the COVID-19 pandemic, we included the survey completion month as a covariate in the final multivariable regression model. Finally, the selection of reference groups in the linear regression models was based on the group that was least likely to report experiences of racial bias. For example, for cultural group comparisons, Asian Indians were selected as the reference group as they reported the lowest proportion of racial bias perceptions among all the Asian American and Native Hawaiian and Pacific Islander subgroups. All statistical tests were 2-sided.

Ethics Approval

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the University of California San Francisco Institutional Review Board (protocol 20-31925).

Results

Participant Characteristics

A total of 5068 participants contributed to the final study sample. As shown in Table 1, 73.97% (3749/5068) of the participants reported having ≥1 of the 9 COVID-19 racial bias beliefs in the past 6 months (during the COVID-19 pandemic). The mean CRBS score was 2.59 (SD 0.96, range 1-5). The participants primarily identified as ethnic Chinese (including individuals from China, Hong Kong, and Taiwan; 1786/5068, 35.24%), Korean (1132/5068, 22.34%), and Vietnamese (953/5068, 18.8%). The participants had a mean age of 45.4 (SD 16.4, range 18-97) years. Most participants were male (3237/5068, 63.87%), identified as heterosexual (4618/5068, 91.12%), and were married or living with a partner (3288/5068, 64.88%). Approximately 22.99% (1165/5068) of participants had LEP, and approximately two-thirds were foreign born and resided in the Western United States region (3288/5068, 64.88%).

Only 3.39% (172/5068) of the study sample reported testing positive for COVID-19. In bivariate analyses, except for COVID-19 positivity, all demographic characteristics, other

COVID-19–related measures (eg, length of SIP, perceived severity of COVID-19, and changes because of COVID-19), and month of survey completion were significantly associated with having COVID-19–related racial bias beliefs.

Table 1. COVID-19 Effects on the Mental and Physical Health of Asian Americans and Pacific Islanders Survey Study participant characteristics (N=5068).

Variables	All	Having racial bias belief		*P* value
		Yes[a] (n=3749)	No (n=1319)	
CRBS[b] score, mean (SD; range)	2.59 (0.96; 1-5)[a]	2.97 (0.76; 1-5)	1.53 (0.62; 1-5)	N/A[c]
Sociodemographic characteristics, n (%)				
Cultural group				
Asian Indian	298 (5.88)	132 (44.3)	166 (55.7)	<.001
Ethnic Chinese[d]	1786 (35.24)	1508 (84.4)	278 (15.6)	<.001
Filipino	176 (3.47)	148 (84.1)	28 (15.9)	<.001
Hmong	110 (2.17)	104 (94.6)	6 (5.4)	<.001
Japanese	220 (4.34)	190 (86.4)	30 (13.6)	<.001
Korean	1132 (22.34)	687 (60.7)	445 (33.7)	<.001
Native Hawaiian and Pacific Islander	136 (2.68)	63 (46.3)	73 (53.7)	<.001
Vietnamese	953 (18.80)	690 (72.4)	263 (27.6)	<.001
Multicultural	172 (3.4)	159 (92.4)	13 (7.6)	<.001
Other	85 (1.67)	68 (80)	17 (20)	<.001
Age (years)				
Values, mean (SD; range)	45.4 (16.4; 18-97)	43.2 (16.0; 18-95)	51.7 (16.0; 18-97)	N/A
<30	1109 (21.9)	984 (88.7)	125 (11.3)	<.001
30-39	881 (17.4)	682 (77.4)	199 (22.6)	<.001
40-49	923 (18.2)	685 (74.2)	238 (25.8)	<.001
50-59	1048 (20.7)	737 (70.3)	311 (29.7)	<.001
≥60	1107 (21.8)	661 (59.7)	446 (33.8)	<.001
Sex, n (%)				
Female	3237 (63.87)	2469 (76.3)	768 (23.7)	<.001
Male	1788 (35.28)	1249 (69.9)	539 (30.1)	<.001
Other or decline to state	43 (0.8)	31 (72.1)	12 (27.9)	<.001
Sexual orientation, n (%)				
Heterosexual	4618 (91.12)	3388 (73.4)	1230 (26.6)	<.001
Not heterosexual	228 (4.49)	214 (93.9)	14 (6.1)	<.001
Decline to state	222 (4.38)	147 (66.2)	75 (33.8)	<.001
Education, n (%)				
High school or less	808 (16.2)	526 (65.1)	282 (34.9)	<.001
Some college or technical school	593 (11.9)	457 (77.1)	136 (22.9)	<.001
Bachelor’s degree	1836 (36.8)	1420 (77.3)	416 (22.7)	<.001
Master’s degree or higher	1748 (35.1)	1298 (74.3)	450 (25.7)	<.001
Annual household income (US$), n (%)				
≤25,000	847 (15.7)	549 (64.8)	298 (35.2)	<.001
>25,000-75,000	1389 (27.4)	1031 (74.2)	358 (25.8)	<.001
>75,000-150,000	1249 (24.6)	972 (77.8)	277 (22.2)	<.001
>150,000	991 (19.6)	771 (77.8)	220 (22.2)	<.001
Decline to state	592 (11.7)	426 (72)	166 (28)	<.001
Employment status, n (%)				

Variables	All	Having racial bias belief		*P* value
		Yes[a] (n=3749)	No (n=1319)	
Full-time	2334 (46.05)	1793 (76.8)	541 (23.2)	<.001
Part-time	870 (17.16)	637 (73.2)	233 (26.8)	<.001
Homemaker	430 (8.48)	282 (65.6)	148 (34.4)	<.001
Unemployed	529 (10.44)	425 (80.3)	104 (19.7)	<.001
Retired	570 (11.25)	365 (64)	205 (36)	<.001
Other or decline to state	335 (6.68)	247 (73.7)	88 (26.3)	<.001
Marital status, n (%)				
Single	1398 (27.58)	1184 (84.7)	214 (15.3)	<.001
Married or living with partner	3288 (64.88)	2295 (69.8)	993 (30.2)	<.001
Separated, divorced, or widowed	339 (6.68)	239 (70.5)	100 (29.5)	<.001
Declined	43 (0.8)	31 (72.1)	12 (27.9)	<.001
LEP[e], n (%)				
Yes	1165 (22.98)	657 (56.4)	508 (43.6)	<.001
No	3903 (77.01)	3092 (79.2)	811 (20.8)	<.001
Nativity, n (%)				
US-born	1761 (34.75)	1565 (88.9)	196 (11.1)	<.001
Foreign-born	3236 (63.85)	2126 (65.7)	1110 (34.3)	<.001
Life in the United States (%), n (%)				
≤25	670 (13.22)	431 (64.3)	239 (35.7)	<.001
>25 to ≤50	1014 (20)	305 (59.7)	409 (40.3)	<.001
>50 to ≤75	979 (19.32)	653 (12.9)	326 (33.3)	<.001
>75 to <100	508 (10.0)	419 (82.5)	89 (17.5)	<.001
100	1761 (34.74)	1565 (88.9)	196 (14.9)	<.001
Census region, n (%)				
West	3288 (64.88)	2436 (74.1)	852 (25.9)	<.001
Midwest	432 (8.52)	349 (80.8)	83 (19.2)	<.001
Northeast	623 (12.29)	459 (73.7)	164 (26.3)	<.001
South	720 (14.21)	503 (69.9)	217 (30.1)	<.001
Month and year of survey completion, n (%)				
October 2020	390 (7.69)	335 (85.9)	55 (14.1)	<.001
November 2020	680 (13.42)	505 (74.3)	175 (25.7)	<.001
December 2020	1350 (26.64)	1027 (76.1)	323 (23.9)	<.001
January 2021	2317 (45.72)	1684 (72.7)	633 (27.3)	<.001
February 2021	331 (6.53)	198 (59.8)	133 (40.2)	<.001
COVID-19-related experiences and impacts				
COVID-19 positivity, n (%)				
Yes	172 (3.39)	131 (76.2)	41 (23.8)	<.001
No	4542 (89.62)	3398 (74.8)	1144 (25.2)	<.001
Unsure	286 (5.64)	181 (63.3)	106 (36.7)	<.001
The severity of COVID-19 where they live, n (%)				
A lot less	447 (8.82)	303 (67.8)	144 (32.2)	<.001

Variables	All	Having racial bias belief		P value
		Yes[a] (n=3749)	No (n=1319)	
Somewhat less	869 (17.15)	659 (75.8)	210 (24.2)	<.001
About the same	1114 (21.98)	785 (70.5)	329 (29.5)	<.001
Somewhat more	1494 (29.47)	1114 (74.6)	380 (25.4)	<.001
A lot more	1121 (22.12)	872 (77.8)	249 (22.2)	<.001
Length of SIP[f] order				
No order	355 (7.0)	178 (50.1)	177 (49.9)	<.001
<1 month	272 (5.36)	175 (64.3)	97 (35.7)	<.001
1 to <2 months	574 (11.33)	458 (79.8)	116 (20.2)	<.001
2 to <3 months	579 (11.42)	482 (83.3)	97 (16.7)	<.001
≥3 months	2826 (55.76)	2140 (75.7)	686 (24.3)	<.001
Do not know	442 (8.72)	304 (68.8)	138 (31.2)	<.001
COVID-19 effect on family income or employment, n (%)				
No change	2086 (41.16)	1438 (68.9)	648 (31.1)	<.001
Mild change	1542 (30.43)	1198 (77.7)	344 (26.3)	<.001
Moderate change	1254 (24.74)	975 (77.8)	279 (21.3)	<.001
Severe change	166 (3.28)	128 (77.1)	38 (22.9)	<.001
COVID-19 effect on social support, n (%)				
No change	1046 (20.64)	623 (59.6)	423 (40.4)	<.001
Mild change	2201 (43.43)	1623 (73.7)	578 (26.3)	<.001
Moderate change	1526 (30.11)	1263 (82.8)	263 (17.2)	<.001
Severe change	249 (4.91)	213 (85.5)	36 (14.5)	<.001
COVID-19 effect on medical health care access, n (%)				
No change	1876 (37.01)	1288 (68.7)	588 (31.3)	<.001
Mild change	1574 (31.05)	1208 (76.8)	366 (23.2)	<.001
Moderate change	1070 (21.11)	855 (79.9)	215 (20.1)	<.001
Severe change	88 (1.73)	74 (84.1)	14 (15.9)	<.001
Not applicable	435 (8.58)	307 (70.6)	128 (29.4)	<.001
COVID-19 effect on mental health treatment access, n (%)				
No change	2271 (44.81)	1576 (69.4)	695 (30.6)	<.001
Mild change	554 (10.93)	459 (82.9)	95 (17.1)	<.001
Moderate change	240 (4.74)	179 (74.6)	61 (25.4)	<.001
Severe change	85 (1.67)	65 (86.7)	10 (13.3)	<.001
Not applicable	1882 (37.13)	1443 (76.7)	439 (23.3)	<.001

[a] For descriptive purposes only, we dichotomized responses to the CRBS items and categorized participants who responded *strongly agree* or *agree* to any of the 9 items as having COVID-19–related racial bias beliefs and those who did not as not having the perceptions.

[b] CRBS: Coronavirus Racial Bias Scale.

[c] N/A: not applicable.

[d] Ethnic Chinese includes mainland Chinese, Hongkonger, Taiwanese, and Huaren individuals.

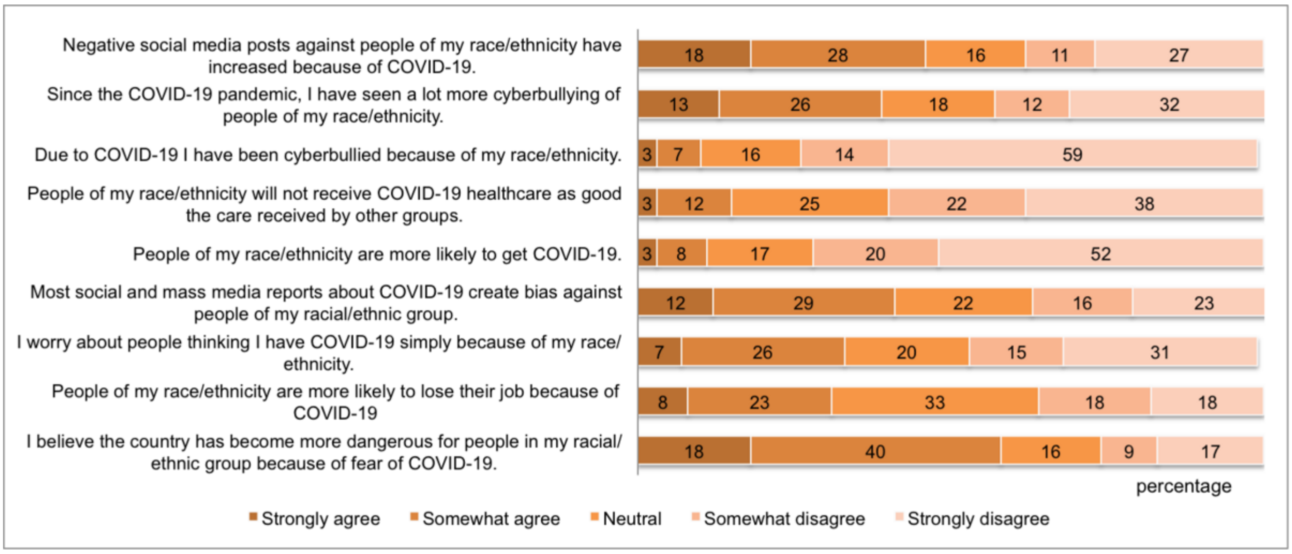
[e] LEP: limited English proficiency. Categorized as limited (yes) if speaking, reading or writing English were indicated as *some*, *a little*, or *not at all*.

[f] SIP: shelter-in-place.

COVID-19 Racial Bias Beliefs

Figure 1 shows the proportions of agreement or disagreement by CRBS item, which ranged from 10% of those who responded *agree* or *strongly agree* to the item "Due to COVID-19, I have been cyberbullied because of my race/ethnicity" to 58% of those who said they *agree* or *strongly agree* to the item "I believe the country has become more dangerous for my ethnic group."

Figure 1. Participants' beliefs of racial bias against Asian American, Native Hawaiian, and Pacific Islander individuals because of the COVID-19 pandemic, by proportion.



Correlates of Racial Bias Beliefs Against Asian Americans, Native Hawaiians, and Pacific Islanders

Multimedia Appendix 1 shows the unadjusted and adjusted associations between each correlate and the mean CRBS score. In the fully adjusted models, the associations remained significant for all variables in the model, except for some sociodemographic variables (marital status, employment status, and education) and perceived severity of COVID-19.

Many sociodemographic characteristics were independent correlates of the CRBS scores. In the fully adjusted model, compared with Asian Indians, ethnic Chinese (β=.79, 95% CI 0.69-0.89), Filipinos (β=.63, 95% CI 0.48-0.78), Hmong (β=.86, 95% CI 0.66-1.06), Japanese (β=.66, 95% CI 0.51-0.81), Korean (β=.15, 95% CI 0.04-0.26), Vietnamese (β=.37, 95% CI 0.26-0.49), and participants who identified as *other/more than one cultural group* (β=.71, 95% CI 0.57-0.85) had significantly higher mean CRBS scores. The mean difference in the CRBS score was not significantly different for Native Hawaiian and Pacific Islander individuals compared with Asian Indians. Compared with participants who were aged ≥60 years, those who were aged <30 years (β=.34, 95% CI 0.2-0.45), 30 to 39 years (β=.23, 95% CI 0.14-0.32), 40 to 49 years (β=.17, 95% CI 0.09-0.26), and 50 to 59 years (β=.10, 95% CI 0.02-018) had significantly higher mean CRBS scores. Women had higher mean CRBS scores than men (β=.05, 95% CI 0.01-0.10), and nonheterosexual individuals had higher mean CRBS scores than heterosexual individuals (β=.23, 95% CI 0.11-0.34). US-born participants had significantly higher mean CRBS scores than those who were foreign born (β=.19, 95% CI 0.13-0.25; β=.19, 95% CI 0.13-0.25), and those with LEP had higher mean CRBS scores than those with no LEP (β=.23, 95% CI 0.16-0.30). There were significant differences by region, with participants from the Midwestern (β=.21, 95% CI 0.12-0.31), Northeastern (β=.25, 95% CI 0.18-0.33), and Southern (β=.11, 95% CI 0.04-0.18) regions having significantly higher mean CRBS scores than the Western region. Compared with those who completed the survey in October 2020, those who completed the survey after that had significantly lower CRBS scores.

Most of the COVID-19–specific experiences were associated with the mean CRBS score, except for COVID-19 positivity (not significant in the unadjusted model and therefore not included in the full model) and perceived severity of COVID-19 where participants resided (became nonsignificant in the adjusted model). In the fully adjusted model, compared with participants without an SIP order, individuals who lived in places with <1 month (β=.09, 95% CI –0.04 to 0.22), 1 to <2 months (β=.24, 95% CI 0.13-0.35), 2 to <3 months (β=.27, 95% CI 0.16-0.38), and >3 months (β=.23, 95% CI 0.14-0.33) of SIP order had significantly higher mean CRBS scores. Higher mean CRBS scores were also observed for those who reported mild to severe changes in their family income, employment, medical health care access, mental health treatment access, and social support related to COVID-19 than for those who reported no change.

Discussion

Principal Findings

COMPASS was among the first nationwide surveys conducted in English and multiple other languages with >5000 Asian American and Native Hawaiian and Pacific Islander individuals and examined racial bias beliefs specifically related to the COVID-19 pandemic. Our findings revealed that 73.97% (3749/5068) of the survey respondents perceived racial bias within the past 6 months because of being *Asian American, Native Hawaiian, and Pacific Islander* during the COVID-19 pandemic. The unique contribution of this study is the documentation of disaggregated data of Asian groups in racial bias experiences reported during a 5-month period from October 2020 to February 2021, approximately 7-11 months after the start of the COVID-19 pandemic in the United States. Importantly, this study also identified multilevel factors associated with racial bias beliefs. These findings allow a comprehensive understanding of the sociodemographic contexts of being Asian American or Native Hawaiian and Pacific Islander and relevant COVID-19–related experiences that may have made Asian American and Native Hawaiian and Pacific Islander individuals vulnerable to perceived negative social attitudes, as exacerbated by the COVID-19 pandemic.

During the COVID-19 pandemic, our nationwide sample of Asian American and Native Hawaiian and Pacific Islander respondents shared that racial bias occurred in various forms. Among the most shared beliefs were perceiving that the United States had become more dangerous for individuals who identified with their ethnic group and observing negative social media posts against people of one's own ethnic group, which was reported by more than half of the respondents. Other racial bias experiences spanned across employment, perceived stigma of having COVID-19, exposure to social and mass media platforms, and observed or directly experienced cyberbullying. Consistent with the noticeable increase of anti-Asian hate incidents [6] and anxiety associated with being *Asian American*, particularly among individuals of Chinese or East Asian descent [3], these findings highlight the disproportionate burdens experienced by Asian American individuals during the pandemic, and such beliefs were shared across social contexts, physical environment, and cyberspace. When racial bias is perceived across multiple everyday contexts, such cumulative stressors and induced fears may have long-term psychological and physiological consequences [12,28,29], which warrants further studies on the long-term health and mental health consequences of COVID-19–related racial bias.

Our results showed that racial bias beliefs were particularly pronounced among some Asian American and Native Hawaiian and Pacific Islander groups and among those who were younger, a sexual minority, had higher income, were United States born, and were English proficient. Racial bias was perceived by many Asian American groups during the pandemic, ranging from 44.3% (132/298) among Asian Indian individuals to 94.5% (104/110) among Hmong. High levels of racial bias beliefs were also reported by ethnic Chinese, Filipino, and Japanese (>80%) individuals and those of multiple Asian descents (>90%). Although Asian Indians and Native Hawaiian and Pacific Islander individuals reported less racial bias beliefs compared with all other groups, perceiving racial bias was still reported by a large portion of individuals from these cultural groups: 44.3% (132/298) and 46.3% (63/136) for Asian Indian and Native Hawaiian and Pacific Islander individuals, respectively. These findings are consistent with the hate incidents reported by Asian American and Native Hawaiian and Pacific Islander individuals during the pandemic [3]. Although the observed cultural group differences persisted even after adjusting for other sociodemographic correlates, it should be noted that all Asian Indian participants and most Native Hawaiian and Pacific Islander participants completed the survey in English. Thus, these findings may not be generalizable to Asian Indian and Native Hawaiian and Pacific Islander participants with LEP. Nonetheless, a recent survey of Asian American, Native Hawaiian, and Pacific Islander women revealed similar levels of racism or discrimination incidents or harassment experienced across East Asian, Southeast Asian, South Asian, and Native Hawaiian and Pacific Islander individuals at a similarly high level (72% among East Asian individuals, to 80% among Native Hawaiian and Pacific Islander individuals) during the pandemic [8]. These findings underscore the urgency of addressing racial bias beliefs, discrimination, or related experiences among Asian American, Native Hawaiian, and Pacific Islander, particularly exacerbated during the pandemic.

Among those traditionally marginalized, Asian American or Native Hawaiian and Pacific Islander individuals who were nonheterosexual overwhelmingly (214/228, 93.8%) reported having racial bias beliefs during the pandemic. Those who had higher incomes, who were English proficient, and those who were United States born reported more racial bias beliefs. The reasons for this were unclear from the study data; a speculation could be related to increased access to and awareness of racial bias reports via media platforms in English [2,30-32]. Other studies have similarly shown that younger Asian American individuals and those with higher education perceived greater racism during the pandemic than older Asian American individuals and those with less education, respectively [5]. Finally, Asian American and Native Hawaiian and Pacific Islander respondents from all regions experienced more racial bias than those in the West, which had the highest population density of Asian American and Native Hawaiian and Pacific Islander individuals. This is in contrast to other data indicating that Asian American individuals on the west coast have reported high rates of discrimination [33]. However, locations such as New York reported an 833% change in anti-Asian crimes from 2019 to 2020 [34].

Although COVID-19 positivity and perceived severity of COVID-19 where respondents live were not relevant factors in racial bias beliefs with other factors adjusted, experiences of COVID-19–specific impacts were associated with COVID-19–related racial bias beliefs. In particular, respondents who had SIP orders for ≥1 month reported higher racial bias beliefs. COVID-19–specific sequelae with negative impacts on family income or employment, family and social support access, medical health care, and mental health care access were found to have contributed to Asian American and Native Hawaiian and Pacific Islander individuals' experiences and perceptions of racial bias. Income, social support, and health care access are all social determinants of health [35]. These findings also suggest a pathway through which COVID-19 affects higher racial bias beliefs and, thereby, health consequences. In addition, these findings revealed that the experience of COVID-19–specific impacts across each area was uniquely associated with and additive in contributing to COVID-19–related racial bias beliefs. These observed associations reflected the differential impacts of COVID-19–specific sequelae that have significant policy implications. Our findings may also reflect a consequence of pre-existing systemic and structural racism that was further exacerbated differentially at multiple levels during the pandemic [36-38].

Although the findings suggested COVID-19–related racial bias beliefs appeared to reduce after October 2020 over the 5-month data collection period of COMPASS, the list of sociodemographic correlates and COVID-19–specific experiences remain significant contributors to racial bias beliefs with the time of data collection adjusted. The findings from this study further underscore the importance of addressing these needs in our respective communities in a timely manner, with attention to cultural contexts and language needs. Moreover, 58% of Asian American individuals said that from March 2020 to March 2021, reports about discrimination and violence against Asian American individuals affected their own mental health [39], although the sequelae have not been well studied, which is a significant concern, given that Asian American individuals underuse mental health care [40,41]. Moving forward, racially motivated hate requires interventions at the structural (eg, hate crime laws, public awareness campaigns, and public health surveillance), interpersonal (eg, hate crime–specific training such as a semester-long college course and intergroup contact programs such as contact and skill-based prejudice reduction programs), and individual levels (eg, self-reflection exercises) [42]. This multilevel intervention was proposed by Cramer et al [42], given the rise of racially motivated hate against several populations, including Asian American individuals, and is based on a socioecological summary of hate-motivated behavior impacts and causes, risk factors for commission, and potential solutions. Cramer et al [42] offered the proposed solutions found in the literature.

Strengths and Limitations

COMPASS is one of the few nationwide surveys focused on the impact of COVID-19 on Asian American and Native Hawaiian and Pacific Islander populations and conducted in multiple Asian languages, which included a large, diverse sample of participants from multiple cultural groups. Data disaggregation is crucial in studies with Asian American and Native Hawaiian and Pacific Islander individuals and yet is rarely performed [43]. The survey was made accessible by smartphones, tablets, or computers, as well as by phone (and in person, if available), with the assistance of a staff member to remove barriers to participation. The study survey had a major focus on inclusion and was available in 5 languages; however, individuals with LEP may have been excluded by not offering the survey in other languages that were not supported.

Owing to the cross-sectional nature of this analysis, our study was limited to reporting experiences at one point in time and did not allow us to assess changes over time during the pandemic. It is possible that the participants experienced racial bias after completing the survey, which would not be reflected in our results. There may also be variation in participants' interpretations of racial bias based on acculturation, which was not assessed as part of the survey. Future research should consider assessing acculturation and the social desirability bias scale [44] to account for potential biases. The responses may have been affected by varying SIP and social distancing policies across regions and states in this nationwide survey.

Conclusions

Our study revealed that most Asian American and Native Hawaiian and Pacific Islander individuals reported racial bias beliefs due to being Asian American and Native Hawaiian and Pacific Islander during the COVID-19 pandemic. In the context of the sharp rise in anti-Asian incidents and hate crimes reported during the pandemic [4,5,9], the findings of this study further underscore the devastating impact of COVID-19 among Asian American, Native Hawaiian, and Pacific Islander individuals. Relevant sociodemographic contexts and pre-existing and COVID-specific impacts were associated with perceived racial bias of *being Asian American, Native Hawaiian, and Pacific Islander* during the pandemic. Racial biases have significant impacts on one's lived experiences, spanning across various everyday contexts that contribute to disparities in access to or opportunities for education, work, and health care [11-13]. Identifying, creating, and allocating culturally and linguistically appropriate resources dedicated to addressing such burdens on Asian American, Native Hawaiian, and Pacific Islander communities because of racial bias among other COVID-19–related sequelae are priorities of urgency. Future interventions such as those proposed earlier require significant resources (eg, academic-community partnerships and state-wide task forces) and funding, as well as comprehensive evaluations of each antihate intervention [42], all of which are necessary to combat this serious public health issue of racially motivated hate.

Acknowledgments

This study was supported by the National Institute on Aging, National Institutes of Health (3R24AG063718-02S1, R24AG063718, and R56AG069130). The authors wish to express their sincere gratitude to the National Pacific Center on Aging, International Children Assistance Network, and other community partners and members of the Community Advisory Board for assistance with the outreach and administration of the COVID-19 Effects on the Mental and Physical Health of Asian Americans and Pacific Islanders Survey Study (COMPASS).

The funder of the study had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Associations of participant characteristics with mean Coronavirus Racial Bias Scale score: results from unadjusted and fully adjusted linear regression analyses.
[DOCX File , 32 KB - jmir_v24i8e38443_app1.docx]

References

1. Dhanani LY, Franz B. Why public health framing matters: an experimental study of the effects of COVID-19 framing on prejudice and xenophobia in the United States. Soc Sci Med 2021 Jan;269:113572. [doi: 10.1016/j.socscimed.2020.113572] [Medline: 33321405]
2. Hswen Y, Xu X, Hing A, Hawkins JB, Brownstein JS, Gee GC. Association of "#covid19" Versus "#chinesevirus" With Anti-Asian Sentiments on Twitter: March 9-23, 2020. Am J Public Health 2021 May;111(5):956-964. [doi: 10.2105/AJPH.2021.306154] [Medline: 33734838]
3. Tessler H, Choi M, Kao G. The anxiety of being Asian American: hate crimes and negative biases during the COVID-19 pandemic. Am J Crim Justice 2020;45(4):636-646 [FREE Full text] [doi: 10.1007/s12103-020-09541-5] [Medline: 32837158]
4. Ellerbeck A. Survey: more than 30 percent of Americans have witnessed COVID-19 bias against Asians. The Center for Public Integrity. URL: https://publicintegrity.org/health/coronavirus-and-inequality/survey-majority-of-asian-americans-have-witnessed-covid-19-bias/ [accessed 2020-05-01]
5. Many Black and Asian Americans say they have experienced discrimination amid the COVID-19 outbreak. Pew Research Center. URL: https://www.pewresearch.org/social-trends/2020/07/01/many-black-and-asian-americans-say-they-have-experienced-discrimination-amid-the-covid-19-outbreak/ [accessed 2021-04-27]
6. STOP AAPI HATE National Report. STOP AAPI HATE. URL: https://stopaapihate.org/national-report-through-september-2021/ [accessed 2022-01-28]
7. Ta Park VM, Dougan MM, Meyer OL, Nam B, Tzuang M, Park LG, et al. Discrimination experiences during COVID-19 among a national, multi-lingual, community-based sample of Asian Americans and pacific islanders: compass findings. Int J Environ Res Public Health 2022 Jan 14;19(2):924 [FREE Full text] [doi: 10.3390/ijerph19020924] [Medline: 35055744]
8. Pillai D, Lindsey A. The state of safety for Asian American and pacific islander women in the U.S. The Asian American Association. URL: https://www.napawf.org/assets/download/napawf-state-of-safety-report.pdf [accessed 2022-04-29]
9. Lee S, Waters SF. Asians and Asian Americans’ experiences of racial discrimination during the COVID-19 pandemic: impacts on health outcomes and the buffering role of social support. Stigma Health 2021 Feb;6(1):70-78. [doi: 10.1037/sah0000275]
10. Williams S. Bias, race. In: Encyclopedia of Child Behavior and Development. Boston, MA: Springer; 2011.
11. Glaser J, Spencer K, Charbonneau A. Racial bias and public policy. Policy Insights Behav Brain Sci 2014 Oct 01;1(1):88-94 [FREE Full text] [doi: 10.1177/2372732214550403]
12. Williams DR, Wyatt R. Racial bias in health care and health: challenges and opportunities. JAMA 2015 Aug 11;314(6):555-556. [doi: 10.1001/jama.2015.9260] [Medline: 26262792]
13. Williams RL, Romney C, Kano M, Wright R, Skipper B, Getrich CM, et al. Racial, gender, and socioeconomic status bias in senior medical student clinical decision-making: a national survey. J Gen Intern Med 2015 Jun 27;30(6):758-767 [FREE Full text] [doi: 10.1007/s11606-014-3168-3] [Medline: 25623298]
14. Braveman P, Egerter S, Williams DR. The social determinants of health: coming of age. Annu Rev Public Health 2011;32:381-398. [doi: 10.1146/annurev-publhealth-031210-101218] [Medline: 21091195]
15. D'Anna LH, Ponce NA, Siegel JM. Racial and ethnic health disparities: evidence of discrimination's effects across the SEP spectrum. Ethn Health 2010 Apr;15(2):121-143. [doi: 10.1080/13557850903490298] [Medline: 20131130]
16. Hahm HC, Ozonoff A, Gaumond J, Sue S. Perceived discrimination and health outcomes a gender comparison among Asian-Americans nationwide. Womens Health Issues 2010 Sep;20(5):350-358 [FREE Full text] [doi: 10.1016/j.whi.2010.05.002] [Medline: 20800771]
17. Zhang W, Hong S. Perceived discrimination and psychological distress among Asian Americans: does education matter? J Immigr Minor Health 2013 Oct 6;15(5):932-943 [FREE Full text] [doi: 10.1007/s10903-012-9676-5] [Medline: 22767300]
18. Language spoken at home by ability to speak English for the population 5 years and over. United States Census Bureau. URL: https://data.census.gov/cedsci/table?q=B16001&tid=ACSDT5Y2020.B16001 [accessed 2022-05-31]
19. Ta Park VM, Meyer OL, Tsoh JY, Kanaya AM, Tzuang M, Nam B, et al. The Collaborative Approach for Asian Americans and Pacific Islanders Research and Education (CARE): a recruitment registry for Alzheimer's disease and related dementias, aging, and caregiver-related research. Alzheimers Dement 2022 Apr 14 (forthcoming). [doi: 10.1002/alz.12667] [Medline: 35420258]
20. Process of translation and adaptation of instruments. World Health Organization. URL: http://www.who.int/substance_abuse/research_tools/translation/en/ [accessed 2018-09-11]
21. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

22. 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]
23. Fisher C, Tao X, Yip T. The effects of COVID-19 victimization distress and racial bias on mental health among AIAN, Asian, Black, and Latinx young adults. Cultur Divers Ethnic Minor Psychol 2022 Apr 07 (forthcoming) [FREE Full text] [doi: 10.1037/cdp0000539] [Medline: 35389692]
24. PhenX Toolkit homepage. PhenX Toolkit. URL: https://www.phenxtoolkit.org/index.php [accessed 2022-04-01]
25. Census Regions and Divisions of the United States. www2censusgov. URL: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf [accessed 2021-06-09]
26. Cawthon P, Orwoll E, Ensrud K, Cauley JA, Kritchevsky SB, Cummings SR, et al. Assessing the impact of the COVID-19 pandemic and accompanying mitigation efforts on older adults. J Gerontol A Biol Sci Med Sci 2020 Sep 16;75(9):e123-e125 [FREE Full text] [doi: 10.1093/gerona/glaa099] [Medline: 32307522]
27. Coronavirus impact scale. NIH U.S. National Library of Medicines. URL: https://www.nlm.nih.gov/dr2/Coronavirus_Impact_Scale.pdf [accessed 2022-07-06]
28. Williams DR, Neighbors HW, Jackson JS. Racial/ethnic discrimination and health: findings from community studies. Am J Public Health 2003 Feb;93(2):200-208. [doi: 10.2105/ajph.93.2.200] [Medline: 12554570]
29. Heard-Garris N, Cale M, Camaj L, Hamati M, Dominguez T. Transmitting Trauma: a systematic review of vicarious racism and child health. Soc Sci Med 2018 Feb;199:230-240. [doi: 10.1016/j.socscimed.2017.04.018] [Medline: 28456418]
30. Cho H, Li W, Cannon J, Lopez R, Song C. Testing three explanations for stigmatization of people of Asian descent during COVID-19: maladaptive coping, biased media use, or racial prejudice? Ethn Health 2021 Jan 15;26(1):94-109. [doi: 10.1080/13557858.2020.1830035] [Medline: 33059486]
31. Yang C, Tsai J, Pan S. Discrimination and well-being among Asians/Asian Americans during COVID-19: the role of social media. Cyberpsychol Behav Soc Netw 2020 Dec 01;23(12):865-870. [doi: 10.1089/cyber.2020.0394] [Medline: 32762541]
32. Yu N, Pan S, Yang C, Tsai J. Exploring the role of media sources on COVID-19-related discrimination experiences and concerns among Asian people in the United States: cross-sectional survey study. J Med Internet Res 2020 Nov 06;22(11):e21684 [FREE Full text] [doi: 10.2196/21684] [Medline: 33108307]
33. Fact Sheet: anti-Asian prejudice March 2021. Center for the Study of Hate & Extremism. URL: https://www.csusb.edu/sites/default/files/FACT%20SHEET-%20Anti-Asian%20Hate%202020%20rev%203.21.21.pdf [accessed 2022-07-06]
34. Farivar M. Hate crimes targeting Asian Americans spiked by 150% in major US cities. VOA. URL: https://www.voanews.com/a/usa_race-america_hate-crimes-targeting-asian-americans-spiked-150-major-us-cities/6202736.html [accessed 2022-03-31]
35. Palmer RC, Ismond D, Rodriquez EJ, Kaufman JS. Social determinants of health: future directions for health disparities research. Am J Public Health 2019 Jan;109(S1):S70-S71. [doi: 10.2105/AJPH.2019.304964] [Medline: 30699027]
36. Muramatsu N, Chin MH. Battling structural racism against Asians in the United States: call for public health to make the "Invisible" visible. J Public Health Manag Pract 2022;28(Suppl 1):S3-S8. [doi: 10.1097/PHH.0000000000001411] [Medline: 34797254]
37. Parolin Z, Lee EK. The role of poverty and racial discrimination in exacerbating the health consequences of COVID-19. Lancet Reg Health Am 2022 Mar;7:100178 [FREE Full text] [doi: 10.1016/j.lana.2021.100178] [Medline: 35018358]
38. Saw A, Yi SS, Đoàn LN, Tsoh JY, Yellow Horse AJ, Kwon SC, et al. Improving Asian American health during the Syndemic of COVID-19 and racism. EClinicalMedicine 2022 Mar;45:101313 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101313] [Medline: 35233516]
39. Findling M, Blendon RJ, Benson J, Koh H. COVID-19 has driven racism and violence against Asian Americans: perspectives from 12 national polls. Health Affairs. URL: https://www.healthaffairs.org/do/10.1377/forefront.20220411.655787/ [accessed 2022-05-16]
40. Cook BL, Trinh N, Li Z, Hou SS, Progovac AM. Trends in racial-ethnic disparities in access to mental health care, 2004-2012. Psychiatr Serv 2017 Jan 01;68(1):9-16 [FREE Full text] [doi: 10.1176/appi.ps.201500453] [Medline: 27476805]
41. Yang KG, Rodgers CR, Lee E, Lê Cook B. Disparities in mental health care utilization and perceived need among Asian Americans: 2012-2016. Psychiatr Serv 2020 Jan 01;71(1):21-27. [doi: 10.1176/appi.ps.201900126] [Medline: 31575351]
42. Cramer RJ, Fording RC, Gerstenfeld P, Kehn A, Marsden J, Deitle C. Hate-motivated behavior: impacts, risk factors, and interventions. Health Affairs. URL: https://www.healthaffairs.org/do/10.1377/hpb20200929.601434/full/ [accessed 2022-05-16]
43. Byon A. Everyone deserves to be seen. SEARAC. URL: https://www.searac.org/wp-content/uploads/2020/05/ihep_aapi_brief.pdf [accessed 2022-03-31]
44. Gee GC, Spencer MS, Chen J, Takeuchi D. A nationwide study of discrimination and chronic health conditions among Asian Americans. Am J Public Health 2007 Jul;97(7):1275-1282. [doi: 10.2105/AJPH.2006.091827] [Medline: 17538055]

Abbreviations

CARE: Collaborative Approach for Asian Americans and Pacific Islanders Research and Education

COMPASS: COVID-19 Effects on the Mental and Physical Health of Asian Americans and Pacific Islanders Survey Study
CRBS: Coronavirus Racial Bias Scale
LEP: limited English proficiency
REDCap: Research Electronic Data Capture
SIP: shelter-in-place

Edited by C Basch; submitted 05.04.22; peer-reviewed by A Bhagavathula, A Mukherjea; comments to author 25.04.22; revised version received 16.05.22; accepted 01.06.22; published 09.08.22.

Please cite as:
Park VT, Tsoh JY, Dougan M, Nam B, Tzuang M, Park LG, Vuong QN, Bang J, Meyer OL
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
URL: https://www.jmir.org/2022/8/e38443
doi:10.2196/38443
PMID:35658091

Original Paper

Physicians' Perceptions as Predictors of the Future Use of the National Death Information System in Peru: Cross-sectional Study

Javier Vargas-Herrera[1*], PhD; Giovanni Meneses[1*], PhD; Juan Cortez-Escalante[2*], PhD

[1]Department of Preventive Medicine and Public Health, National University of San Marcos, Lima, Peru

[2]Pan American Health Organization, Brasilia, Brazil

[*]all authors contributed equally

Corresponding Author:
Javier Vargas-Herrera, PhD
Department of Preventive Medicine and Public Health
National University of San Marcos
Av German Amezaga 375
Lima, 15081
Peru
Phone: 51 945029342
Email: jvargash@unmsm.edu.pe

Abstract

Background: A computer application called the National Death Information System (SINADEF) was implemented in Peru so that physicians can prepare death certificates in electronic format and the information is available online. In 2018, only half of the estimated deaths in Peru were certified using SINADEF. When a death is certified in paper format, the probability being entered in the mortality database decreases. It is important to know, from the user's perspective, the factors that can influence the successful implementation of SINADEF. SINADEF can only be successfully implemented if it is known whether physicians believe that it is useful and easy to operate.

Objective: The aim of this study was to identify the perceptions of physicians and other factors as predictors of their behavioral intention to use SINADEF to certify a death.

Methods: This study had an observational, cross-sectional design. A survey was provided to physicians working in Peru, who used SINADEF to certify a death for a period of 12 months, starting in November 2019. A questionnaire was adapted based on the Technology Acceptance Model. The questions measured the dimensions of subjective norm, image, job relevance, output quality, demonstrability of results, perceived usefulness, perceived ease of use, and behavioral intention to use. Chi-square and logistic regression tests were used in the analysis, and a confidence level of 95% was chosen to support a significant association.

Results: In this study, 272 physicians responded to the survey; 184 (67.6%) were men and the average age was 45.3 (SD 10.1) years. The age range was 24 to 73 years. In the bivariate analysis, the intention to use SINADEF was found to be associated with (1) perceived usefulness, expressed as "using SINADEF avoids falsifying a death certificate" (P<.001), "using SINADEF reduces the risk of errors" (P<.001), and "using SINADEF allows for filling out a certificate in less time" (P<.001); and (2) perceived ease of use, expressed as "I think SINADEF is easy to use" (P<.001). In the logistic regression, perceived usefulness (odds ratio [OR] 8.5, 95% CI 2.2-32.3; P=.002), perceived ease of use (OR 10.1, 95% CI 2.4-41.8; P=.001), and training in filling out death certificates (OR 8.3, 95% CI 1.6-42.8; P=.01) were found to be predictors of the behavioral intention to use SINADEF.

Conclusions: The behavioral intention to use SINADEF was related to the perception that it is an easy-to-use system, the belief that it improves the performance of physicians in carrying out the task at hand, and with training in filling out death certificates.

(J Med Internet Res 2022;24(8):e34858) doi:10.2196/34858

KEYWORDS

death certificates; health information system; mortality; vital statistics; Technology Acceptance Model; model; acceptance model; certificates; information system; physicians; predictors; cross-sectional study; analysis; death

Introduction

Medical death certification is the main source of information on causes of death in a population [1]. Various studies worldwide report that there is low coverage of deaths that have medical certification of death [2,3], and those deaths that have certification of the causes of death do not have the desired quality [4,5]. In 2016, the coverage of deaths with medical certification of the causes of death in Peru was 56% [6] and the quality, measured as the proportion of codes of causes of death in the International Classification of Diseases classified as "garbage code," was among the highest in the world [7]. In 2017, a computer application called the National Death Information System (SINADEF) [8] was implemented in Peru, which is used by physicians to prepare death certificates online. This system was the result of coordinated work between the National Institute of Statistics and Informatics, the National Registry of Identification and Civil Status, and the Ministry of Health, supported by the Bloomberg "Data for Health" initiative. Before implementing SINADEF, physicians performed death certification on paper forms, which were entered into computer applications installed on computers with local databases, which were then sent through email messages at the regional and national levels. In addition to the lack of opportunity for data availability, this system has many other problems: it does not verify the identity of the deceased by consulting a database, it does not identify the hospital where the person died, it does not unequivocally locate the district where the death occurred, and often, the cause of death written by the doctor is illegible.

With SINADEF, physicians can certify a death online, verify the identity of the deceased and the hospital where the death occurred, consult online databases, and more accurately record the place of death. Further, the problem of the lack of legibility of the doctor's handwriting, which is very important in the registry of causes of death, disappears. From the perspective of electronic government, information and communication technology (ICT) can facilitate the services provided to citizens. It allows health authorities at all levels of government to monitor mortality indicators in a timely manner and improve prevention and control actions, as is happening now with the monitoring of the COVID-19 pandemic. Likewise, it allows other state agencies to be more efficient in public policy management processes. These aspects justify taking the greatest precautions to guarantee its successful implementation.

A paper [6] that reported the first results of the implementation of SINADEF showed that in July 2018, only half of the estimated deaths were certified using that system, and it identified some gaps and barriers that could limit the achievement of adequate coverage and quality of the registry of causes of death. One of them is the modality used by the doctor to certify a death [6]. When a doctor certifies a death on paper forms, the data may not enter the system. The paper forms can remain with the relative of the deceased or in a drawer of a hospital office, in funeral agencies, in the civil registry, etc. In July 2018, half of the estimated deaths were not entered in SINADEF. Some studies about the map of death registration processes carried out in Peru indicate that deaths registered in paper format are less likely to enter the mortality database [9], which highlights the importance of identifying factors linked with physicians, thereby improving the use of SINADEF to certify a death. Furthermore, SINADEF has played a fundamental role in documenting the excess mortality that occurred during the COVID-19 pandemic in Peru. Several studies [10-12] used SINADEF to warn, from the beginning of the first wave of transmission of COVID-19 in Peru, a significant difference in deaths that the physicians who used SINADEF attributed to COVID-19 and that were not reported by the system of epidemiological surveillance. In addition, they coincide in pointing out the importance of the quality of the data provided by SINADEF and the need to strengthen the entire death registration process so that it is useful for containing the health emergency. Recent studies using SINADEF reported that excess mortality from all causes in 2020 was more than 100,000 deaths and that more realistic figures have been obtained owing to coordination initiatives between those responsible for epidemiological surveillance, recording of vital events, and diagnosis [13]. However, they did not notice that a significant proportion of deaths, especially those that occur in the community, are not reported through SINADEF; therefore, it is important to evaluate the factors that could limit its real use.

Intuitively, it may seem that the information production process supported by paper format, when replaced by an electronic registry, as is the case with SINADEF, will immediately produce obvious improvements such as the reduction of errors due to data transcription. However, there are also bad experiences in the implementation of ICT projects in health services, which in theory were very robust, probably because the user experience was not considered in the design of the applications [14]. Therefore, it is necessary to know the determinants of the future use of a technology.

All innovation in health requires the intensive participation of the people who will be affected by the processes in which it will intervene. To expand the use of online death certification, it is important to know what the physicians think about the system. Some barriers related to the system itself are dependence on internet connectivity, limited availability of computers and printers in hospitals, and failures in technical support and application maintenance. Although the advantages include their use even from mobile devices, physicians may not be willing to use them. To understand why users accept or use new technologies, a predictive evaluation methodology of technology use known as the Technology Acceptance Model (TAM) has been proposed, which has been tested and validated in different contexts and studies [15,16]. The TAM has been used in the evaluation and implementation of ICT in the health field [17-20]. The TAM is based on user perceptions. If people believe that an ICT application helps them to do their job better ("perceived usefulness") and, at the same time, that it does not require additional effort ("perceived easy to use"), they will end up adopting that technology [15].

Rahimi et al [21] conducted a systematic review of 134 studies published between 1999 and 2017 that used TAM to evaluate ICT applications in the health field after their initial introduction. The areas of ICT studied were general information technologies, health information systems, electronic health records or electronic medical records, electronic prescription systems (e-prescription), pocket computers, telemedicine, mobile health, and personal health records. An online system for medical certification of death can be considered an electronic medical record component, but in this review, no study specifically evaluated an information system like SINADEF. The studies evaluated are widely distributed throughout the world, but they are scarce in Latin America, and none were reported in Peru. The main findings were that the application of TAM was quite heterogeneous. Most studies used extensions of the original TAM, which suggests that there is no optimal version of TAM to use in the field of health. Holden and Karsh [16] also carried out a review of TAM and highlight the importance of developing "the left part of the model," that is, the determinants with specific, contextualized, and actionable constructs.

In Peru, we have not identified previous research that describes the use of TAM to evaluate ICT in health; only 1 study evaluated the intended use of mobile banking services [22]. Our study aims to identify the perceptions of physicians about the usefulness and ease of use of SINADEF and other factors such as predictors of the behavioral intention to use SINADEF, contextualizing the predictor variables with specific propositions about the use of SINADEF to certify deaths.

Methods

Design

An observational, cross-sectional study was carried out that consisted of the application of a survey directed to physicians who used the national computer system of deaths (SINADEF) to produce a death certificate. The study population was made up of physicians with a professional practice in Peru and who, as of 2017, had the chance to carry out, at least once, the death certification process online through SINADEF.

Recruitment

Between November 24, 2019, and November 18, 2020, an electronic form was sent to 26,185 physicians who were registered as SINADEF users until December 31, 2017. Of them, 505 physicians opened the electronic form and 424 agreed to participate in the study; 111 physicians were excluded because they reported that they had no experience in the use of SINADEF, and 41 physicians were excluded because they had contradictory answers about their degree of agreement with the use of SINADEF to certify a death. Therefore, the sample consisted of 272 physicians.

TAM

Questionnaire

The questions in the questionnaire were based on TAM. TAM proposes that the behaviors of individuals depend on their beliefs and subjective norms and that the use of a new technology depends on 2 variables in particular: the perceived ease of use of the technology and the perceived usefulness. In turn, these 2 variables will directly influence the attitude of the individual toward the actual use of technology [15]. In a simpler way, if an individual has the belief that a technology is easy to use and at the same time, it is useful for the work that it has to fulfill, he or she will have a favorable attitude to use the technology effectively. In this study, the TAM2 model proposed by Venkatesh and Davis [23] was used, which incorporates the following variables: subjective norm, image, job relevance, output quality, and results demonstrability. The measurement scale used was based on the original proposals of Venkatesh and Davis [23] for the evaluation of information technology in general but adapted to the context of its use in the medical certification of deaths in Peru, following the recommendations of Holden and Karsh [16] on adapting the model specifically to the health context by using belief induction methods. In the propositions of beliefs, reference is made to the regulatory provisions of the Ministry of Health in Peru or the comparison of the use of SINADEF with the traditional way of using paper forms. The following operational definitions were used for each of the explanatory variables proposed in the model used (Textbox 1).

In addition, variables external to the model are included that can be explanatory or that intervene in the attitude or behavior of intention to use the technology of physicians, such as age, sex, medical specialty, workplace, time spent in service, and training in the filling of death certificates.

Textbox 1. Operational definitions of the Technology Acceptance Model used.

Subjective norm: Subjective perception of the individual on social pressures, which includes the perception of the beliefs of relevant people, for the adoption of the behavior of use of the National Death Information System (SINADEF), through their opinion of agreement or disagreement on a Likert scale of the following propositions:

- "I have to use SINADEF to certify deaths because it's already established that way."
- "The authorities of my hospital and the Ministry of Health could sanction me if I do not use SINADEF to certify deaths."
- "I have to use SINADEF because everyone already uses it to certify deaths."

Image: Refers to the user's self-perception, visual, or mental representation produced by the use of SINADEF. It was measured by their opinion of agreement or disagreement on a Likert scale of the following propositions:

- "I feel comfortable with information and communication technologies."
- "I consider myself a person open to change."
- "I have good adaptability."

Job relevance: Refers to the perceived attribution of the user that SINADEF is important for the performance of work tasks. It was measured by their opinion of agreement or disagreement on a Likert scale of the following propositions:

- "When SINADEF is implemented, the causes of death of the population will be known in a timelier manner."
- "The use of SINADEF will help the directors or managers of health organizations make decisions."
- "If SINADEF is consolidated, it will be possible to quickly consult the data of the deceased."

Output quality: Refers to the real or perceived attribution of the SINADEF user of the quality of the report provided by the technology. It was measured by their opinion of agreement or disagreement on a Likert scale of the following proposition:

- "The SINADEF death certificate is of higher quality than the paper format."

Perceived usefulness: Refers to an individual's perception that the use of SINADEF will improve job performance. It was measured by their opinion of agreement or disagreement on a Likert scale of the following propositions:

- "Using SINADEF avoids falsifying a death certificate."
- "Using SINADEF reduces the risk of errors in death certificates."
- "Using SINADEF allows me to produce a death certificate in less time than doing it in paper format."

Perceived ease of use: Refers to an individual's perception that using SINADEF does not require effort, through their opinion of agreement or disagreement on a Likert scale of the following propositions:

- "I think SINADEF is easy to use."
- "It's easy to get a password to use SINADEF."
- "It's easy to recover the SINADEF password when you forget it or it's blocked."
- "It's easy to get technical support from SINADEF when you need it."

Behavioral intention to use: Refers to the motivation or willingness of an individual to make the effort to use SINADEF to certify a death, through their opinion of agreement or disagreement on a Likert scale of the following proposition:

- "When I need to certify a death, I will use the SINADEF."
- "In any circumstance, if I need to certify a death, I will not use SINADEF."

Results demonstrability: Thinking that SINADEF allows you to demonstrate the results, through their opinion of agreement or disagreement on a Likert scale of the following proposition:

- "I can easily communicate the results of my experience using SINADEF."
- "For me, the results of using SINADEF will be visible."

Instrument Validation

The content of the proposed instrument was validated through an expert judgment process [24]. Six expert researchers were selected—2 of them experts in measurement and evaluation. The index of agreement of the judges on the propositions of the TAM variables was 84.9%, with a κ index of 0.7304. All elements with a κ index less than 0.6 were removed and replaced. Items with a κ index of 0.6571 were reviewed and their formulation was paraphrased or modified based on the judges' recommendations. The reliability of the 20-item measurement scale for the 8 variables that were included in the study was evaluated with the Cronbach α coefficient. The value obtained for the Cronbach α coefficient was .874.

Statistical Analysis

Overview

The data collected through the electronic forms were analyzed using the SPSS statistics package (IBM Corp). A description of the variables was made, presenting arithmetic means and standard deviations of the quantitative variables, such as age and years of medical work, and frequency tables of qualitative variables such as sex, workplace, and medical specialty. A bivariate analysis was performed between the independent variables such as perceived usefulness or perceived ease of use and the dependent variable behavioral intention to use. Subsequently, a multivariate analysis was performed using binary logistic regression with the Wald successive steps method with a likelihood ratio, evaluating models that included both descriptive and explanatory variables as independent variables, with the dependent variable behavioral intention to use. To carry out the bivariate analysis and binary logistic regression analysis, the responses to the propositions of each of the variables on a Likert scale were transformed into 2 alternatives: “agree” and “disagree.” In the initial model, all the variables of TAM that were found associated with the behavioral intention to use in the bivariate analysis were considered as independent or predictive variables. Additionally, the variables sex, age, main work center, and training in filling out death certificates were entered into the model. Following the TAM2 model, the variables subjective norm, image, job relevance, output quality, demonstrability of results, and perceived ease of use were crossed with perceived usefulness. In the successive steps, the variables with a Wald value less than 1 and with statistical significance less than .05 were eliminated.

Power

In the bivariate analysis, the chi-square test was used to establish the existence of an association between the variables, choosing a confidence level of 95%, a level of statistical significance of (α) equal to or less than .05, and a power of 80% (1-β) equal to or less than .20. To identify the variables associated with behavioral intention to use in the multivariate analysis, the logistic regression with Wald method (backward) of successive steps with the likelihood ratio was used. A confidence level of 95% and a significance level (α) equal to or less than .05 were chosen.

Ethics Approval

The research protocol and the informed consent signed by the participants were approved by the Research Ethics Committee of the San Marcos University School of Medicine (#19-0027).

Results

Sample

In this study, 272 physicians, who were SINADEF users since 2017, were studied. Most (184/272, 67.6%) were male. The average age was 45.3 (SD 10.1) years. The minimum age was 24 years and the maximum age was 73 years. Most of the respondents (100/272, 36.8%) were in the age group of 40 to 49 years. Only among young physicians between the ages of 24 and 30 years did women predominate (10/17, 58.8%), while in the rest of the age groups, there was a predominance of male physicians.

Description of the Population Studied

The most frequent participants in this study were general practitioners (70/272, 25.7%), pediatricians or neonatologists (30/272, 11%), internists (25/272, 9.2%), and obstetrician-gynecologists (15/272, 5.5%). Forensic physicians constituted 4.4% (12/272) of the participants. The study included graduates of all universities that have a faculty of medicine in Peru. Approximately 5.1% (14/272) of the participants studied abroad. Most of the study participants worked in hospitals as their main workplace, followed by those who worked in private clinics and health centers. Most of the participants (221/272, 81.3%) had worked for 10 or more years. Approximately 71% (193/272) stated that they had received some type of training in filling death certificates and among these, the majority did so through conferences given at the same hospital, through health services networks, or through a combination of training modalities (see Table 1).

Figure 1 shows the perceptions classified by their level of agreement among the respondents, and those with the highest agreement were perceived as people open to change, who considered the usefulness of the system to quickly consult information about the deceased, and the ability to adapt to innovative methods. Those who had less agreement corresponded to the ease of obtaining passwords or technical support and the possible sanctions that their nonuse would cause.

Table 1. Description of the main characteristics of the studied population (N=272).

Characteristics	Values, n (%)
Sex	
Male	184 (67.6)
Female	88 (32.4)
Age group (years)	
24-30	17 (6.3)
30-39	60 (22.1)
40-49	100 (36.8)
50-59	70 (25.7)
>59	25 (9.2)
Main workplace	
Hospital	190 (69.9)
Health center/post	32 (11.8)
Private clinic/physician's office	29 (1.7)
Medical-legal division	16 (5.9)
Other	5 (1.8)
Years of medical work	
Less than 10	51 (18.8)
10-19	99 (36.4)
20-29	83 (3.5)
More than 29	39 (14.3)
Training in filling death certificate	
Yes	193 (71)
No	79 (29)
Training modality (n=193)	
A conference at the health center or through a health services network	58 (3.1)
During undergraduate studies	28 (14.5)
During postgraduate studies	20 (1.4)
A virtual course	14 (7.3)
Combined (more than one of the above)	73 (37.8)

Figure 1. Level of agreement (%) in the physicians' perceptions about the use of the National Death Information System. MINSA: Ministry of Health; SINADEF: National Death Information System.

Bivariate Analysis

Physicians' perceptions that were found to be associated with behavioral intention to use ($P<.05$) were perceived usefulness as shown by "using SINADEF avoids falsifying a death certificate," "using SINADEF reduces the risk of errors," and "using SINADEF allows for filling out a certificate in less time," and perceived ease of use as shown by "I think SINADEF is easy to use," as well as other perceptions related to image, job relevance, output quality, and results demonstrability. As a subjective norm, no proposition was found associated with behavioral intention to use (see Table 2). There was also no significant association between age, sex, occupation characteristics, or training in filling out death certificates with behavioral intention to use.

Table 2. Behavioral intention to use according to the perceptions of the respondents.

Behavioral intention to use	Agree (n=256)		Disagree (n=16)		*P* value[a]
	Agree, n (%)	Disagree, n (%)	Agree, n (%)	Disagree, n (%)	
Perceived usefulness					
Using SINADEF[b] avoids falsifying a death certificate	220 (97.3)	36 (78.3)	6 (2.7)	10 (21.7)	<.001
Using SINADEF reduces the risk of errors in death certificates	221 (97.8)	35 (76.1)	5 (2.2)	11 (23.9)	<.001
Using SINADEF allows me to produce a death certificate in less time than doing it in paper format	207 (97.6)	49 (81.7)	5 (2.4)	11 (18.3)	<.001
Perceived ease of use					
I think SINADEF is easy to use.	236 (97.1)	20 (68.9)	7 (2.9)	9 (31)	<.001
It's easy to get a password to use SINADEF.	148 (96.1)	108 (91.5)	6 (3.9)	10 (8.5)	.11
It's easy to get technical support from SINADEF when you need it	93 (95.9)	163 (93.1)	4 (4.1)	12 (6.9)	.36
It's easy to recover the SINADEF password when you forget it or it's blocked	129 (95.6)	127 (92.7)	6 (4.4)	10 (7.3)	.25
Subjective norm					
I have to use SINADEF to certify deaths because it's already established that way	208 (94.5)	48 (92.3)	12 (5.5)	4 (7.7)	.54
The authorities of my hospital and the Ministry of Health could sanction me if I do not use SINADEF to certify deaths	81 (93.1)	175 (94.6)	6 (6.9)	10 (5.4)	.63
I have to use SINADEF because everyone already uses it to certify deaths	195 (95.6)	61 (89.7)	9 (4.4)	7 (10.3)	.07
Image					
I feel comfortable with information and communication technologies	233 (95.5)	23 (82.1)	11 (4.5)	5 (17.9)	.004
I consider myself a person open to change	254 (94.8)	2 (50)	14 (5.2)	2 (50)	.02[c]
I have good adaptability	245 (94.6)	11 (84.6)	14 (5.4)	2 (15.4)	.14
Job relevance					
When SINADEF is implemented, the causes of death of the population will be known in a timelier manner	237 (95.9)	19 (76)	10 (4.1)	6 (24)	<.001
The use of SINADEF will help the directors or managers of health organizations make decisions	203 (96.7)	53 (85.5)	7 (3.3)	9 (14.5)	.001
If SINADEF is consolidated, it will be possible to quickly consult the data of the deceased	250 (96.2)	6 (50)	10 (3.8)	6 (50)	<.001
Output quality					
The SINADEF death certificate is of higher quality than the paper format	224 (87.5)	7 (43.8)	32 (12.5)	9 (56.3)	<.001
Results demonstrability					
I can easily communicate the results of my experience using SINADEF	224 (87.5)	7 (43.8)	32 (12.5)	9 (56.3)	<.001
For me, the results of using SINADEF will be visible	217 (84.8)	7 (43.8)	39 (15.2)	9 (56.3)	<.001

[a] Chi-square test was performed.

[b] SINADEF: National Death Information System.

[c] Fisher exact test was performed.

Multivariate Analysis

Three models were tested that had perceived usefulness as a dependent variable, corresponding to the 3 perceptions that evaluated perceived usefulness: (1) SINADEF avoids death certificate forgery, (2) SINADEF reduces the possibility of errors, and (3) SINADEF is faster than filling out paper forms. We use as perceived usefulness predictors the perceptions of the variables proposed in the TAM2 model: subjective norm, image, job relevance, output quality, and results demonstrability. The first model reached an overall percentage of correctly classified cases of 84.2% and found subjective norm, image, job relevance, and results demonstrability as predictors significantly associated with perceived usefulness but did not find perceived ease of use as a variable significantly related to perceived usefulness. The second model reached a global percentage of correctly classified cases of 88.2% and found perceived ease of use, subjective norm, image, and job relevance as predictors of perceived usefulness. The third model reached a global percentage of correctly classified cases of 82.7% and, in addition to perceived ease of use, only found job relevance and results demonstrability as perceived usefulness predictors. The second model was chosen because it showed the highest percentage of prediction and included 3 of the 5 variables proposed in the TAM2 model as significant predictors of perceived usefulness. The second model was based on the perception that SINADEF reduces the possibility of making errors when making a death certificate as an indicator of perceived usefulness. It was found that they were significantly associated with perceived usefulness: subjective norms such as the belief that SINADEF should be used because everyone already uses it to certify deaths (odds ratio [OR] 2.407, 95% CI 1.008-5.75; *P*=.048); image, the perception of being comfortable with ICT (OR 5.363, 95% CI 1.886-15.255; *P*=.002); job relevance such as the perception that the use of SINADEF will help health directors to make decisions (OR 4.49, 95% CI 1.978-10.2; *P*<.001); and perceived ease of use as the perception that SINADEF is easy to use (OR 18.95, 95% CI 6.634-54.156; *P*<.001). The model reached a predictive ability of 88.2% (see Table 3).

The final model found the following statistically associated predictors with behavioral intention to use: perceived usefulness, measured by the belief that SINADEF reduces the risk of error in death certificate (OR 8.515, 95% CI 2.242-32.3; *P*=.002) and perceived ease of use, due to the belief that SINADEF is easy to use (OR 10.116, 95% CI 2.443-41.883; *P*=.001). Also included in this model was the variable training in the filling of death certificates, a variable external to the TAM2 model but which had a very important contribution (OR 8.324, 95% CI 1.615-42.895; *P*=.01). The final model reached a predictive ability of 95.6% (see Table 4 and Figure 2).

Table 3. Logistic regression of perceived usefulness predictors.

Predictor	β	Odds ratio (95% CI)	*P* value
Subjective norm: I have to use SINADEF[a] because everyone already uses it to certify deaths	.879	2.407 (1.008-5.75)	.048
Image: I feel comfortable with information and communication technologies	1.502	5.363 (1.886-10.2)	.002
Job relevance: The use of SINADEF will help the directors or managers of health organizations make better decisions	1.045	4.491 (1.978- 5.931)	<.001
Perceived ease of use: I think SINADEF is easy to use	2.942	18.955 (6.634-54.156)	<.001
Constant	–10.153	0.000	<.001

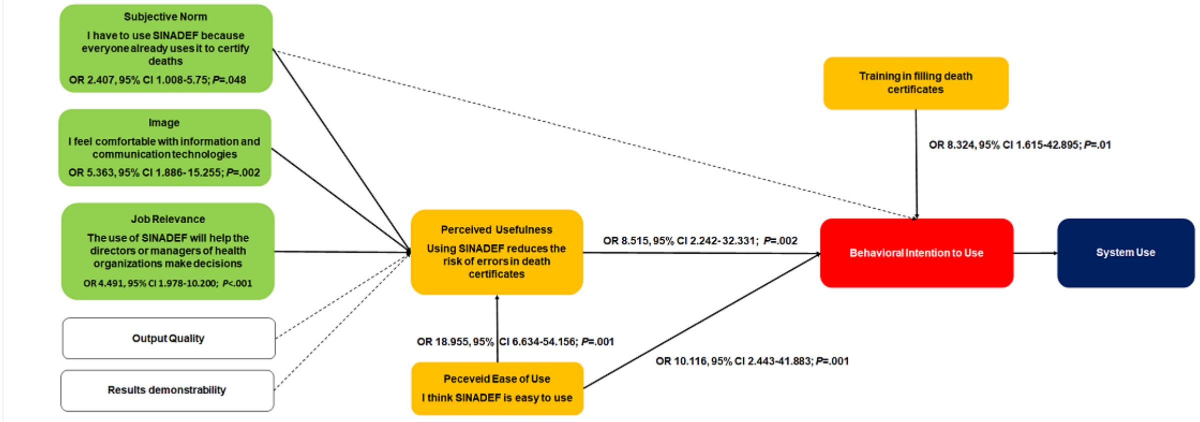
[a] SINADEF: National Death Information System.

Table 4. Logistic regression of behavioral intention to use predictors.

Predictor	β	Odds ratio (95% CI)	*P* value
Perceived usefulness: Using SINADEF[a] reduces the risk of errors in death certificates	2.142	8.515 (2.242-32.331)	.002
Perceived ease of use: I think SINADEF is easy to use	2.314	10.116 (2.443-41.883)	.001
External variable: Training in filling death certificate	2.119	8.324 (1.615-42.895)	.01
Constant	–10.313	0.000	<.001

[a] SINADEF: National Death Information System.

Figure 2. Summary of SINADEF behavioral intention to use predictors in the logistic regression models used. OR: odds ratio; SINADEF: National Death Information System.



Discussion

Principal Results

Although it may seem that the implementation of an ICT project will be successful, this is not always the case. Projects can be successful, considering their management (scope, costs, and deadlines) from their gestation to their implementation [25]. However, in the case of ICT projects in health, their final adoption will also depend on the beliefs of health personnel, in which the real barriers and facilitators for their use can be identified [16]. SINADEF has contributed to improving the coverage [6] and the quality of information on causes of death [5]. However, until the end of 2017, its use rate was only 50%; therefore, it was pertinent to evaluate the intention of system use.

Most of the physicians interviewed considered themselves open to change. They believed that SINADEF would allow them to quickly consult the data of deceased people or that they could adopt innovative methods (95% or more). However, more than half of them did not agree with the idea that if they do not use the SINADEF, they should be sanctioned (68%) or that it is easy to obtain technical help when they have problems with the system (64%). Bivariate analysis revealed that perceived ease of use, measured by the belief that SINADEF is easy to use, was significantly associated with behavioral intention to use (*P*<.001). Perceived ease of use was measured on 4 propositions: the system is easy to use, it is easy to get a password, it is easy to get technical support, and it is easy to recover a password. Only the belief that the system is easy to use had a degree of agreement of 89.3%, while the other proposals had an agreement of less than 60%. Furthermore, in the bivariate analysis, only the perception that the system is easy to use was associated with behavioral intention to use; therefore, it was also part of the final logistic regression model to predict behavioral intention to use. In turn, perceived usefulness was measured with any of 3 ideas—SINADEF avoids falsifying a death certificate, fewer errors are made with SINADEF when creating a death certificate, or a death certificate is produced faster with SINADEF—and it was found to be associated with behavioral intention to use (*P*<.001). In addition, most or all the image, job relevance, output quality, and results demonstrability proposals were significantly associated with behavioral intention to use (*P*<.05).

In the multivariate analysis, among the 3 propositions that measured perceived usefulness, the proposition "using SINADEF reduces the risk of errors in death certificates" had the highest predictive capacity (88.2%). Subjective norm measured as "I have to use the SINADEF because everyone already uses it to certify deaths" (OR 2.4, 95% CI 1.0-5.8; *P*=.048), image measured as "I feel comfortable with technology" (OR 5.4, 95% CI 1.9-10.2; *P*=.002), job relevance measured as "the use of SINADEF will help directors or managers of health organizations to make decisions" (OR 4.5, 95% CI 2.0-5.9; *P*<.001), and perceived ease of use measured as "I believe that SINADEF is easy to use" (OR 19, 95% CI 6.6-54.2; *P*<.001) were significantly associated with perceived usefulness. In the final model, we found a significant association between perceived usefulness measured as "using SINADEF reduces the risk of errors in death certificates" (OR 8.5, 95% CI 2.2-32.3; *P*=.002), perceived ease of use measured as "I think that SINADEF is easy to use" (OR 10.1, 95% CI 2.4-41.9; *P*=.001), and training to fill out the death certificate (OR 8.3, 95% CI 1.9-42-9; *P*=.01), with behavioral intention to use. The correct classification of cases by the final model was 95.6%.

Limitations of This Study

A limitation of this study is the selection of the participating physicians. The sample was made up of physicians who responded to an invitation sent to the email registered in SINADEF, and the response rate was less than 2%. This is because most physicians did not read the messages because they have an email account assigned by the institution where they work, which they usually do not use, or because the message went to the "spam" folder. Regarding the participants, there is a probability that they have incurred a social desirability bias. Despite this, we believe the sample to be representative of physicians in terms of graduation from universities in the country, specialties, sex, and age. In addition, the sample size was adequate for carrying out the proposed statistical models. Another aspect to consider is that a significant proportion of the surveys was completed in 2020, the initial year of the COVID-19 pandemic in Peru, in which SINADEF played an important role in monitoring mortality in Peru, and it is possible that this factor, not measured in this study, influenced the favorable opinion of physicians toward intention to use. It is also possible that the proposals to be accepted or rejected by the participants do not represent all the factors that may be linked to the process of the medical certification of deaths in Peru. It should also be considered that, as in all observational studies, it should not be inferred that there is a causal relationship between perceived usefulness and perceived ease of use and training with behavioral intention to use but only an association between them. This study also did not evaluate the performance or the real and objective effectiveness of the technology, but only the beliefs of the users about its usefulness that will motivate its future use and its final adoption.

Comparison With Prior Work

There is a need to institutionalize SINADEF to intensify the process of improving the coverage and quality of information on causes of death. The use of paper forms, from the logistics of their supply to the procedures that ensure their entry into a database, can be a barrier to improving death certification [6]. This situation was verified during the COVID-19 pandemic by observing that the use of paper forms to certify a death decreased the probability that the data would enter the mortality database [9]. In addition, considering that there are 18 mortality information subsystems in Peru, most of which use paper forms and none of which have complete information [26], the consolidation of SINADEF as the main information system would reduce the fragmentation of currently existing data.

The usefulness that SINADEF has shown for mortality surveillance during the COVID-19 pandemic is another aspect that highlights the importance of contributing to its institutionalization. During the COVID-19 pandemic, SINADEF was used in several studies [10-12,27] as an important source of information to document excess mortality from all causes in Peru and was recognized as the most reliable way to measure the severity and the impact of COVID-19 on the population. With the support of SINADEF, an excess mortality of 371.9 per 100,000 inhabitants in 2020 was documented in Peru [28], and in general, all the researchers agreed in highlighting the importance of strengthening the mortality documentation system in Peru.

We have not identified studies of the application of TAM to evaluate acceptability, specifically in mortality information systems, but there are several publications that have studied the technological acceptance of health information systems that include electronic medical records or electronic health records. As in other studies, we found that perceived usefulness and perceived ease of use are powerful predictors of the intention to use SINADEF. These findings are consistent with most of the previous research synthesized in systematic reviews. Gagnon et al [29] carried out a systematic review of the factors that influence the adoption of ICT in health and found that the most common direct determinants of adoption were usefulness and ease of use. Later, Rahimi et al [21] reviewed 134 publications on the use of TAM to find out the perceptions of users of health information systems as predictors of the use of technology. Although the reference framework used or the methods for analyzing the results may differ, most of the studies reviewed agree in confirming that the perceived usefulness and the perceived ease of use are important predictors of the intention to use the technology.

In our study, the direct predictors of behavioral intention to use were perceived usefulness, perceived ease of use, and training in filling out death certificates, and the predictors of perceived usefulness were subjective norm, image, and job relevance. When reviewing the predictor variables of behavioral intention to use, in the cases under study, considerable heterogeneity was observed. Some studies found only 1 primary variable from TAM, such as perceived usefulness or perceived ease of use, associated with intention to use [30,31], and other studies reported subjective norm or job relevance directly related to behavioral intention to use, without being intermediated by perceived usefulness [30-32]. There are also studies that describe predictors of perceived ease of use, such as job relevance, management support, and training or computational self-efficacy [33,34], and some others found that age, sex, or clinical specialty were predictors of both perceived usefulness and perceived ease of use [31]. This variability supports the idea of contextualizing the results of each study not only referring to the type of technology used but also to the organizational culture in which the technology will work.

In our study, subjective norm (“I use SINADEF because everyone already uses it”), image (“I feel comfortable with technology”), and job relevance (“SINADEF helps managers to make decisions”) were predictors of perceived usefulness. This is partially consistent with studies reporting peer influence [35] or computational self-efficacy as perceived usefulness predictors [32,34]. Perceived usefulness measured by the idea that the system helps reduce errors is a performance indicator. In this regard, a study [36] found that improvements in performance are related to the intentions of health professionals to use electronic medical record systems.

In our perceived ease-of-use study, measured in its simplest and most direct way, “SINADEF is easy to use” was a strong and significant determinant of physicians’ intention to use SINADEF and influenced their perceived usefulness. However, the physicians did not agree with the perceived ease of use propositions that “it’s easy to obtain a username and password” or that “it’s easy to obtain technical support,” and indeed, these were not significantly related to the intention to use. In this regard, Boonstra and Broekhuis [37] reported that the lack of technical support to address problems that arise during system operation is a barrier to the adoption of the electronic medical record.

Training in filling out death certificates had a significant direct influence on behavioral intention to use. In this regard, the implementation of SINADEF was accompanied by a training process both in the competence to identify the cause of death and the technical management of the computer application. It is recognized that training to improve physicians’ knowledge regarding the proper filling of death certificates will improve the usability of mortality statistics. This study demonstrated that this activity is a direct predictor of the use of SINADEF in Peru [38]. This is consistent with the identification of the lack of technical knowledge and insufficient skills for the management of ICTs as a barrier to the adoption of technologies by becoming a source of resistance that hinders their adoption [37].

Conclusions

According to our study, it seems that the intention to use SINADEF is related to the perception that it is an easy-to-use system, it is widely accepted, it improves the performance of the physicians who use it, and it helps to manage health services. Additionally, training in the filling of death certificates plays an important role in the intention to use the system. It also informs ICT decision-makers of, for example, opportunities for improvement to address possible barriers that may limit the sustainability of the system, such as deficiencies in technical support and in the timely resolution of emerging problems. This study provides important knowledge based on the opinion of physicians on the intention to use SINADEF, which should effectively contribute to its institutionalization in the country.

Acknowledgments

We thank Vital Strategies and the Bloomberg “Data for Health” initiative for the opportunity to explore this topic. We thank the professors of San Marcos University who participated as reviewers of the manuscript drafts. We thank Carolina Tarqui, Javier Silva, and Diana Kumar for their contributions and comments.

Conflicts of Interest

None declared.

References

1. AbouZahr C, de Savigny D, Mikkelsen L, Setel P, Lozano R, Nichols E, et al. Civil registration and vital statistics: progress in the data revolution for counting and accountability. The Lancet 2015 Oct;386(10001):1373-1385 [FREE Full text] [doi: 10.1016/s0140-6736(15)60173-8]
2. Sibai A. Mortality certification and cause-of-death reporting in developing countries. Bull World Health Organ 2004 Feb;82(2):83 [FREE Full text] [Medline: 15042227]
3. Mathers C, Fat D, Inoue M, Rao C, Lopez A. Counting the dead and what they died from: an assessment of the global status of cause of death data. Bull World Health Organ 2005 Mar;83(3):171-177 [FREE Full text] [Medline: 15798840]
4. Soto Cáceres Cabanillas R, Fernandez- Mogollón J, Daniel V. Calidad de llenado del certificado de defunción en dos hospitales de Chiclayo,Perú, 2016. Rev.exp.med 2019 Jul 31;5(2):76 [FREE Full text] [doi: 10.37065/rem.v5i2.344]
5. Miki J, Rampatige R, Richards N, Adair T, Cortez-Escalante J, Vargas-Herrera J. Saving lives through certifying deaths: assessing the impact of two interventions to improve cause of death data in Perú. BMC Public Health 2018 Dec 03;18(1):1329 [FREE Full text] [doi: 10.1186/s12889-018-6264-1] [Medline: 30509233]
6. Vargas-Herrera J, Ruiz KP, Nuñez GG, Ohno JM, Pérez-Lu JE, Huarcaya W, et al. [Preliminary results of the strengthening of the national death registry information system]. Rev Peru Med Exp Salud Publica 2018;35(3):505-514 [FREE Full text] [doi: 10.17843/rpmesp.2018.353.3913] [Medline: 30517488]
7. Naghavi M, Makela S, Foreman K, O'Brien J, Pourmalek F, Lozano R. Algorithms for enhancing public health utility of national causes-of-death data. Popul Health Metr 2010 May 10;8:9 [FREE Full text] [doi: 10.1186/1478-7954-8-9] [Medline: 20459720]
8. Resolución Ministerial N° 280-2016-RM. Ministry of Health. 2016 Apr 25. URL: https://www.gob.pe/es/l/862704 [accessed 2019-10-28]
9. Silva-Valencia J, Adair T, Hart J, Meza G, Vargas Herrera J. How has COVID-19 impacted the civil registration and vital statistics system in Loreto, Perú? Evidence using process mapping and qualitative analysis. BMJ Open 2021 Nov 19;11(11):e055024 [FREE Full text] [doi: 10.1136/bmjopen-2021-055024] [Medline: 34799366]
10. Villarreyes E, Luna A, Soriano A. Estimating the number of deaths due to COVID-19 in Lima and Peru during March and April 2020 using ARIMA time series and modeling Internet. Research Square. Preprint posted online Sept 1, 2020. URL: https://www.researchsquare.com/article/rs-62317/v1 [accessed 2022-04-08]
11. Yanez J, Chung S, Inga-Berrospi F, Mejia C. Demographic and Geographic COVID-19 Death Risk Factors in Peru. A Nationwide Analysis. SSRN Journal 2020 Jul 23:1-29 [FREE Full text] [doi: 10.2139/ssrn.3648543]
12. Hernández-Vásquez A, Azañedo D. COVID-19 in Peru: the need to pay attention to the high number of police deaths due to the pandemic. J Public Health (Oxf) 2020 Aug 18;42(3):640-641 [FREE Full text] [doi: 10.1093/pubmed/fdaa079] [Medline: 32529227]
13. Ramírez-Soto M, Ortega-Cáceres G. Analysis of Excess All-Cause Mortality and COVID-19 Mortality in Peru: Observational Study. Tropical Medicine and Infectious Disease 2022 Mar 05:44 [FREE Full text] [doi: 10.3390/tropicalmed7030044]
14. Monteagudo J, Serrano L, Hernández SC. Telemedicine cience or fiction? An Sist Sanit Navar 2005;28(3):309 [FREE Full text]
15. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
16. Holden R, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Mar;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
17. Chismar W, Wiley-Patton S. Test of the technology acceptance model for the internet in pediatrics. Proc AMIA Symp 2002:155-159 [FREE Full text] [Medline: PMC2244480]
18. Duyck P, Pynoo B, Devolder P, Voet T, Adang L, Vercruysse J. User Acceptance of a Picture Archiving and Communication System. Methods Inf Med 2018 Jan 18;47(02):149-156 [FREE Full text] [doi: 10.3414/me0477]
19. Asua J, Orruño E, Reviriego E, Gagnon M. Healthcare professional acceptance of telemonitoring for chronic care patients in primary care. BMC Med Inform Decis Mak 2012 Nov 30;12:139 [FREE Full text] [doi: 10.1186/1472-6947-12-139] [Medline: 23194420]
20. Hsiao C, Chang J, Tang K. Exploring the influential factors in continuance usage of mobile social Apps: Satisfaction, habit, and customer value perspectives. Telematics and Informatics 2016 May;33(2):342-355 [FREE Full text] [doi: 10.1016/j.tele.2015.08.014]
21. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A Systematic Review of the Technology Acceptance Model in Health Informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
22. Yamakawa P, Guerrero C, Rees G. Factors that influence the use of mobile banking services in Peru. Revista Universidad y Empresa 2013 Jul;15(25):131 [FREE Full text]

23. Venkatesh V, Davis F. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204 [FREE Full text] [doi: 10.1287/mnsc.46.2.186.11926]
24. Galicia Alarcón L, Balderrama Trápaga J, Edel Navarro R. Content validity by experts judgment: Proposal for a virtual tool. Ap 2017 Oct 1;9(2):42-53 [FREE Full text] [doi: 10.32870/ap.v9n2.993]
25. Varajão J, Magalhães L, Freitas L, Ribeiro P, Ramos J. Implementing Success Management in an IT project. Procedia Computer Science 2018;138:891-898 [FREE Full text] [doi: 10.1016/j.procs.2018.10.116]
26. Cobos MD, Sant FC, Miki J, Vargas-Herrera J, Botchway E, Woode S, et al. Mapping sources and silos of mortality data: Case studies in Peru and Ghana. Bloomberg Philanthropies Data for Health Initiative, Civil Registration and Vital Statistics Improvement, University of Melbourne. 2019 May. URL: https://crvsgateway.info/file/11077/3229 [accessed 2020-06-15]
27. Sempé L, Lloyd-Sherlock P, Martínez R, Ebrahim S, McKee M, Acosta E. Estimation of all-cause excess mortality by age-specific mortality patterns for countries with incomplete vital statistics: a population-based study of the case of Peru during the first wave of the COVID-19 pandemic. Lancet Reg Health Am 2021 Oct;2 [FREE Full text] [doi: 10.1016/j.lana.2021.100039] [Medline: 34693394]
28. Ramírez-Soto M, Ortega-Cáceres G. Analysis of Excess All-Cause Mortality and COVID-19 Mortality in Peru: Observational Study. Trop Med Infect Dis 2022 Mar 05:44 [FREE Full text] [doi: 10.3390/tropicalmed7030044]
29. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Mar;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
30. Ketikidis P, Dimitrovski T, Lazuras L, Bath P. Acceptance of health information technology in health professionals: an application of the revised technology acceptance model. Health Informatics J 2012 Jun;18(2):124-134 [FREE Full text] [doi: 10.1177/1460458211435425] [Medline: 22733680]
31. Hwang H, Dutta B, Chang H. The Differing Effect of Gender and Clinical Specialty on Physicians' Intention to Use Electronic Medical Record. Methods Inf Med 2019 Dec;58(S 02):e58-e71 [FREE Full text] [doi: 10.1055/s-0039-1695718] [Medline: 31499570]
32. Gagnon M, Ghandour E, Talla P, Simonyan D, Godin G, Labrecque M, et al. Electronic health record acceptance by physicians: testing an integrated theoretical model. J Biomed Inform 2014 Apr;48:17-27 [FREE Full text] [doi: 10.1016/j.jbi.2013.10.010] [Medline: 24184678]
33. Abdekhoda M, Dehnad A, Zarei J. Determinant factors in applying electronic medical records in healthcare. East Mediterr Health J 2019 Mar 19;25(1):24-33 [FREE Full text] [doi: 10.26719/emhj.18.007] [Medline: 30919922]
34. Dünnebeil S, Sunyaev A, Blohm I, Leimeister J, Krcmar H. Determinants of physicians' technology acceptance for e-health in ambulatory care. Int J Med Inform 2012 Nov;81(11):746-760. [doi: 10.1016/j.ijmedinf.2012.02.002] [Medline: 22397989]
35. Hsieh P. Physicians' acceptance of electronic medical records exchange: an extension of the decomposed TPB model with institutional trust and perceived risk. Int J Med Inform 2015 Jan;84(1):1-14. [doi: 10.1016/j.ijmedinf.2014.08.008] [Medline: 25242228]
36. Kim S, Lee K, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016 Jan 30;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
37. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010 Aug 06;10:231 [FREE Full text] [doi: 10.1186/1472-6963-10-231] [Medline: 20691097]
38. Akhade S, Dash S, Akhade K. The knowledge assessment and reducing the errors of medical certificate of cause of death with sensitization training of physicians: A quality improvement intervention study. J Educ Health Promot 2022;11:19 [FREE Full text] [doi: 10.4103/jehp.jehp_502_21] [Medline: 35281401]

Abbreviations

ICT: information and communication technology
OR: odds ratio
SINADEF: National Death Information System
TAM: Technology Acceptance Model

Edited by R Kukafka; submitted 10.11.21; peer-reviewed by S Nagavally, F Velayati, KM Kuo; comments to author 05.12.21; revised version received 27.04.22; accepted 30.05.22; published 15.08.22.

Please cite as:
Vargas-Herrera J, Meneses G, Cortez-Escalante J
Physicians' Perceptions as Predictors of the Future Use of the National Death Information System in Peru: Cross-sectional Study
J Med Internet Res 2022;24(8):e34858
URL: https://www.jmir.org/2022/8/e34858
doi:10.2196/34858
PMID:35969435

Original Paper

Web-Based Risk Prediction Tool for an Individual's Risk of HIV and Sexually Transmitted Infections Using Machine Learning Algorithms: Development and External Validation Study

Xianglong Xu[1,2,3], PhD; Zhen Yu[2,3,4], MSc; Zongyuan Ge[4], PhD; Eric P F Chow[1,2,5], PhD; Yining Bao[3], MSc; Jason J Ong[1,2,3], PhD; Wei Li[6], PhD; Jinrong Wu[7], MSc; Christopher K Fairley[1,2,3], PhD; Lei Zhang[1,2,3], PhD

[1]Melbourne Sexual Health Centre, Alfred Health, Melbourne, Australia
[2]Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia
[3]China Australia Joint Research Center for Infectious Diseases, Xi'an Jiaotong University Health Science Centre, Xi'an, China
[4]Monash e-Research Centre, Faculty of Engineering, Airdoc Research, Nvidia AI Technology Research Centre, Monash University, Melbourne, Australia
[5]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia
[6]School of Public Health, Southeast University, Nanjing, China
[7]Research Centre for Data Analytics and Cognition, La Trobe University, Melbourne, Australia

Corresponding Author:
Lei Zhang, PhD
Melbourne Sexual Health Centre
Alfred Health
580 Swanston Street
Melbourne, 3053
Australia
Phone: 61 3 9341 6230
Fax: 61 3 9341 6264
Email: lei.zhang1@monash.edu

Abstract

Background: HIV and sexually transmitted infections (STIs) are major global public health concerns. Over 1 million curable STIs occur every day among people aged 15 years to 49 years worldwide. Insufficient testing or screening substantially impedes the elimination of HIV and STI transmission.

Objective: The aim of our study was to develop an HIV and STI risk prediction tool using machine learning algorithms.

Methods: We used clinic consultations that tested for HIV and STIs at the Melbourne Sexual Health Centre between March 2, 2015, and December 31, 2018, as the development data set (training and testing data set). We also used 2 external validation data sets, including data from 2019 as external "validation data 1" and data from January 2020 and January 2021 as external "validation data 2." We developed 34 machine learning models to assess the risk of acquiring HIV, syphilis, gonorrhea, and chlamydia. We created an online tool to generate an individual's risk of HIV or an STI.

Results: The important predictors for HIV and STI risk were gender, age, men who reported having sex with men, number of casual sexual partners, and condom use. Our machine learning–based risk prediction tool, named MySTIRisk, performed at an acceptable or excellent level on testing data sets (area under the curve [AUC] for HIV=0.78; AUC for syphilis=0.84; AUC for gonorrhea=0.78; AUC for chlamydia=0.70) and had stable performance on both external validation data from 2019 (AUC for HIV=0.79; AUC for syphilis=0.85; AUC for gonorrhea=0.81; AUC for chlamydia=0.69) and data from 2020-2021 (AUC for HIV=0.71; AUC for syphilis=0.84; AUC for gonorrhea=0.79; AUC for chlamydia=0.69).

Conclusions: Our web-based risk prediction tool could accurately predict the risk of HIV and STIs for clinic attendees using simple self-reported questions. MySTIRisk could serve as an HIV and STI screening tool on clinic websites or digital health platforms to encourage individuals at risk of HIV or an STI to be tested or start HIV pre-exposure prophylaxis. The public can use this tool to assess their risk and then decide if they would attend a clinic for testing. Clinicians or public health workers can use this tool to identify high-risk individuals for further interventions.

(J Med Internet Res 2022;24(8):e37850) doi:10.2196/37850

KEYWORDS

HIV; sexually transmitted infections; syphilis; gonorrhea; chlamydia; sexual health; sexual transmission; sexually transmitted; prediction; web-based; risk assessment; machine learning; model; algorithm; predictive; risk; development; validation

Introduction

HIV and sexually transmitted infections (STIs) are major global public health concerns [1,2]. The World Health Organization (WHO) estimated that over 1 million curable STIs occur every day among people aged 15 years to 49 years worldwide [3]. An estimated 29,090 people have been infected with HIV in Australia as of the end of 2020, with an HIV prevalence rate of 0.14% among people over 15 years old [4]. The estimated undiagnosed HIV infection rate among all people living with HIV in Australia was about 9% in 2020 [4]. Gonorrhea, chlamydia, and early syphilis can be asymptomatic. There were large increases in STIs in Australia between 2013 and 2017. The notification rates of STIs for chlamydia increased from 302.2/100,000 to 394.9/100,000 in men and from 430.7/100,000 to 441.8/100,000 in women, gonorrhea increased from 91.1/100,000 to 174.2/100,000 in men and from 39.6/100,000 to 61.8/100,000 in women, and syphilis increased from 12.3/100,000 to 31.1/100,000 in men and from 1.4/100,000 to 5.5/100,000 in women [5]. In addition, STIs account for a large health and economic burden in limited-income countries [6].

In response to the rising rates of STIs, the WHO proposed the “Global health sector strategy on Sexually Transmitted Infections, 2016-2021,” which aimed to end STI epidemics as public health concerns by 2030. This specifically includes a 90% reduction in gonorrhea incidence globally from the 2018 global baseline and achieving a rate of ≤50 congenital syphilis cases per 100,000 live births in 80% of countries [7]. In 2018, the United Nations proposed “The 2030 Agenda for Sustainable Development,” which called for an end to the AIDS epidemic by 2030 [8]. Key to the effective control of these infections is accessible health care and, in particular, frequent testing because treated infections rapidly become noninfectious [2]. Screening of asymptomatic individuals is important for diagnosis, treatment, prevention, and control of HIV and STIs [9]. Barriers to testing include misjudgment of an individual's HIV and STI risk, limited availability of testing, and high cost of testing [10]. Therefore, developing innovative tools will help individuals accurately judge their risk of HIV and STIs, hence increasing screening in high-risk individuals.

An easily accessible and user-friendly tool that accurately identifies an individual's risk of infection could form part of a web-based risk prediction program and play a role in risk prediction and personalized risk management [11]. Providing the public with risk prediction tools to assist them in estimating the risk of HIV and STIs may encourage those individuals at high risk to test more regularly. A previous study showed that increased risk perceptions were associated with greater STI health care use (eg, testing) [12]. An HIV and STI risk prediction tool may increase risk perceptions and motivate individuals to seek HIV and STI testing or treatment. Another review study suggested that web-based screening apps can effectively increase the uptake of health screening in the general population [13]. However, there is no web-based tool we could identify that provides users with an individual’s current quantitative risk of HIV and STIs (gonorrhea, chlamydia, and syphilis) using self-reported questions.

A number of mathematical techniques can be used to generate an individual’s risk of HIV and STIs. Logistic regression has limitations in predictive analysis that uses complex and big data. Logistic regression methods require strong assumptions and cannot easily deal with nonlinear relationships, interactions, and multicollinearity [14,15]. In contrast, nonlinear machine learning approaches can address these limitations and have numerous advantages (eg, capturing nonlinear relationships and interactions) in predictive analysis using big data [16]. Machine learning also can identify rare health outcomes with high accuracy [17]. Ensemble learning is also a machine learning approach that combines multiple machine learning algorithms to improve the model's performance [18].

Despite the advantages of machine learning approaches, there is an absence of individual risk prediction tools for HIV and STI risk using machine learning models. Existing studies using machine learning algorithms to predict HIV and STI acquisition mainly focus on HIV [19-30], and few focus on STIs [19,21,31]. Of these HIV prediction studies, 4 studies focused on high-risk individuals (such as men who have sex with men [MSM] [20,21,24,29]), 2 studies used imaging or clinical text data [22,30], 4 studies used more than 40 predictors [23,26-28], and 2 studies assessed future but not current HIV prediction [19,25]. Of the STI prediction studies, 1 study was conducted with MSM [21], and the other 2 studies focused on future STI prediction [19,31]. These studies also found that nonlinear machine learning models (eg, random forest [RF], gradient boosting machine [GBM], and neural networks) performed better than logistic regression in HIV and STI prediction [19,21,24,31]. These published studies highlight a lack of machine learning models that use simple self-reported questions, predict both the risk of HIV and STIs, and can be used by both men and women. Therefore, to address the current lack of studies that predict the risk of both STIs and HIV, particularly in lower-risk heterosexual individuals, we aimed to use a stacking ensemble learning framework and self-reported questions to predict HIV and 3 common STIs (gonorrhea, chlamydia, and syphilis) in both men and women and a subsequent web-based HIV and STI risk prediction tool.

Methods

Study Population

The Melbourne Sexual Health Centre (MSHC) is the largest public sexual health center in Victoria, Australia and offers free HIV and STI testing and management [32]. At the MSHC, individuals' demographic information and sexual practices are recorded using a computer-assisted self-interview (CASI) at each visit, at least 3 months apart [33]. We used clinical consultation data from the electronic health record (EHR) at MSHC to develop and validate the risk prediction model. We chose March 2, 2015, as the commencement date because this date was when we adopted a new testing platform for gonorrhea and chlamydia (Aptima Combo, Hologic, Marlborough, MA). Our study data included men and women aged 18 years and older who was tested for HIV or an STI at the MSHC between March 2, 2015, and January 29, 2021. We excluded transgender people and individuals aged younger than 18 years.

We used data from March 2, 2015, to December 31, 2018, as the development data set (training and testing data set). The HIV study data set included training and testing data (88,642 consultations). The syphilis, gonorrhea, and chlamydia study data sets had 92,291, 97,473, and 115,845 consultations, respectively.

We used temporal validation as the external validation to evaluate the transportability and generalizability of our risk prediction models. The COVID-19 epidemic may potentially have changed the demographics of those who attend the MSHC [34]. We performed 2 temporal validations to validate our models further and reduce the possible bias caused by COVID-19. The 2 external validation data sets included data from 2019 as external "validation data 1" and data from January 2020 and January 2021 as external "validation data 2." For HIV, the first external validation data set contained 28,875 consultations, and the second external validation data set contained 18,052 consultations. For syphilis, the first external validation data set contained 30,302 consultations, and the second external validation data set contained 19,150 consultations. For gonorrhea, the first external validation data set contained 36,805 consultations, and the second external validation data set contained 22,886 consultations. For chlamydia, the first external validation data set contained 36,393 consultations, and the second external validation data set contained 22,615 consultations.

Ethical Approval

Ethical approval was granted by the Alfred Hospital Ethics Committee, Melbourne, Australia (project number: 124/18). All methods were carried out following relevant guidelines and regulations of the Alfred Hospital Ethics Committee. As this was a retrospective study involving minimal risk to the privacy of the study participants, the need for informed consent was waived by the Alfred Hospital Ethics Committee. All identifying details of the study participants were removed before any computational analysis.

Predictors

The data fields we selected for inclusion as predictors were informed by literature review, expert opinion, and prior work [21]. The predictors were self-reported questions from the EHR, including demographics, sexual practices, STI history, and STI contact history (summarized in Table 1 and Tables S1-S5 in Multimedia Appendix 1).

Table 1. Characteristics of clinic consultations in the training and testing data set.

Variables	HIV (n=88,642 consultations)	Syphilis (n=92,291 consultations)	Gonorrhea (n=97,473 consultations)	Chlamydia (n=115,845 consultations)
Gender, n (%)				
Female	26,651 (30.1)	27,134 (29.4)	31,282 (32.1)	38,548 (33.3)
Male	61,991 (69.9)	65,157 (70.6)	66,191 (67.9)	77,297 (66.7)
Age at consultation (years), median (IQR)	29.0 (24.0-35.0)	29.0 (25.0-35.0)	28.0 (24.0-35.0)	28.0 (24.0-34.0)
Country of birth, n (%)				
Australia	39,148 (44.2)	40,990 (44.4)	43,881 (45.0)	51,162 (44.2)
Overseas	46,003 (51.9)	47,670 (51.7)	49,835 (51.1)	60,272 (52.0)
Missing	3491 (3.9)	3631 (3.9)	3757 (3.9)	4411 (3.8)
STI[a] symptoms, n (%)				
No	56,175 (63.4)	57,413 (62.2)	54,595 (56.0)	68,584 (59.2)
Yes	25,067 (28.3)	27,150 (29.4)	34,751 (35.7)	38,930 (33.6)
Missing	7383 (8.3)	7728 (8.4)	8127 (8.3)	8331 (7.2)
Men who have sex with men, n (%)				
Not applicable (female)	26,651 (30.1)	27,134 (29.4)	31,282 (32.1)	38,548 (33.3)
No	16,508 (18.6)	17,089 (18.5)	15,245 (15.6)	26,975 (23.3)
Yes	45,483 (51.3)	48,068 (52.1)	50,946 (52.3)	50,322 (43.4)

[a] STI: sexually transmitted infection.

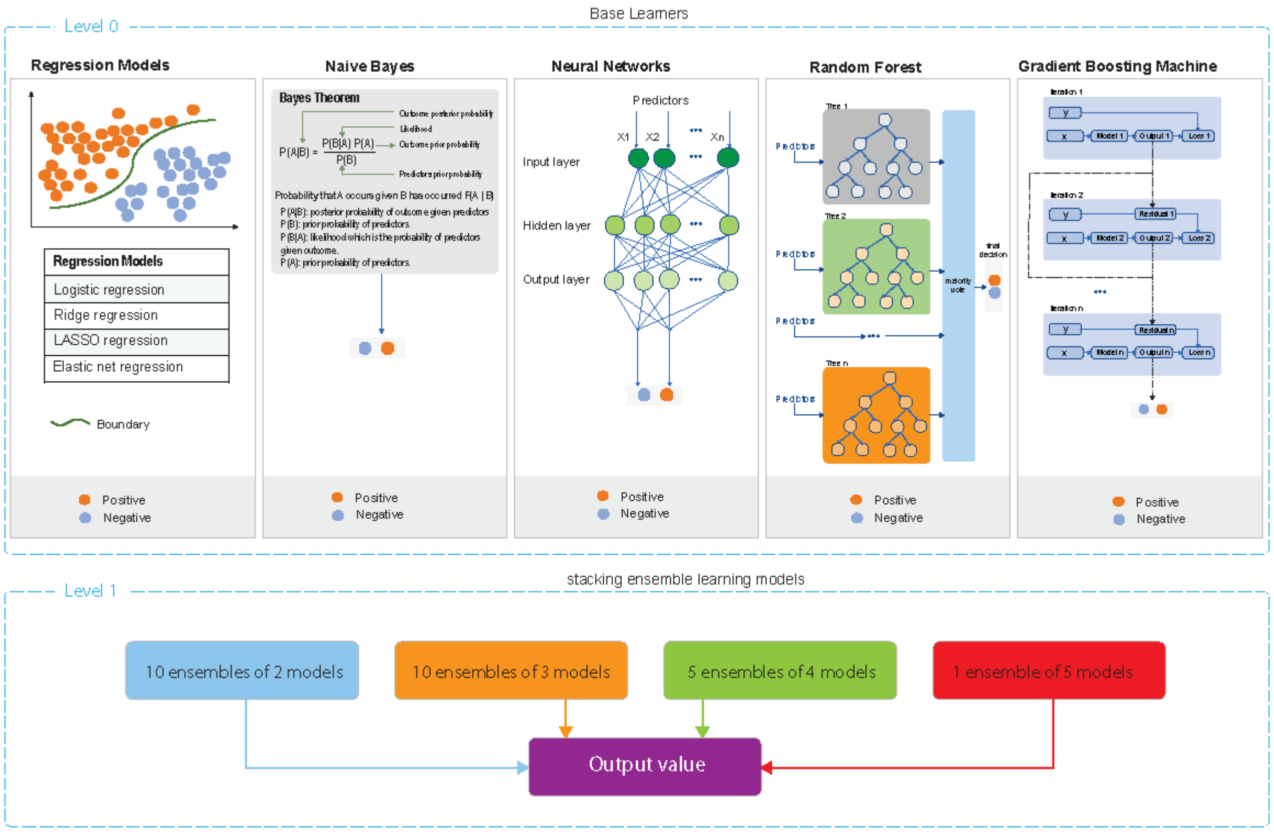
Measurement of Outcomes

HIV infection was defined as a new diagnosis of HIV based on serology. Syphilis infection was defined as a new diagnosis of early syphilis (primary, secondary, and early latent [<2 years]) using a blood test or nucleic amplification test (NAAT). Gonorrhea infection was defined as a new diagnosis of gonorrhea using culture or NAAT at any anatomical site. In the clinic, gonorrhea testing initially occurs with NAAT, and culture is mostly used after a positive NAAT. Chlamydia infection was defined as a new diagnosis using NAAT at any anatomical site. Our previous publications report the diagnostic methods in detail [19,21].

Risk Assessment Model Development

We developed 34 machine learning models to assess the risk of acquiring HIV, syphilis, gonorrhea, and chlamydia (details in Figure 1).

Figure 1. Development of machine learning algorithms. The architecture of the gradient boosting machine was adapted from Feng et al [35]. LASSO: least absolute shrinkage and selection operator.



Base Learner

Logistic regression has been widely used to predict the risk of incident STIs and HIV [36,37]. GBM uses boosting based on decision trees by adjusting the parameters to minimize a loss function and determine the optimal point with the smallest error [38]. RF comprises an ensemble of decision trees using bootstrap aggregation and randomization of predictors to achieve a high degree of predictive accuracy [39]. Naive Bayes (NB) is simple, has high accuracy and speed in large databases, and has been widely used for disease classification [40]. Deep learning (DL) has effectively solved many medical problems and utilizes a hierarchical level of an artificial neural network to perform the classification process [41].

We first established 4 regression models, including logistic regression, ridge regression, least absolute shrinkage and selection operator (LASSO) regression, and elastic net regression (ENR). Based on the preliminary results of the 4 regression analyses, we found that ENR was better than the other 3 regression analyses (details in Multimedia Appendix 1). Considering our previous machine learning study among MSM [21] and the advantages of NB (eg, high accuracy and speed in large databases), we developed 5 base models, including the aforementioned ENR, NB, DL (neural networks), RF, and GBM.

Stacking Ensemble Learning

Stacking ensemble learning is an ensemble learning method that trains a new model based on the combined predictions of 2 (or more) previous machine learning models. Stacking ensemble learning often performs better than individual machine learning techniques [42]. We systematically established 26 ensemble learning models by combining the aforementioned 5 base models to improve the performance of predicting HIV and STIs. Details are in Multimedia Appendix 1 (summarized in Table S6).

Machine Learning Training Techniques

Our models used a one-hot encoding scheme for data classification. We did not impute missing data but created a binary feature vector indicating missing values. The data were considered “imbalanced” given that each of the 4 infections was <10%. Imbalanced data may cause either overfitted or underperformed predictive results [43]. We used 5 x 10 (5 outer folds, 10 inner folds) nested cross-validation (CV) for model selection and training [21,44]. The outer 5-fold CV was used to address the selection bias caused by using a single data set. The inner 10-fold CV was used on the training data set to perform the hyperparameter tuning of machine learning models. We used the area under the curve (AUC) to select the best model. An AUC of 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and >0.9 is considered outstanding [45]. Machine learning models were built using the *h2o* package (version 3.32.1.2) in R software (3.6.1 and R studio 1.2.5019).

Estimating the Risk of HIV and STIs

Our machine learning models predicted the probability of HIV or an STI with a normalized distribution between values 0 and 1. The model-predicted probability was calibrated to the actual prevalence level of HIV and STIs. We used a logistic function to provide a fitting curve for each model-predicted probability and infection prevalence. We regarded the estimated infection prevalence as the “calibrated risk” of infection and presented it in the risk report. We used MATLAB R2019a (MathWorks, Natick, MA) to calibrate the model-predicted probability to the actual prevalence level. The method is described in detail in our previous paper [19]. We classified the calibrated risk of HIV or an STI into 3 risk levels: HIV (low, <0.1%; medium, 0.1%-1.0%; and high, >1.0%), syphilis (low, <0.2%; medium, 0.2%-5.0%; and high, >5.0%), gonorrhea (low, <0.1%; medium, 0.1%-1.0%; and high, ≥1.0%), and chlamydia (low, <2.0%; medium, 2.0%-15.0%; and high, >15.0%).

Establishment of a HIV and STI Risk Prediction Tool

To investigate the effect of predictors, we used the best base machine learning model to calculate the variable importance for HIV, syphilis, gonorrhea, and chlamydia infection. We identified and selected predictors that accounted for more than 80.0% of the overall model performance for each infection. We retrained, retested, and revalidated the best performing model based on these predictors. We compared the AUC, sensitivity, and specificity to re-evaluate the model performance with the shortlisted predictors. We also used the AUC to evaluate the change in performance in the best machine learning model before and after predictor shortlisting (details in Multimedia Appendix 1). We formed a new questionnaire by pooling the important predictors to develop a web-based tool for HIV and STI risk prediction.

Results

Characteristics of the Study Data

Our training and testing data included 216 (0.2% of 88,642 consultations) HIV infections, 787 (1.9% of 92,291 consultations) syphilis infections, 7581 (7.8% of 97,473 consultations) gonorrhea infections, and 10,217 (8.8% of 115,845 consultations) chlamydia infections. The proportion of each of the 4 infection data sets that was men was between 66.7% (77,297/115,845) and 70.6% (65,157/92,291). Further details are provided in Table 1 and Table S1 in Multimedia Appendix 1. The characteristics of the external validation data are shown in Tables S2-S5 in Multimedia Appendix 1.

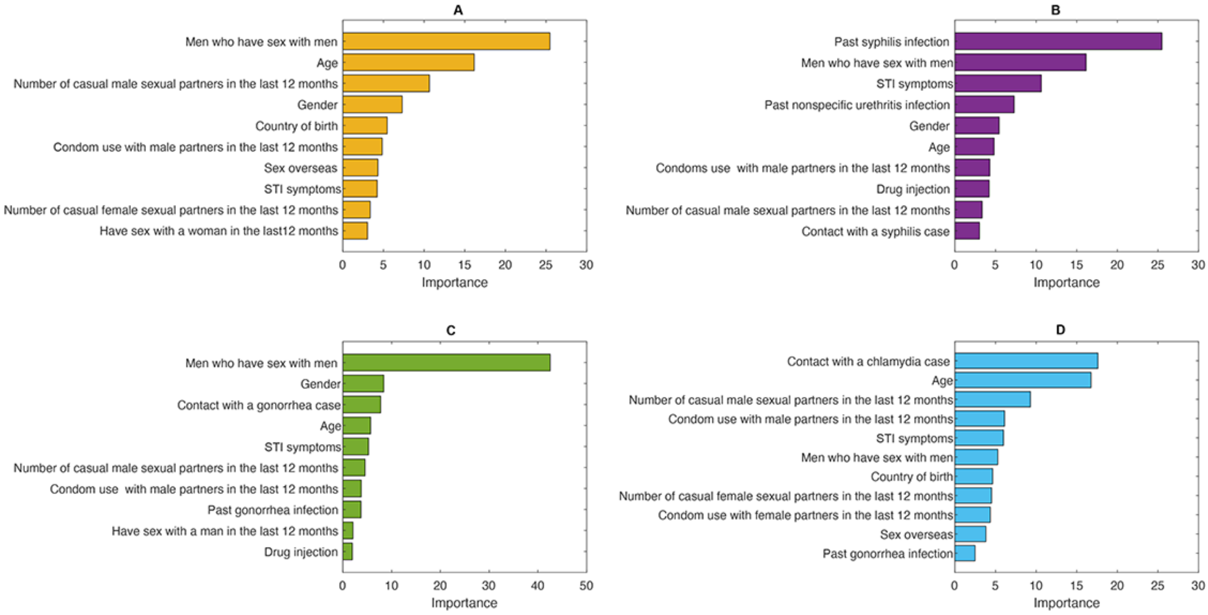
Selecting the Best ML Model for the HIV and STI Risk Prediction Tool

Our results demonstrated that the ensemble learning models performed better than individual machine learning models. Of all 34 models, our best model (ensemble ENR+GBM+RF) provided acceptable or excellent performance on testing data for predicting HIV (AUC=0.78), syphilis (AUC=0.84), gonorrhea (AUC=0.78), and chlamydia (AUC=0.70; Figures S1-S3 in Multimedia Appendix 1). Details on the testing data analysis are provided in Tables S7-S22 in Multimedia Appendix 1. Our external validation results showed very comparable AUCs (0.69-0.85) to the testing data analysis. Details on the external validation analysis are provided in Tables S7-S22 in Multimedia Appendix 1.

Selecting the Most Important Predictors for the HIV and STI Risk Prediction Tool

The top 10 predictors for each of the 4 infections accounted for >80.0% of the overall HIV and STI model performance. These predictors included gender, presence of STI symptoms, MSM, age, country of birth, having sex with a man in the last 12 months, the number of casual male sexual partners in the last 12 months, condom use with male partners in the last 12 months, the number of casual female sexual partners in the last 12 months, drug injection in the last 12 months, sex overseas in the last 12 months, past gonorrhea infection, past nonspecific urethritis infection, past syphilis infection, contact with a gonorrhea case, contact with a chlamydia case, and contact with a syphilis case (Figure 2). We formed the final HIV and STI risk prediction questionnaire with the top 10 predictors for each infection.

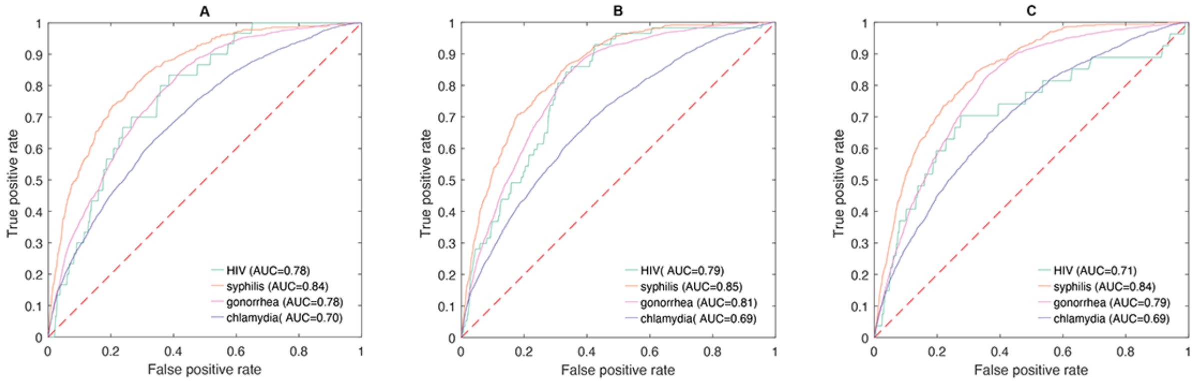
Figure 2. Importance of the top 10 predictors in the prediction of HIV or sexually transmission infections (STIs) using a gradient boosting machine, for detecting (A) HIV, (B) syphilis, (C) gonorrhea, and (D) chlamydia.



Establishment and Evaluation of the HIV and STI Risk Prediction Tool, MySTIRisk

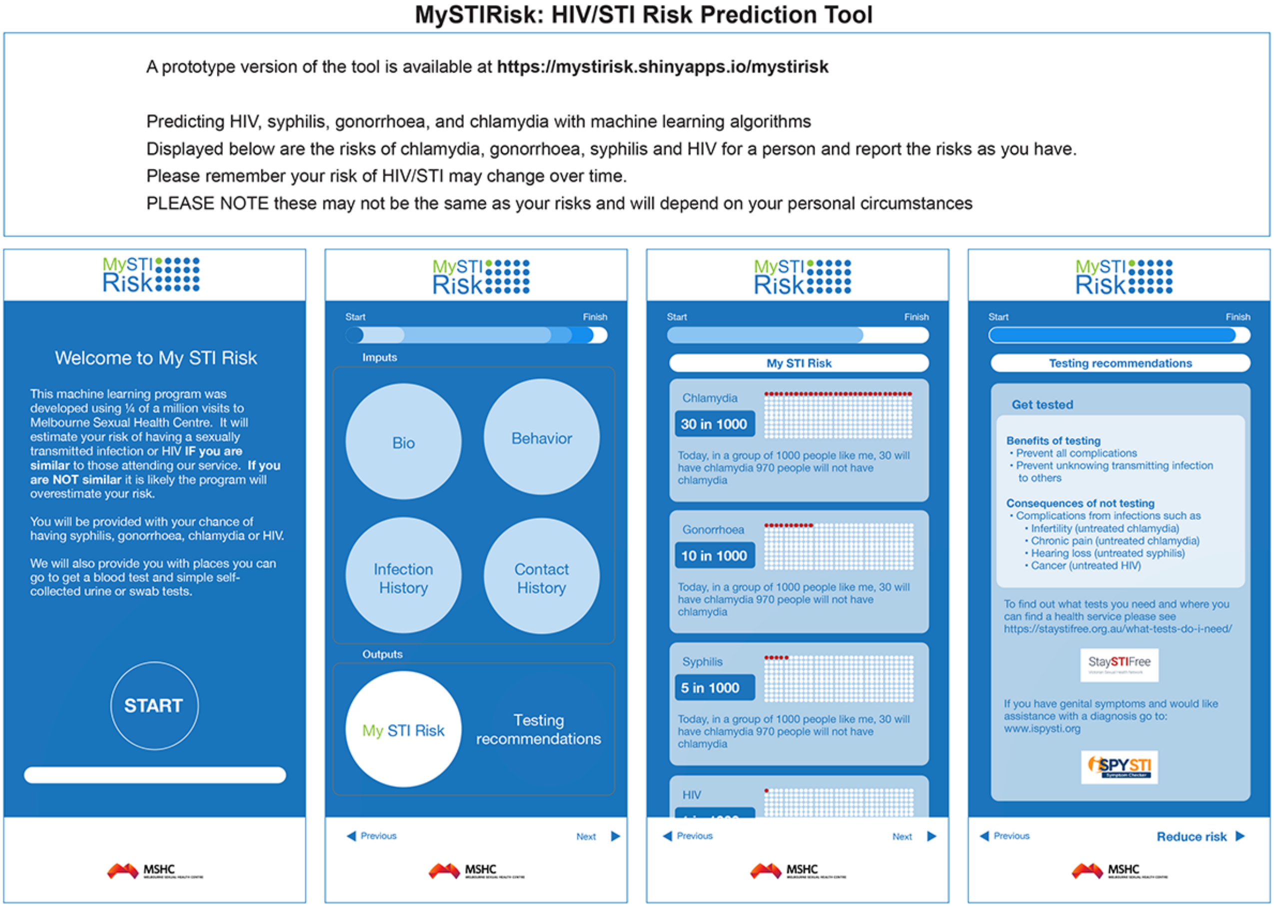
Based on the selected most important predictors and the best model (ensemble ENR+GBM+RF), we built a HIV and STI risk prediction tool, named *MySTIRisk*. We examined *MySTIRisk* and demonstrated its performance on testing to be acceptable or excellent (AUC for HIV=0.78; AUC for syphilis=0.84; AUC for gonorrhea=0.78; AUC for chlamydia=0.70), similar to its original model based on predictors. Our risk prediction tool obtained stable performance on external validation data from 2019 (AUC for HIV=0.79; AUC for syphilis=0.85; AUC for gonorrhea=0.81; AUC for chlamydia=0.69). Our risk prediction tool also achieved stable performance on external validation data from 2020-2021 (AUC for HIV=0.71; AUC for syphilis=0.84; AUC for gonorrhea=0.79; AUC for chlamydia=0.69; Figure 3 and Tables S23-S26 in Multimedia Appendix 1). Using the selected predictors, our risk prediction tool showed comparable AUCs to the best machine learning model using all predictors (Table S27 in Multimedia Appendix 1).

Figure 3. Receiver operating characteristic curve performance of the HIV and sexually transmitted infection (STI) risk prediction tool on (A) testing data analysis from 2015-2018, (B) external data validation analysis from 2019, and (C) external data validation analysis from 2020-2021. AUC: area under the curve.



To estimate the risk of HIV or an STI, we fitted the data using a logistic function to provide a fitting curve for each model-predicted probability and infection prevalence (Figures S4-S7 in Multimedia Appendix 1). Then, a prototype version of the tool was created with R Shiny [46,47] to allow for individual input and HIV and STI risk computation. A prototype version of the tool is available online [48]. The graphical user interface elements of the tool are summarized in Figure 4. The web application collects individual characteristics, processes the collected characteristics, loads the trained machine learning models, calculates a quantitative HIV and STI risk, and displays the results of the risk and recommendations. The web application's input was designed using previous successful websites or internal CASI questionnaires (60,000 entries a year) that operate at MSHC and used individual characteristic data, including demographics, sexual practices, STI history, and STI contact history. The web application's output includes HIV and STI risk prediction results and recommendations that were developed in consultation with Professor Jon Emery at the University of Melbourne, who is an expert in the communication of risk (see the Acknowledgments section). We acknowledge that this is a prototype and that further development will take place in optimizing this output for accurate risk communication.

Figure 4. Graphical user interface elements of the HIV and sexually transmitted infection (STI) risk prediction tool, called MySTIRisk. A prototype version of the tool is available at [48]. Machine learning algorithms are used to predict a person’s risk of chlamydia, gonorrhea, syphilis, and HIV.



MySTIRisk: HIV/STI Risk Prediction Tool

A prototype version of the tool is available at **https://mystirisk.shinyapps.io/mystirisk**

Predicting HIV, syphilis, gonorrhoea, and chlamydia with machine learning algorithms
Displayed below are the risks of chlamydia, gonorrhoea, syphilis and HIV for a person and report the risks as you have.
Please remember your risk of HIV/STI may change over time.
PLEASE NOTE these may not be the same as your risks and will depend on your personal circumstances

Welcome to My STI Risk

This machine learning program was developed using ¼ of a million visits to Melbourne Sexual Health Centre. It will estimate your risk of having a sexually transmitted infection or HIV **IF you are similar** to those attending our service. **If you are NOT similar** it is likely the program will overestimate your risk.

You will be provided with your chance of having syphilis, gonorrhoea, chlamydia or HIV.

We will also provide you with places you can go to get a blood test and simple self-collected urine or swab tests.

Chlamydia: 30 in 1000. Today, in a group of 1000 people like me, 30 will have chlamydia 970 people will not have chlamydia

Gonorrhoea: 10 in 1000. Today, in a group of 1000 people like me, 30 will have chlamydia 970 people will not have chlamydia

Syphilis: 5 in 1000. Today, in a group of 1000 people like me, 30 will have chlamydia 970 people will not have chlamydia

Get tested

Benefits of testing
- Prevent all complications
- Prevent unknowing transmitting infection to others

Consequences of not testing
- Complications from infections such as
  - Infertility (untreated chlamydia)
  - Chronic pain (untreated chlamydia)
  - Hearing loss (untreated syphilis)
  - Cancer (untreated HIV)

To find out what tests you need and where you can find a health service please see https://staystifree.org.au/what-tests-do-i-need/

If you have genital symptoms and would like assistance with a diagnosis go to: www.ispysti.org

These are examples of the HIV and STI risk prediction results:

> *Your HIV risk is about 2/1000. In a group of 1000 people like me, 2 will have HIV. 998 people will not have HIV.*
>
> *Your syphilis risk is about 10/1000. In a group of 1000 people like me, 10 will have syphilis. 990 people will not have syphilis.*
>
> *Your gonorrhea risk is about 30/1000. In a group of 1000 people like me, 30 will have gonorrhea. 970 people will not have gonorrhea.*
>
> *Your chlamydia risk is about 50/1000. In a group of 1000 people like me, 50 will have chlamydia. 950 people will not have chlamydia.*

The following examples describe testing recommendations:

- Benefits of testing: Prevent all complications and prevent unknowingly transmitting infection to others.
- Consequences of not testing: Complications from infections such as infertility (untreated chlamydia), chronic pain (untreated chlamydia), hearing loss (untreated syphilis), and cancer (untreated HIV).

Discussion

Principal Findings

This is the first web-based risk prediction tool based on machine learning algorithms and self-reported data to accurately identify HIV and syphilis, gonorrhea, and chlamydia infection in men and women and was stable on external validation. Our findings showed that machine learning algorithms could predict HIV and STIs in clinic attendees. Our results also showed that stacking ensemble learning algorithms perform better than individual machine learning models to predict HIV and STIs. We then developed a web-based application to provide an immediate and individualized assessment for the risk of a positive diagnosis of HIV and 3 STIs. Our application could be a part of clinic websites or digital health platforms to identify individuals with a higher risk of HIV and STIs or potential candidates for HIV pre-exposure prophylaxis (PrEP). Further validation studies in other countries can assess the usefulness of this risk prediction tool, which helps reduce HIV and STI incidence and the cost of HIV and STI screening, which requires expensive equipment and specialized expertise.

Comparison With Prior Work

Our results showed that nonlinear machine learning algorithms provided better performance than the conventional logistic regression for predicting HIV and STIs in men and women. Our findings are consistent with the results of previous machine learning predictive models for HIV and STIs [19,21,24,31]. Bao et al [21] showed that a GBM model performed better than logistic regression in MSM. Our study suggests that nonlinear machine learning models (eg, GBM, RF) could provide better performance than conventional logistic regression even without ensemble learning.

Our results showed that the stacking ensemble machine learning techniques outperform individual machine learning models. We systematically developed and tested 34 machine learning models and found that stacking ensemble learning technology outperformed individual machine learning models [18]. Previous studies have used ensemble learning models to predict an individual's HIV risk [19,25]; however, no study has looked at the risk of gonorrhea, chlamydia, or syphilis using ensemble learning models. The only study we could identify was one that had predicted the risk of a repeat STI with ensemble learning. Elder et al [31] showed that an ensemble of models could perform better for 2 or more repeat STIs within 730 days of follow-up than the individual classifiers (AUC=0.76). Our results found that stacking ensemble techniques could also be applied to enhance the performance of HIV prediction. The AUC of our ensemble HIV model (AUC=0.78, 95% CI 0.74-0.83) was higher than that in a similar study in Kenya and Uganda for HIV risk prediction (AUC=0.73, 95% CI 0.71-0.76) [25]. We also found that the combinations of more individual machine learning models do not necessarily lead to a better stacking ensemble model. For example, in our study, the stacking ensemble learning of 4 models for syphilis was not higher than a stacking ensemble learning of 3 models. We also found that a better performing stacking ensemble model always included GBM. The findings of our stacking ensemble learning strategies may have implications for future stacking ensemble learning frameworks.

Our models have several strengths compared with previous machine learning models for predicting HIV and STIs. First, our predictive models were not limited to high-risk groups (such as MSM). HIV and STI risk prediction models have been published previously but mainly for high-risk individuals, such as MSM [20,21,24,29]. Our models could predict HIV and STI acquisition in both men and women, including homosexual and heterosexual individuals. Second, our predictive models only used self-reported and simple questions to develop models. Previously published studies used numerous predictors for their models [23,26-28]. Third, we systematically developed 26 ensemble models. In our study, we tested all possible combinations of 5 base models. The final strength of our research is that we performed 2 external validation analyses of each model.

We were unable to locate any web-based, publicly available tool to quantify STI risk. We identified some available web-based HIV prediction tools, such as the “HIV risk prediction tool” [49], “HIV/AIDS Risk Calculator” [50], and “Online Risk Assessment” [51]. We also identified some available web-based STI prediction tools, such as “Find out if you need to get tested for an STD” [52], “Online STI Testing” [53], and “Take a free test” [54]. These HIV and STI prediction tools provide only subjective terms such as “high” risk or “You are advised to take an HIV/STI test.” Our risk prediction tool could quantify the risk of HIV and STIs. In addition, our artificial intelligence (AI)–based risk prediction tool can simultaneously provide risk scores for HIV and 3 common STIs (gonorrhea, chlamydia, and syphilis) for men and women aged 18 years and older.

Implications

Our web-based HIV and STI risk prediction tool can be used as a screening tool to potentially increase HIV and STI testing and encourage access to testing and health care (Figure S8 in Multimedia Appendix 1). The tool could be used on clinic websites so the public could assess their risk and then decide if they would attend a clinic for testing. It may also be used within a clinic to identify and triage those at higher risk of HIV and STIs if the demand in the clinic is too great to see everyone who attended. However, an AI-based risk prediction tool cannot replace formal HIV and STI testing and treatment in clinical settings, but it would allow individuals to understand their own risks and increase testing uptake. Our tool could increase risk perception and concern about infection, thus increasing HIV and STI testing. A study in the British population showed that increased risk perceptions are associated with greater STI health care use [12]. Further external validation of our AI-based risk prediction tool in other countries or regions, such as low- and middle-income countries, may provide an opportunity to reduce the cost of HIV and STI screening by better focusing testing on those at highest risk [55].

There are many possible ways that our web-based risk prediction tool could be potentially used, including as part of a behavioral intervention to control HIV and STIs or to help clinicians or public health workers identify high-risk individuals for risk management or further interventions. An example of this exists in adolescent health risk behaviors. Researchers used an individual’s risk behavior scores and personalized feedback as part of an intervention for health behaviors, including nutritional behaviors, physical activity, and sleep [56]. In this randomized clinical trial, the youths in the intervention group significantly reduced their risk behavior scores at 3 months compared with the control group [56]. Our web-based risk prediction tool could serve as a behavioral intervention tool in the same way.

Future work will investigate the effectiveness of this web-based HIV and STI risk prediction tool for behavioral change (ie, uptake of PrEP or condom promotion) and STI service utilization behaviors (timely clinic attendance and HIV and STI testing uptake) after receiving risk prediction results and testing recommendations. Implementing this web-based HIV and STI prediction tool may encourage individuals with STI symptoms or those at high risk without symptoms to attend health services for timely testing and regular testing. Since February 2009, the MSHC has offered MSM regular SMS reminders for STI screening [57]. For example, providing an estimated risk of HIV and STIs and risk reduction advice (ie, uptake of PrEP or condom promotion) among high-risk populations (eg, MSM) in an SMS reminder message may encourage testing and behavioral changes.

Limitations

This study has some limitations. First, the predictive factors depend on self-reported information from the CASI system, which is subject to the participants' recall, nonresponse, and social desirability bias. For example, MSM who declined to report the number of male partners were at a higher risk of chlamydia [58]. There has been substantial work undertaken on the CASI system's validity and accuracy [59]. Second, machine learning models may suffer from overfitting. We used repeated CV to tackle the overfitting problem. We also used ensemble learning methods to enhance the model's generalizability. Third, the generalizability of our models to those not attending the clinic or to other countries or regions is limited because it was derived from a single sexual health service. Thus, if it is used in other countries and regions, further validation is required. Finally, the risks of HIV have changed rapidly over this time by introducing PrEP, so future models will need to include this question, given how the potency of this single preventive strategy.

Conclusions

This is the first web-based risk assessment tool using machine learning algorithms and self-reported data to identify HIV, syphilis, gonorrhea, and chlamydia in men and women. Our online risk prediction tool could accurately predict the risk of HIV and STIs in clinic attendees with a simple self-administered questionnaire. Our risk prediction tool could be part of clinic websites or digital health platforms. The public can use this risk prediction tool to assess their HIV and STI risk to inform testing. Clinicians or public health workers can use this risk prediction tool to identify high-risk individuals for further interventions.

Acknowledgments

EC and JJO are supported by Australian National Health and Medical Research Council Emerging Leadership Investigator Grants (GNT1172873 and GNT1193955, respectively). CKF is supported by an Australian National Health and Medical Research Council Leadership Investigator Grant (GNT1172900). LZ is supported by the National Natural Science Foundation of China (Grant number: 81950410639); Outstanding Young Scholars Support Program (Grant number: 3111500001); Xi’an Jiaotong University Basic Research and Profession Grant (Grant number: xtr022019003, xzy032020032); Epidemiology modeling and risk assessment (Grant number: 20200344); and Xi’an Jiaotong University Young Scholar Support Grant (Grant number: YX6J004). The authors want to acknowledge Afrizal Afrizal from the Melbourne Sexual Health Centre (MSHC) for data extraction. The authors thank Glenda Fehler for her contribution to data cleaning. The authors also would like to acknowledge Jon Emery from the University of Melbourne for an insightful discussion on risk prediction tools (eg, Figure 4). We thank Mark Chung at the MSHC for his assistance in preparing Figure 4.

Authors' Contributions

XX, CKF, and LZ conceived and designed the study. XX cleaned the data, established the models and coding, wrote the first draft, and edited the manuscript. WL, EC, CKF, and LZ contributed to data cleaning. XX, ZG, ZY, YB, and LZ contributed to establishing the models and coding. JW and XX developed the web-based application. CKF and LZ contributed to establishing the web-based application. EC, CKF, and LZ contributed to data verification and supervision. EC, YB, ZY, ZG, JJO, WL, CKF, and LZ contributed to the interpretation of data and manuscript revision. All authors contributed to the preparation of the manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 728 KB - jmir_v24i8e37850_app1.docx]

References

1. Ramchandani MS, Golden MR. Confronting rising STIs in the era of PrEP and treatment as prevention. Curr HIV/AIDS Rep 2019 Jun;16(3):244-256 [FREE Full text] [doi: 10.1007/s11904-019-00446-5] [Medline: 31183609]
2. Chow EPF, Grulich AE, Fairley CK. Epidemiology and prevention of sexually transmitted infections in men who have sex with men at risk of HIV. Lancet HIV 2019 Jun;6(6):e396-e405. [doi: 10.1016/S2352-3018(19)30043-8] [Medline: 31006612]
3. Report on global sexually transmitted infection surveillance 2018. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/277258/9789241565691-eng.pdf [accessed 2019-05-04]
4. HIV, viral hepatitis and sexually transmissible infections in Australia Annual surveillance report 2021. Kirby Institute. 2021. URL: https://kirby.unsw.edu.au/sites/default/files/kirby/report/Annual-Suveillance-Report-2021_HIV.pdf [accessed 2022-04-06]

5. HIV, viral hepatitis and sexually transmissible infections in Australia: Annual surveillance report 2018. Kirby Institute. 2018. URL: https://kirby.unsw.edu.au/report/hiv-viral-hepatitis-and-sexually-transmissible-infections-australia-annual-surveillance [accessed 2019-05-08]
6. Mayaud P, Mabey D. Approaches to the control of sexually transmitted infections in developing countries: old problems and modern challenges. Sex Transm Infect 2004 Jun 01;80(3):174-182 [FREE Full text] [doi: 10.1136/sti.2002.004101] [Medline: 15169997]
7. Global health sector strategy on Sexually Transmitted Infections, 2016-2021. World Health Organization. 2016 Oct 03. URL: https://www.who.int/publications/i/item/WHO-RHR-16.09 [accessed 2021-04-13]
8. UNAIDS data 2018. UNAIDS. 2018. URL: https://www.unaids.org/sites/default/files/media_asset/unaids-data-2018_en.pdf [accessed 2021-07-12]
9. Levy SB, Gunta J, Edemekong P. Screening for sexually transmitted diseases. Prim Care 2019 Mar;46(1):157-173. [doi: 10.1016/j.pop.2018.10.013] [Medline: 30704656]
10. Vermund SH, Wilson CM. Barriers to HIV testing-where next? The Lancet 2002 Oct;360(9341):1186-1187. [doi: 10.1016/s0140-6736(02)11291-8]
11. Collins IM, Bickerstaffe A, Ranaweera T, Maddumarachchi S, Keogh L, Emery J, et al. iPrevent®: a tailored, web-based, decision support tool for breast cancer risk assessment and management. Breast Cancer Res Treat 2016 Feb;156(1):171-182 [FREE Full text] [doi: 10.1007/s10549-016-3726-y] [Medline: 26909793]
12. Clifton S, Mercer CH, Sonnenberg P, Tanton C, Field N, Gravningen K, et al. STI risk perception in the British population and how it relates to sexual behaviour and STI healthcare use: findings from a cross-sectional survey (Natsal-3). EClinicalMedicine 2018 Aug;2-3:29-36 [FREE Full text] [doi: 10.1016/j.eclinm.2018.08.001] [Medline: 30320305]
13. Ooi CY, Ng CJ, Sales AE, Lim HM. Implementation strategies for web-based apps for screening: scoping review. J Med Internet Res 2020 Jul 20;22(7):e15591 [FREE Full text] [doi: 10.2196/15591] [Medline: 32706655]
14. Eftekhar B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Med Inform Decis Mak 2005 Feb 15;5:3 [FREE Full text] [doi: 10.1186/1472-6947-5-3] [Medline: 15713231]
15. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina (Kaunas) 2020 Sep 08;56(9):1 [FREE Full text] [doi: 10.3390/medicina56090455] [Medline: 32911665]
16. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods 2018 Apr 3;15(4):233-234 [FREE Full text] [doi: 10.1038/nmeth.4642] [Medline: 30100822]
17. Garg R, Dong S, Shah S, Jonnalagadda S. A bootstrap machine learning approach to identify rare disease patients from electronic health records. arXiv 2016 Sep 06:1-8.
18. Latha CBC, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked 2019;16:100203. [doi: 10.1016/j.imu.2019.100203]
19. Xu X, Ge Z, Chow EPF, Yu Z, Lee D, Wu J, et al. A machine-learning-based risk-prediction tool for HIV and sexually transmitted infections acquisition over the next 12 months. J Clin Med 2022 Mar 25;11(7):1818 [FREE Full text] [doi: 10.3390/jcm11071818] [Medline: 35407428]
20. Dong Y, Liu S, Xia D, Xu C, Yu X, Chen H, et al. Prediction model for the risk of HIV infection among MSM in China: validation and stability. Int J Environ Res Public Health 2022 Jan 17;19(2):1010 [FREE Full text] [doi: 10.3390/ijerph19021010] [Medline: 35055826]
21. Bao Y, Medland NA, Fairley CK, Wu J, Shang X, Chow EPF, et al. Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using machine learning approaches. J Infect 2021 Jan;82(1):48-59. [doi: 10.1016/j.jinf.2020.11.007] [Medline: 33189772]
22. Turbé V, Herbst C, Mngomezulu T, Meshkinfamfard S, Dlamini N, Mhlongo T, et al. Deep learning of HIV field-based rapid tests. Nat Med 2021 Jul;27(7):1165-1170 [FREE Full text] [doi: 10.1038/s41591-021-01384-9] [Medline: 34140702]
23. Duthe J, Bouzille G, Sylvestre E, Chazard E, Arvieux C, Cuggia M. How to identify potential candidates for HIV pre-exposure prophylaxis: an AI algorithm reusing real-world hospital data. Stud Health Technol Inform 2021 May 27;281:714-718. [doi: 10.3233/SHTI210265] [Medline: 34042669]
24. Xiang Y, Fujimoto K, Li F, Wang Q, Del Vecchio N, Schneider J, et al. Identifying influential neighbors in social networks and venue affiliations among young MSM: a data science approach to predict HIV infection. AIDS 2021 May 01;35(Suppl 1):S65-S73 [FREE Full text] [doi: 10.1097/QAD.0000000000002784] [Medline: 33306549]
25. Balzer LB, Havlir DV, Kamya MR, Chamie G, Charlebois ED, Clark TD, et al. Machine learning to identify persons at high-risk of human immunodeficiency virus acquisition in rural Kenya and Uganda. Clin Infect Dis 2020 Dec 03;71(9):2326-2333 [FREE Full text] [doi: 10.1093/cid/ciz1096] [Medline: 31697383]
26. Gruber S, Krakower D, Menchaca J, Hsu K, Hawrusik R, Maro J, et al. Using electronic health records to identify candidates for human immunodeficiency virus pre-exposure prophylaxis: An application of super learning to risk prediction when the outcome is rare. Stat Med 2020 Oct 15;39(23):3059-3073 [FREE Full text] [doi: 10.1002/sim.8591] [Medline: 32578905]

27. Marcus JL, Hurley LB, Krakower DS, Alexeeff S, Silverberg MJ, Volk JE. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. The Lancet HIV 2019 Oct;6(10):e688-e695. [doi: 10.1016/s2352-3018(19)30137-7]
28. Krakower DS, Gruber S, Hsu K, Menchaca JT, Maro JC, Kruskal BA, et al. Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: a modelling study. Lancet HIV 2019 Oct;6(10):e696-e704 [FREE Full text] [doi: 10.1016/S2352-3018(19)30139-0] [Medline: 31285182]
29. Xiang Y, Fujimoto K, Schneider J, Jia Y, Zhi D, Tao C. Network context matters: graph convolutional network model over social networks improves the detection of unknown HIV infections among young men who have sex with men. J Am Med Inform Assoc 2019 Nov 01;26(11):1263-1271 [FREE Full text] [doi: 10.1093/jamia/ocz070] [Medline: 31197365]
30. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using clinical notes and natural language processing for automated HIV risk assessment. J Acquir Immune Defic Syndr 2018 Feb 01;77(2):160-166 [FREE Full text] [doi: 10.1097/QAI.0000000000001580] [Medline: 29084046]
31. Elder HR, Gruber S, Willis SJ, Cocoros N, Callahan M, Flagg EW, et al. Can machine learning help identify patients at risk for recurrent sexually transmitted infections? Sex Transm Dis 2021 Jan;48(1):56-62. [doi: 10.1097/OLQ.0000000000001264] [Medline: 32810028]
32. Vodstrcil LA, Fairley CK, Williamson DA, Bradshaw CS, Chen MY, Chow EPF. Immunity to hepatitis A among men who have sex with men attending a large sexual health clinic in Melbourne, Australia, 2012-2018. Sex Transm Infect 2020 Jun;96(4):265-270. [doi: 10.1136/sextrans-2019-054327] [Medline: 32169881]
33. Misson J, Chow EPF, Chen MY, Read TRH, Bradshaw CS, Fairley CK. Trends in gonorrhoea infection and overseas sexual contacts among females attending a sexual health centre in Melbourne, Australia, 2008-2015. Commun Dis Intell (2018) 2018;42:1 [FREE Full text] [Medline: 30626294]
34. Chow E, Hocking J, Ong J, Phillips T, Fairley C. Sexually transmitted infection diagnoses and access to a sexual health service before and after the national lockdown for COVID-19 in Melbourne, Australia. Open Forum Infect Dis 2021 Jan;8(1):ofaa536 [FREE Full text] [doi: 10.1093/ofid/ofaa536] [Medline: 33506064]
35. Feng J, Xu Y, Jiang Y, Zhou ZH. Soft gradient boosting machine. arXiv 2020 Jun 07:1-16 [FREE Full text]
36. Caron M, Allard R, Bédard L, Latreille J, Buckeridge DL. Enteric disease episodes and the risk of acquiring a future sexually transmitted infection: a prediction model in Montreal residents. J Am Med Inform Assoc 2016 Nov;23(6):1159-1165. [doi: 10.1093/jamia/ocw026] [Medline: 27026613]
37. Powers KA, Price MA, Karita E, Kamali A, Kilembe W, Allen S, et al. Prediction of extended high viremia among newly HIV-1-infected persons in sub-Saharan Africa. PLoS One 2018;13(4):e0192785 [FREE Full text] [doi: 10.1371/journal.pone.0192785] [Medline: 29614069]
38. Na K. Prediction of future cognitive impairment among the community elderly: A machine-learning based approach. Sci Rep 2019 Mar 04;9(1):3335 [FREE Full text] [doi: 10.1038/s41598-019-39478-7] [Medline: 30833698]
39. Rigatti SJ. Random forest. J Insur Med 2017;47(1):31-39. [doi: 10.17849/insm-47-01-31-39.1] [Medline: 28836909]
40. Venkata Ramana B, Babu MP, Venkateswarlu N. A critical study of selected classification algorithms for liver disease diagnosis. IJDMS 2011 May 31;3(2):101-114. [doi: 10.5121/ijdms.2011.3207]
41. Jang H, Cho K. Applications of deep learning for the analysis of medical data. Arch Pharm Res 2019 Jun;42(6):492-504. [doi: 10.1007/s12272-019-01162-9] [Medline: 31140082]
42. Pari R, Sandhya M, Sankar S. A multitier stacked ensemble algorithm for improving classification accuracy. Comput. Sci. Eng 2020 Jul 1;22(4):74-85. [doi: 10.1109/mcse.2018.2873940]
43. Menardi G, Torelli N. Training and assessing classification rules with imbalanced data. Data Min Knowl Disc 2012 Oct 30;28(1):92-122. [doi: 10.1007/s10618-012-0295-5]
44. Shehzad A, Rockwood K, Stanley J, Dunn T, Howlett SE. Use of Patient-Reported Symptoms from an Online Symptom Tracking Tool for Dementia Severity Staging: Development and Validation of a Machine Learning Approach. J Med Internet Res 2020 Nov 11;22(11):e20840 [FREE Full text] [doi: 10.2196/20840] [Medline: 33174853]
45. Mandrekar J. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology 2010 Sep;5(9):1315-1316 [FREE Full text] [doi: 10.1097/jto.0b013e3181ec173d]
46. Beeley C. Web Application Development with R Using Shiny: Build stunning graphics and interactive data visualizations to deliver cutting-edge analytics. Birmingham, UK: Packt Publishing Ltd; 2016.
47. Gregorich M, Heinzel A, Kammer M, Meiselbach H, Böger C, Eckardt K, et al. A prediction model for the decline in renal function in people with type 2 diabetes mellitus: study protocol. Diagn Progn Res 2021 Nov 18;5(1):19 [FREE Full text] [doi: 10.1186/s41512-021-00107-5] [Medline: 34789343]
48. MySTIRisk. URL: https://mystirisk.shinyapps.io/mystirisk [accessed 2022-03-08]
49. HIV Risk Reduction Tool. Centers for Disease Control and Prevention. URL: https://hivrisk.cdc.gov/risk-estimator-tool/#-sb [accessed 2022-04-06]
50. HIV/AIDS Risk Calculator. URL: https://www.medindia.net/patients/calculators/hiv-risk-calculator.asp [accessed 2022-04-06]
51. Online Risk Assessment. URL: https://aidsconcern.org.hk/en/testing-service/assess/
52. Find out if you need to get tested for an STD. URL: https://stdwizard.com/#/home
53. Online STI Testing. URL: https://www.getthefacts.health.wa.gov.au/online-sti-testing

54. Take a free test. URL: https://www.couldihaveit.com.au/Take-a-free-test
55. Alami H, Rivard L, Lehoux P, Hoffman SJ, Cadeddu SBM, Savoldelli M, et al. Artificial intelligence in health care: laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. Global Health 2020 Jun 24;16(1):52 [FREE Full text] [doi: 10.1186/s12992-020-00584-1] [Medline: 32580741]
56. Richardson LP, Zhou C, Gersh E, Spielvogle H, Taylor JA, McCarty CA. Effect of electronic screening with personalized feedback on adolescent health risk behaviors in a primary care setting: a randomized clinical trial. JAMA Netw Open 2019 May 03;2(5):e193581 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.3581] [Medline: 31074815]
57. Zou H, Fairley CK, Guy R, Bilardi J, Bradshaw CS, Garland SM, et al. Automated, computer generated reminders and increased detection of gonorrhoea, chlamydia and syphilis in men who have sex with men. PLoS One 2013;8(4):e61972 [FREE Full text] [doi: 10.1371/journal.pone.0061972] [Medline: 23613989]
58. Chow E, Carlin J, Read T, Chen M, Bradshaw C, Sze J, et al. Factors associated with declining to report the number of sexual partners using computer-assisted self-interviewing: a cross-sectional study among individuals attending a sexual health centre in Melbourne, Australia. Sex. Health 2018;15(4):350 [FREE Full text] [doi: 10.1071/sh18024]
59. Fairley CK, Sze JK, Vodstrcil LA, Chen MY. Computer-assisted self interviewing in sexual health clinics. Sex Transm Dis 2010 Nov;37(11):665-668. [doi: 10.1097/OLQ.0b013e3181f7d505] [Medline: 20975481]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
CASI: computer-assisted self-interview system
CV: cross-validation
DL: deep learning
EHR: electronic health records
ENR: elastic net regression
GBM: gradient boosting machine
LASSO: least absolute shrinkage and selection operator
MSHC: Melbourne Sexual Health Centre
MSM: men who have sex with men
NAAT: nucleic amplification tests
NB: Naive Bayes
PrEP: pre-exposure prophylaxis
RF: random forest
RR: ridge regression
STI: sexually transmitted infection
WHO: World Health Organization

Edited by R Kukafka; submitted 09.03.22; peer-reviewed by X Zou, X Ma, S El kefi; comments to author 01.04.22; revised version received 13.04.22; accepted 28.07.22; published 25.08.22.

Please cite as:
Xu X, Yu Z, Ge Z, Chow EPF, Bao Y, Ong JJ, Li W, Wu J, Fairley CK, Zhang L
Web-Based Risk Prediction Tool for an Individual's Risk of HIV and Sexually Transmitted Infections Using Machine Learning Algorithms: Development and External Validation Study
J Med Internet Res 2022;24(8):e37850
URL: https://www.jmir.org/2022/8/e37850
doi:10.2196/37850
PMID:36006685

Original Paper

Investigating a Work System Approach to Implement an Emergency Department Surge Management System: Case Study

Jennifer Jewer[1,2], BCom, BEd, MASc, PhD

[1]Faculty of Business Administration, Memorial University of Newfoundland, St. John's, NL, Canada

[2]Cross-appointed, Discipline of Emergency Medicine, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

Corresponding Author:
Jennifer Jewer, BCom, BEd, MASc, PhD
Faculty of Business Administration
Memorial University of Newfoundland
300 Prince Philip Drive
St. John's, NL, A1B 3X5
Canada
Phone: 1 7098643094
Email: jenniferj@mun.ca

Abstract

Background: Emergency department (ED) crowding is a global health care issue. eHealth systems have the potential to reduce crowding; however, the true benefits are seldom realized because the systems are not integrated into clinicians' work. We sought a deep understanding of how an eHealth system implementation can be structured to truly integrate the system into the workflow.

Objective: The specific objectives of this study were to examine whether work system theory (WST) is a good approach to structure the implementation of an eHealth system by incorporating the entire work system, and not just the eHealth system, in the implementation framework; identify the role that specific elements of WST's static framework and dynamic work system life cycle model play in the implementation; and demonstrate how WST can be applied in the health care setting to guide the implementation of an eHealth system.

Methods: Through a case study of an ED in a rural hospital, we used a mixed methods approach to examine the implementation of a surge management system through the lens of WST. We conducted 14 hours of observation in the ED; 20 interviews with clinicians, management, and members of the implementation team; and a survey of 23 clinicians; reviewed related documentation; and analyzed ED data to measure wait times. We used template analysis based on WST to structure our analysis of qualitative data and descriptive statistics for quantitative data.

Results: The surge management system helped to reduce crowding in the ED, staff was satisfied with the implementation, and wait time improvements have been maintained for several years. Although study participants indicated changes to their workflow, 72% (13/18) of survey participants were satisfied with their use of the system, and 82% (14/17) indicated that it was integrated with their workflow. Examining the implementation through the lens of WST enabled us to identify the aspects of the implementation that made it so successful. By applying the WST static framework, we saw how the implementation team incorporated the elements of the ED work system, assessed their alignment, and designed interventions to address areas of misalignment. The dynamic work system life cycle model captured how planned and unplanned changes were managed throughout the iterative implementation cycle—83% (15/18) of participants indicated that there was sufficient management support for the changes and 80% (16/20) indicated the change served an important purpose.

Conclusions: The broad scope and holistic approach of WST is well suited to guide eHealth system implementations as it focuses efforts on the entire work system and not just the IT artifact. We broaden the focus of WST by applying it to the implementation of an ED surge management system. These findings will guide further studies and implementations of eHealth systems using WST.

(J Med Internet Res 2022;24(8):e37472) doi:10.2196/37472

KEYWORDS

emergency department; surge management; work system; system implementation; emergency department information system; mobile phone

Introduction

Background

Emergency department (ED) crowding is a major global health care issue [1]. The negative consequences are well established and include adverse patient outcomes and increased mortality [1,2]. ED crowding describes a situation in which the demand for emergency services exceeds the ability to provide care in a reasonable amount of time. When an ED has reached the point of overcapacity, the hospital implements a process, called surge, to allow for decompression. The use of eHealth systems, including systems to manage surge, has the potential to offer numerous benefits for EDs; however, the benefits have often been less than anticipated in many cases owing to implementation difficulties [2]. In particular, the difficulty of integrating the system into clinicians' work is cited as a key barrier to the successful implementation of ED [3] and other eHealth systems [4,5]. Several systematic reviews of eHealth interventions found that workflow was one of the most common barriers to successful implementations [6-8]. Granja et al [6] recommended that there is a critical need to perform in-depth studies of the workflow when implementing eHealth interventions to identify facilitators and barriers at the earliest possible stage of the implementation to ensure that they are defined in the implementation strategy. This highlights the need for those undertaking such projects to understand the factors that affect the staff's work and workflow, so that they can modify and improve the implementation to align with the organization's and staff's requirements. However, we lack a thorough understanding of how the implementation of such systems can be structured to incorporate the broad work system.

Objectives

We propose that using work system theory (WST) [9,10] to guide the implementation of eHealth systems may help to attain desired outcomes by incorporating the entire work system into the implementation framework. To evaluate whether WST will be a good approach, we examined the implementation of a surge management system in an ED of a rural, Canadian, 80-bed hospital through the lens of WST.

The surge management system was designed to track patient demand and capacity in the ED, calculate surge levels, and prescribe volume-based staffing. One of the most important eHealth systems in the health care domain is the ED information system to manage information and workflow and support patient care in the emergency room, and there are numerous studies supporting the advantages of its use [2]. However, although surge management is often a component of many ED information systems and there are a plethora of studies examining many processes to manage surge (ie, lean management [11], small cycles of process changes [12], and implementation of fast-tracking [13]), there are few studies examining eHealth systems specifically for the management of surge in the ED.

WST is a well-established theory for understanding relationships between technology and work systems. A work system is "a system in which human participants and/or machines perform work (processes and activities) using information, technology, and other resources to produce specific products/services for specific internal and/or external customers" (p75) [9]. WST is based on the premise that systems, and the work processes they affect must be properly managed to fit with practice. The adoptive entity is not only the system but also the entire IT-enabled work system. WST views the work system from two perspectives: (1) a static framework, with (2) a dynamic life cycle. The static framework presents a view of the work system at a particular time interval through 9 elements constituting the work system's form, function, and environment. According to WST, the 6 internal elements of the static framework—processes and activities, participants, information, technologies, customers, products, and services—should be balanced. The remaining 3 external elements—environment, infrastructure, and strategies—provide the context in which the work system operates. The dynamic work system life cycle (WSLC) model, presents how the work system changes over time through planned and emergent (unplanned) changes. In this study, we used a mixed methods approach to examine the implementation of the surge management system from a WST perspective and conceptualize the eHealth system as only a component within a broad ecosystem. We propose that WST can be useful for structuring the implementation of such eHealth systems.

Methods

Overview

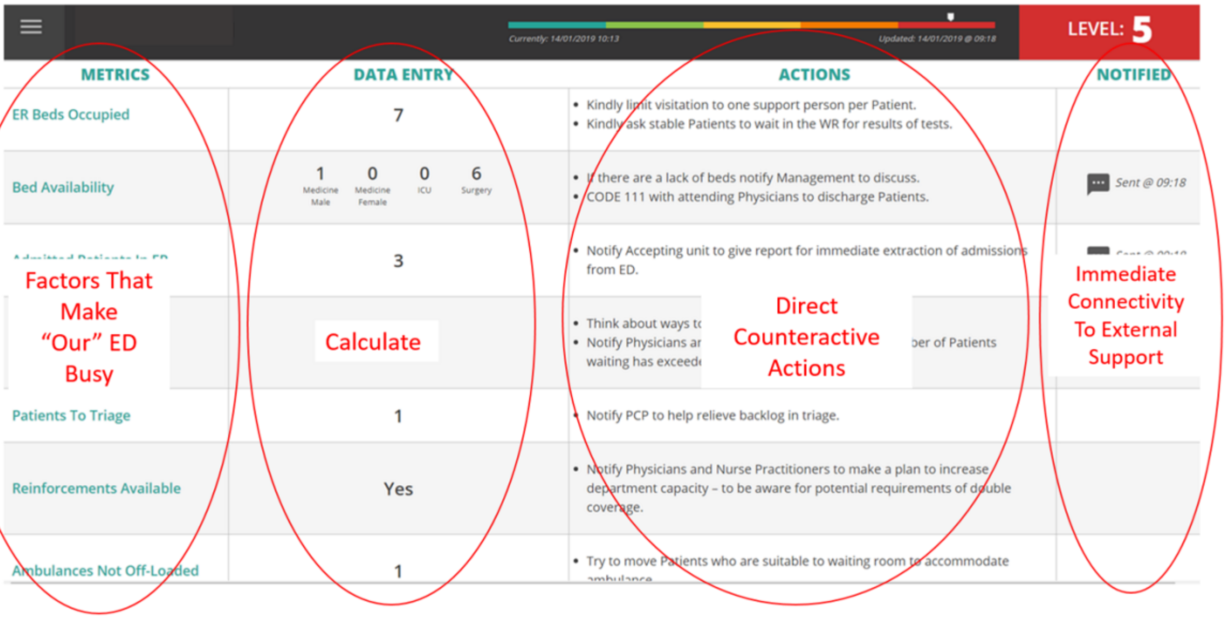
We followed a longitudinal case study methodology over approximately 2 years using a mixed methods approach to collect data during and after the implementation of the surge management system. The ED is located in a rural hospital that provides emergency and inpatient services to a catchment population of approximately 40,000. At any point of time, the ED has 8 stretchers, 3 high-turnover examination beds, and average daily volume of approximately 80 patient visits. The ED team is made up of 1 primary family physician who practices emergency medicine, 1 secondary coverage physician, 1 nurse practitioner, and up to 4 registered nurses per shift. In total, the ED has approximately 50 staff and management directly involved in surge management.

Surge Management System—Background

The surge management system is composed of the eHealth system and related interventions to manage surge levels in the ED (ie, new surge management processes and procedures and new personnel). When we refer to the surge management system in this study, we are not only referring to the eHealth system portion but also to the other components of the new surge management process. This reflects the view of WST that technology is only an element of the work system. The surge management system was implemented iteratively over approximately 5 years, and our study examines the last 2 years of the implementation, and in particular, the implementation of the eHealth system portion of the surge management system. The implementation team did not follow a formal implementation framework; instead, there was a unique situation in which the surge management system was developed by 2 ED clinicians, and they led the implementation.

The eHealth system portion of the surge management system is installed in the ED nursing station, and staff receives notifications of surge levels through a smartphone app via email and SMS text messages (Figure 1). Patient demand and capacity are entered at least every 2 hours for identified metrics that influence waiting times in the ED. Data are entered more frequently if the staff perceives the patient demand to be increasing. A surge level score is calculated in real time using algorithms. Each score has a corresponding set of prescribed volume-based staffing, management, and overcapacity protocols. For example, a total score of 40 triggers the highest level (level 5) and is associated with actions such as sending a text to frontline management and staff, calling 'Surge Level 5,' sending all patients with low acuity to the waiting room, and contacting physicians with potential discharges.

Figure 1. Screenshot of the surge management system.



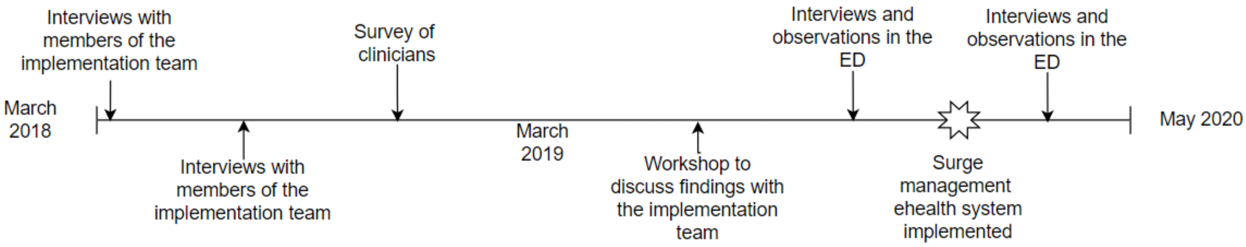
METRICS	DATA ENTRY	ACTIONS	NOTIFIED
ER Beds Occupied	7	• Kindly limit visitation to one support person per Patient. • Kindly ask stable Patients to wait in the WR for results of tests.	
Bed Availability	1 Medicine Male 0 Medicine Female 0 ICU 6 Surgery	• If there are a lack of beds notify Management to discuss. • CODE 111 with attending Physicians to discharge Patients.	Sent @ 09:18
	3	• Notify Accepting unit to give report for immediate extraction of admissions from ED.	
		• Think about ways t • Notify Physicians ar ber of Patients waiting has exceed	
Patients To Triage	1	• Notify PCP to help relieve backlog in triage.	
Reinforcements Available	Yes	• Notify Physicians and Nurse Practitioners to make a plan to increase department capacity – to be aware for potential requirements of double coverage.	
Ambulances Not Off-Loaded	1	• Try to move Patients who are suitable to waiting room to accommodate	

Data Collection and Analysis

We started our primary data collection several months after most of the manual changes to the surge management process were in place and as the team was beginning the implementation of the eHealth system portion. At this time, the clinicians used a manual version of the surge management system, in which they calculated surge levels by hand and manually reported the surge levels. Figure 2 shows a time line of our data collection.

We started by conducting open-ended interviews with 2 key members of the implementation team over a period of several months. These interviewees also worked in the ED and were key participants in the surge management process. This gave us background information about the decision to implement the surge management system, overview of the hospital and ED functioning, and insight into the ongoing and planned implementation processes.

Figure 2. Time line of data collection. ED: emergency department.



After 4 months, we distributed a web-based survey to the clinical staff in the ED to determine their satisfaction with the surge management system, the perceived benefits to the hospital, and the nature of their job level change. We also asked about their perceptions of fairness of the change, management support for the change, and their commitment to the change. Validated measures were used and adapted for our setting where necessary [14-18]. The survey was completed by 23 clinicians who play a direct role in surge management at the hospital. Though the sample size for the surveys appears small (n=23), it represents approximately 50% of the population of potential users. Refer to Multimedia Appendix 1 for the survey questions. To cope with the statistical limitations of the small survey sample size, we confined our quantitative analyses to descriptive statistics. We conducted a workshop with members of the implementation team to discuss the initial observations and to receive feedback, thereby increasing internal validity [19].

Then, we used the findings from the initial interviews and survey to inform our observations in the ED and to probe areas further through additional interviews approximately 9 months later and again after 4 months once all aspects of the eHealth system implementation were complete. We conducted 14 hours of observation over 2 days (1 day during the implementation process and 1 day after the eHealth portion of the surge management system, and all related changes were implemented). Each day included several clinical shifts so that we could see how the process functioned at various times throughout the day and at various surge levels. We had full access to the ED and our observations involved observing the environment, staff interactions and events, activities, and processes occurring in the hospital, which are associated with the use of the surge management system. We took notes on what was observed and asked questions to understand what was happening. Refer to Multimedia Appendix 1 for the observation protocol. During this time, we also conducted 16 interviews with frontline clinicians and management, and members of the implementation team were included in this group. Refer to Multimedia Appendix 1 for the interview guide. We stopped our phases of observations and interviews when data saturation was reached. All notes and recordings of interviews and observations were transcribed, and all sources were anonymized. We reviewed documentation on the system and the implementation, including documentation on the surge management system and the surge protocol, implementation plans, business case for the surge management system, and specification of outcome measures. Table 1 provides an overview of the data collection.

Table 1. Data collection overview.

Data source	Details	Respondent characteristics
Surveys	A total of 23 respondents	• A total of 13 internal support (n=4, 31% primary RNs[a]; n=1, 8% triage RN; n=5, 38% primary care paramedics; and n=3, 23% ED[b] physicians) and 10 external support (n=1, 10% admitting physician; n=3, 30% RN on inpatient unit; and n=6, 60% other)
Observations	A total of 14 hours of observation	• Observed the ED over 2 days and several clinical shifts—1 day during the system implementation and 1 day after the implementation. • Observed clinicians using and interacting with the system and managing flow in the ED. Full access was available wherever required.
Interviews	A total of 20 interviews	• Of the 20 interviews, 4 (20%) were with ED physicians, 11 (55%) were with ED RNs, 2 (10%) were with nurse practitioners, 1 (5%) was with primary care paramedic, 1 (5%) was with ED manager, and 1 (5%) was with patient care facilitator (inpatient beds)
Document review	Surge management system documentation, surge protocol documentation, implementation plans, business case for the surge management system, and specification of outcome measures	N/A[c]
ED wait time data	Patient ED wait times from point of registration to patient departure from the ED, from April 1, 2017, to March 31, 2021	N/A

[a] RN: registered nurse.

[b] ED: emergency department.

[c] N/A: not applicable.

We used template analysis [20] based on the elements of WST to structure the qualitative data analysis. Template analysis forced us to take a well-structured approach to handle the data [20] and allowed us to examine the data according to the elements of WST. We started by coding the data according to the static framework. The 9 elements of the static framework adapted to the implementation of the surge management system in the ED are shown in Figure 3 [9]. The figure shows that the *participants* use *information* and *technologies* in various *activities and processes* to create *products and services* to serve their *customers* (ie, *the 6 internal elements*). The *environment, strategies*, and available *infrastructure (ie, the 3 external elements)* also influence the work system. The arrows indicate that the specific elements in the work system must be in alignment.

We started by producing a list of codes (ie, the template) for each of the 9 elements of the WST and their interactions (ie, a code for participants, a code for activities and processes, and a code for the interaction between these elements). This enabled us to identify conceptual themes and then cluster them into broader groupings. Then, we created a hierarchical organization of codes, with groups of similar codes clustered together to produce more general high-order codes. For example, separate codes relating to different groups of *activities and processes* were incorporated into high-order activities and processes code. Then, this was further subdivided into codes to capture different activities and processes, and these were further divided into factors influencing when and how different activities and processes were followed, codes capturing the challenges with performing the different activities and processes, and codes capturing the interventions that the hospital implementation team conducted to deal with the challenges that arose (ie, changing the triage process in the ED, initiating training, or adding an extra physician). As we coded the transcripts and marked them with the appropriate code, we revised the template as needed. For example, as we identified an issue that was not covered by an existing code, we added a new code; we also deleted codes if we found that there was no need to use it. This was an iterative process of reading the transcripts, assigning codes, and reviewing the coding template until we were confident that the template was sufficiently clear and comprehensive.

Next, we used template analysis to code the WSLC model to capture the iterative process through which the system was implemented from initiation to operation and maintenance to identify planned and unplanned changes and the resulting interventions. Figure 4 [9] shows the WSLC model, adapted to the implementation of the surge management system. We coded the data according to the 4 phases of the WSLC model (ie, initiation, development, implementation, and operation and maintenance), interactions between the phases, planned and emergent changes, and outcomes. Similar to the coding of the static framework, this was an iterative process involving coding and further refinement of the template until we reached a state that enabled us to capture the WSLC model. Once all the transcripts were coded according to the final templates for the static framework and the WSLC model, we reviewed the coded text to identify themes and relationships between them.

Finally, we obtained data from the ED to measure wait times before, during, and after our study period. The data contained the patient ED wait times from patient registration to patient departure from the ED. We analyzed the data from approximately 1 year before our study started to 1 year after our study was completed (April 1, 2017, to March 31, 2021). We collected data for total patient visits, time to provider initial assessment (PIA), length of stay for departed patients (LOSDep), and patients who left without being seen. Time to PIA is the elapsed time from the point a patient first registers at the ED until the designated provider (ie, physician or nurse practitioner) makes contact. LOSDep is the time interval between a patient's arrival to the ED to the time the patient physically leaves the ED. The number of patients who left without being seen is the percentage of patients who have registered at the ED and have been triaged but leave before being seen by a designated provider.

Figure 3. Static framework for the surge management system (adapted from the publication by Alter [9]). ED: emergency department; RN: registered nurse.

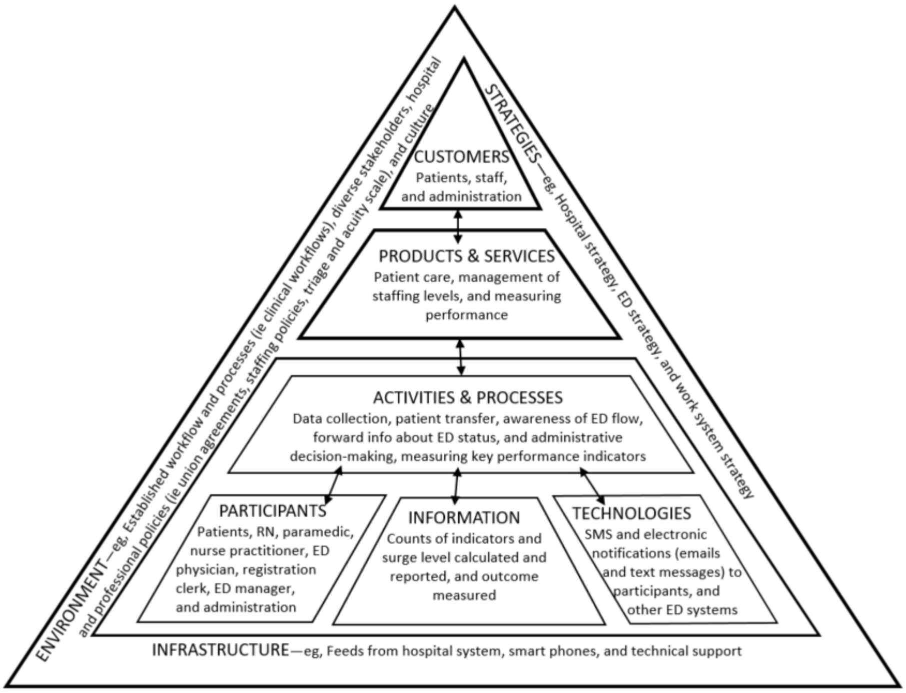
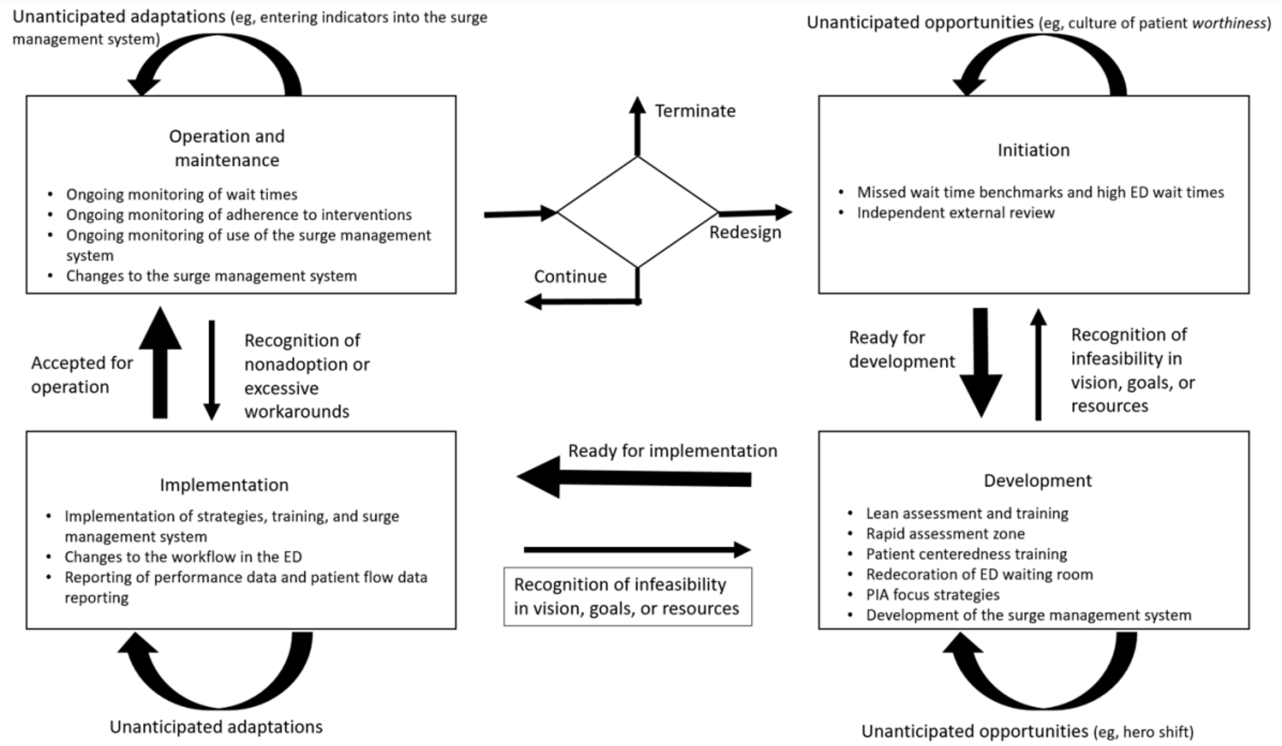


Figure 4. Work system life cycle model for the surge management system (adapted from the publication by Alter [9]). ED: emergency department; PIA: provider initial assessment.



Ethics Approval

Participants provided informed consent, and the Memorial University's Interdisciplinary Committee on Ethics in Human Research approved the study (20190669-BA).

Results

Summary

The results indicate that WST is a useful approach to structure the implementation of an eHealth system because it incorporates the entire work system, and not just the eHealth system, in the implementation framework. In this section, we identify the role that specific elements of WST's static framework and dynamic WSLC model played in the implementation and demonstrate how WST can be applied in the health care setting to guide the implementation of an eHealth system. Although the implementation team did not follow WST specifically, they included many of its elements in their approach.

Static Framework

Overview

We found that the hospital implementation team began the project with an independent external review to assess the organization and function (ie, the elements) of the ED work system. Then, they designed interventions to address the issues that were found. They did not just examine the eHealth system portion of the surge management process, rather, they recognized the radical change that they were trying to implement in the ED and examined the entire work system. A member of the implementation team commented the following during the implementation of the surge management system:

> *I guess the electronic piece [of the surge management system] is one thing, but there's also the strategy that goes along with it and we're examining that in the broader sense. And the strategy is controversial, a little bit. So if you don't have supportive buy-in from some of the frontline staff, and if it's not presented in a way that would be well explained, it could be definitely seen as something that people wouldn't want to do.*

Internal Environment

The implementation team recognized the need to focus their efforts on meeting the needs of their *customers* (ie, the staff, hospital administration, and patients). They involved the staff in the implementation. A staff member commenting on the implementation of the surge management system said the following:

> *Management driven, it's frontline driven from us. And we're like, what can we change about this process.*

The implementation team designed strategies to address their customers' (ie, staff and patient) issues. For example, time to PIA is strongly correlated with patient satisfaction; thus, they developed strategies (ie, briefly assessed patients even when formal assessment space was not immediately available or emergency physicians could triage with nursing staff in the triage room without waiting and discharge if applicable) to reduce time to PIA. As can be seen in Figure 3 [9], this is a basic component of the WST static framework, which places customers at the top of the triangle. An aspect of the implementation that could have been improved in the beginning was staff training on the overall new surge management process. As we describe in the following sections, the implementation team realized that more training was needed during the implementation for staff entering data into the surge management system; however, their approach to training focused primarily on learning the job and did not always seem to provide the staff an understanding of the overall purpose of the surge management changes or their role in the process. There were comments that the way the surge management process functioned depended on who was on shift that day, and that, in particular, physicians can have a big impact on the process. In situations such as the ED, where teamwork is so important, and as one staff member commented, "You're making decisions on the fly and you can't really plan ahead," a shared purpose and approach to surge management is needed. Furthermore, most survey respondents (11/18, 61%) indicated (agreed or strongly agreed) that they thought that the surge management system had improved their productivity and ability to coordinate continuity of care (11/18, 61%) and improved the hospital's patient care delivery, productivity, and clinical outcomes (Table 2—user satisfaction and benefits to hospital); however, in the interviews, most staff indicated that they were not aware of the real impact on patients or efficiency in the ED. They had not seen the time to PIA, LOSDep, number of patients who left without being seen, or surge levels over time; thus, they were not sure what benefits were realized. A staff member commented the following:

> *It'd be nice to see the stats and if [the surge management system] was actually related in respect to getting patients to the floor. Like door-to-doctor, or triage-to-doctor, that kind of stuff...or what's decreasing over time. That would be nice to see.*

The implementation team did not include patients, and the project did not measure patient satisfaction. Following a WST approach may have helped to ensure that they had designed the implementation with a focus on customers (ie, staff and patients), and some of these issues may have been addressed at the onset of the project.

Despite some issues with implementation, many elements of the static framework were applied. We found that rather than just implementing the eHealth system portion of the surge management system (ie, the *technology*), the implementation team recognized that there was a need to change *activities and processes*, adjust *participant* perceptions of how the ED should function, and track and share various *information* such as performance indicators (Figure 3) [9]. The implementation team was redesigning the ED's work system to address areas of misalignment identified between elements of the static framework. For example, when the implementation team recognized that waiting for beds for patient assessment, blood tests, electrocardiograms, and other minor procedures were contributing to overcrowding in the ED, they created a rapid assessment fast-track zone. When they recognized the need to decrease time to PIA, they designed changes in the workflow to reduce the time to PIA. These changes to the work processes helped to address the misalignment between the *activities and processes* and *products and services* elements and focus the work system on providing patient care and decreasing wait times. Interventions to address the misalignment between *participants* and *activitiesand processes* were addressed through lean training for frontline ED staff to encourage them to become active participants in the improvement process and redecoration of the waiting and examination rooms to have a more inviting environment for patients. Regular performance reporting facilitated alignment between *information* and *activities and processes*. The implementation team also used the information obtained from the external review and through the various interventions to design and develop a surge management system that was aligned with the activities and processes in the ED. The team recognized that the eHealth portion of the surge management system was only a component of the ED work system. They understood the impact and importance of the changes on other elements of the work system. A staff commented the following:

> *All of these different things, super track, nurse practitioner, the hero shift...some days we'd sink if we never had it.*

Table 2. Staff perceptions of the surge management system as indicated in the survey.

	Value, mean (SD; range)
User satisfaction[a]	
How satisfied are you with the surge management system?[b]	3.52 (1.08; 1-5)
The surge management system improves my productivity.	3.71 (1.10; 2-5)
The surge management system enhances my ability to coordinate continuity of care.	3.76 (1.18; 2-5)
The surge management system makes my job easier.	3.48 (1.12; 1-5)
The surge management system improves the quality of care that I can provide.	3.52 (1.12; 2-5)
The surge management system improves the quality of my decision-making.	3.48 (1.12; 1-5)
Benefits to hospital[a]	
Using the surge management system has improved patient care delivery.	3.94 (0.97; 1-5)
Using the surge management system has improved clinical outcomes.	3.82 (0.95; 2-5)
The surge management system improves our productivity.	4.12 (0.99; 1-5)
Job level change[a,c]	
I am expected to do more work than I used to.	2.78 (0.85; 2-5)
The nature of my work has changed.	3 (0.90; 2-5)
My job responsibilities have changed.	2.91 (1; 2-5)
I find greater demands placed on me at work because of this change.	2.91 (1; 2-5)
I am experiencing more pressure at work because of this change.	2.74 (1.01; 1-5)
The work processes and procedures I use have changed.	2.96 (0.88; 2-5)
My use of the surge management system is integrated with my workflow.	4.06 (0.83; 2-5)
Change fairness[a]	
Sufficient advanced notice was given to employees affected by the change.	3.96 (1.40; 1-5)
Those affected by the change had ample opportunities for input.	3.78 (1.35; 1-5)
The hospital kept everyone fully informed during the change.	3.65 (1.27; 2-5)
People affected negatively by this change were treated fairly.	3.78 (1; 2-5)
Management support for the change[a]	
Sufficient resources were available to support this change.	4 (0.85; 2-5)
All levels of management were committed to this change.	3.95 (0.84; 2-5)
Management dealt quickly and effectively with surprises during the change.	3.59 (0.91; 2-5)
There was sufficient management support for this change.	3.82 (0.96; 2-5)
Management was supportive of this change.	4.05 (0.67; 3-5)
People in this hospital find their work more interesting.	3.32 (1.13; 1-5)
Most people in this hospital are better off.	3.35 (1.19; 1-5)
People’s quality of life at work has improved.	3.36 (1.14; 2-5)
Commitment to change[a]	
This change serves an important purpose.	4.04 (0.82; 2-5)
I believe in the value of this change.	4 (0.90; 2-5)
This change is a good strategy for this organization.	4.09 (0.73; 3-5)
I think management is making a mistake by introducing this change.[c]	1.91 (0.79; 1-4)
Things would be better without this change.[c]	1.83 (0.72; 1-3)
This change is not necessary.[c]	1.78 (0.74; 1-3)

[a] Scoring: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree.

[b] Scoring: 1=extremely dissatisfied, 2=dissatisfied, 3=neither satisfied nor dissatisfied, 4=satisfied, and 5=extremely satisfied.

[c] Reverse score.

External Environment

According to WST, the 6 elements of the static framework are influenced by the 3 elements of the external environment of the hospital that affect the work system (Figure 3 [9])—the *environment*, *strategies*, and *infrastructure*. First, within the environment, the culture had a major impact on the implementation of the system. The ED staff had a history of resistance to change and certain beliefs about who should be in the ED. Commenting on staff acceptance of the changes related to the surge management system, a staff member said, "I think it was really hard to go from being very rule oriented to being more flexible." The implementation team created patient-centeredness training to address the belief system that patients with low acuity should not seek care in the ED. One of the members of the implementation team commented, "We changed a complex system of beliefs." Other factors such as established processes and professional practices also created some resistance to the new approach. A member of the implementation team commented the following:

> *We have been, since 1998 in Canada, seeing patients based on Canadian triage acuity scale. [The surge management system] kind of meddles with that a little bit.*

These initiatives appeared to bring staff on board and create a culture that was more accepting of change. In addition, a change committee was established to help communicate and identify needs for change on an ongoing basis. Second, the *strategies* at the hospital, ED, and work system were in alignment. The health authority, hospital, and ED were focused on improving patient care in the ED and supportive of innovative approaches. A manager commented the following:

> *[The health authority] has tried to put a more innovative angle in healthcare and [this hospital] been chosen as there are innovative kind of people in the frontline doing some things that can help get innovation into healthcare. Another strategy is probably improving patient satisfaction...We know that usually what happens in the ED is often reflected in how well your hospital works and how well patients are satisfied. If you don't have a good ED, you're not going to be reflected as having a good hospital. So the more that we get things improved in the ED, it tends to transform and cross over to other hospitals or areas.*

The hospital management and ED management were in support of this initiative and devoted staff and funds to support the project, and their strategies were aligned and focused on reducing ED wait times. Third, there were some issues with the infrastructure as they were not able to connect the eHealth portion of the surge management system with the hospital's other systems; therefore, staff had to continue to input the values for the indicators into the surge management system. There appeared to be some confusion regarding the responsibility for this task, and we observed that sometimes, the indicators were not entered regularly. This impaired the use of and possible benefits from the system.

Outcomes of the Implementation

The strength of this whole system approach to structuring the implementation is reflected in the success of the surge management system. The new approach to surge management changed the work system in the ED (Table 2—job level change), with comments from staff such as, "It's a whole new way of thinking." Despite these changes in workflow in the ED, we observed a strong commitment to the system (Table 2—user satisfaction). In total, 72% (13/18) of the survey participants were satisfied or extremely satisfied with their use of the system, 82% (14/17) would like to increase or significantly increase their use of the system in the future, and 82% (14/17) agreed or strongly agreed that it was integrated with their workflow.

We observed that most staff accepted the new collaborative approach to surge management (eg, "People are willing to work together and help"), and many commented that the patients seemed happy (eg, "A lot of people seem happier"). Furthermore, participants indicated that they agreed or strongly agreed that the system has improved patient care delivery (14/17, 82%), clinical outcomes (10/17, 59%), and productivity (15/17, 88%; Table 2—benefits to hospital).

The ED wait time data indicated an improvement in ED wait times at the beginning of the implementation from October 1, 2014, to March 31, 2017. Despite approximately 26% increase in patient volume, the time to PIA decreased by 62.1 minutes, LOSDep decreased by 65 minutes, and patients who left without being seen decreased from 12.1% to 4.6% [21]. Our study began approximately 1 year later when the eHealth system portion of the surge management system was being implemented. Our analysis shows that the wait times have plateaued, but stayed consistent since the time of initial implementation and through the implementation of the eHealth system portion of the surge management system, despite the increase in patient volume. Table 3 shows the characteristics of patient visits to the ED from April 1, 2017, to March 31, 2021. The dramatic decrease in ED wait times after the initial implementation of the surge management system and the ability of the ED to maintain these wait times in the 4 years after the implementation and through the implementation of the eHealth system portion of the surge management system can be seen in Figure 5. Slight variations in the time to PIA and LOSDep and large increase in the number of patients who left without being seen can be seen around the beginning of the COVID-19 pandemic in March 2020; however, since this time, the wait times have returned to their prepandemic values.

Table 3. Characteristics of patient visits to the emergency department (2013-2021).

Characteristics	July 1, 2013, to September 30, 2014[a]	January 1, 2016, to March 31, 2017[a]	April 1, 2017, to March 31, 2018	April 1, 2018, to March 31, 2019	April 1, 2019, to March 31, 2020	April 1, 2020, to March 31, 2021
Total visits, n	23,898	30,031	26,966	29,321	29,014	22,931
Number of daily visits, mean (SD)	52[b]	66[b]	74 (6.54)	80 (6.25)	79 (8.03)	63 (10.05)
Time to PIA[c] (minutes), mean (SD)	104.3 (0.9)	42.2 (8.1)	49.6 (5.5)	48.5 (4)	51.9 (5.7)	41.3 (5.5)
LOSDep[d] (minutes), mean (SD)	199.4 (16.8)	134.4 (14.5)	158.2 (7.3)	139.7 (8.3)	147.8 (10)	145.6 (11.5)
Patients who left without being seen (%), mean (SD)	12.1 (2.2)	4.6 (1.7)	4.1 (0.3)	3.7 (0.2)	3.8 (0.4)	5.0 (1.1)

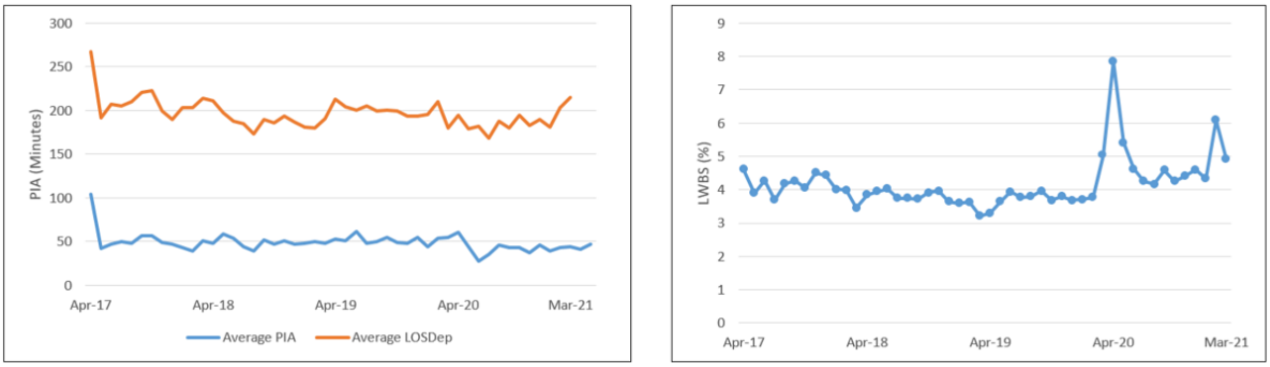
[a] These data were obtained from the publication by Patey et al [21].

[b] SD value is unavailable.

[c] PIA: provider initial assessment.

[d] LOSDep: length of stay for departed patients.

Figure 5. Average time to provider initial assessment (PIA), length of stay for departed patients (LOSDep), and patients who left without being seen.



WSLC Model

In contrast to the static framework that we used to examine the work system at one point in time, we applied the WSLC model to examine the changes to the work system over time through planned and unplanned changes as part of the system's natural evolution. We found that the surge management system project followed the 4 phases of the WSLC: initiation, development, implementation, and operation and maintenance (Figure 4) [9]. The project was initiated to address missed wait time benchmarks and high ED wait times, and it was initiated with the independent external review. As we discussed, this review identified areas for improvement, and interventions for a new surge management system were created in the development phase, planned strategies were implemented in the ED, and surge management system process and outcomes were continuously monitored. Initially, continuous monitoring was performed informally; however, as we mentioned previously, a change committee was created to take a more systematic approach to monitor the surge management system and make the necessary changes over time.

The planned changes included the eHealth system portion of the surge management system itself, along with the interventions described previously; however, in addition, we found that unanticipated opportunities and challenges emerged through the development cycle. For example, during initiation, the implementation team found that in addition to the changes in workflow, training, and creation of the surge management system that the team thought they would have to make, they discovered that they also needed to create a culture of patient *worthiness*, and an intervention to address this change was developed. Throughout the phases of the project, we found that the implementation team followed an iterative approach. When they recognized that changes were necessary in the static framework elements, they had the authority to make those changes. An implementation team member commented the following:

> *It's interim improvements as we go...where we're going to fix what we see is wrong before we go anywhere else.*

During the implementation phase, the team found that during the busy times, despite following the new surge management policies and procedures, staff were still often unable to manage the wait times sufficiently. Therefore, the team went back to the development phase and created a new *hero* shift to provide added ED physician and nursing support during these times. This was an unanticipated opportunity and proved to have beneficial results. In addition, during the operation and maintenance phase, the team saw unanticipated adaptations in the frequency with which staff entered the surge measures in the surge management system, and additional staff training was conducted. The implementation team also saw issues with accessing information for some indicators, and the surge management system was modified accordingly.

We found that the iterative approach helped the team to address issues as they arose. However, the changes required by the new ED surge management process were not easily made, and the implementation team faced some resistance. Staff commented the following:

> *Some people understand the process of flow better than others.*
>
> *And then you've got some doctors, it's like pulling teeth...they're just associated with the old way of doing things.*
>
> *I think it was really hard to go from being very rule-oriented to being more flexible.*

The implementation team recognized the challenges with the new system:

> *The strategy is controversial...it could be definitely seen as something that people wouldn't want to do.*

For example, the changes included broadening the scope of work for the primary care paramedics to include transporting patients to an admitting floor. This could have been perceived as negative, but survey participants indicated a high degree of change fairness (Table 2):

> *People affected negatively by this change were treated fairly.*

The implementation team commented that the challenges to incorporating the surge management system into the daily workflow were exacerbated by the fact that they could not make the use of the system mandatory:

> *It's really tough to do mandatory things. It's not like a factory...most people got practices that are guided by professional bodies...probably the biggest challenge here is that.*

However, we found that, overall, the staff indicated that the changes were managed well, as demonstrated by agreement levels (Table 2—change fairness, management support for the change, and commitment to change). Sufficient resources were available to support the change (14/20, 70% agreed or strongly agreed), there was sufficient management support for the changes (15/19, 79%), and the change served an important purpose (16/20, 80%).

The need for an iterative approach was also reflected in the fact that the ED work system itself was not stable. We saw that there were changes to the surge management system, new staff, new activities and process, and other new systems. An ED staff member commented the following:

> *...Change is now part of regular work...and here it seems to just be part of what you do now. It's actually become just regular work.*

This reflects the need for continuous monitoring of the surge management process and illustrates how the WSLC model can be used to manage the implementation to recognize unanticipated opportunities and adaptations over time.

Discussion

Principal Findings

The broad scope and holistic approach of WST is well suited to guide eHealth system implementations in the ED as it has the potential to address a key barrier to successful implementations—integrating the system into clinicians' work. EDs are complex systems that involve a variety of groups responsible for guiding patients through different organizational and clinical processes during their care. System implementation in such a complex setting requires focus on the work system, and not just the IT artifact. The surge management system is an example of a successful implementation, as the surge management system and related interventions improved key ED wait times. In our study, we demonstrate what made this implementation a success, through a WST analysis lens. It is important to understand what aspects of the implementation made it a success because if we understand what works, other implementation teams also can use this approach. We found that the key success factors were the incorporation of the entire work system into the implementation framework and the iterative approach.

Although the implementation team did not follow WST, we propose that it can be applied in the implementation of eHealth systems in such environments. The WST's static framework can help to broaden the focus of implementation from just the eHealth system to a view that considers the eHealth system as a part of a large work system, in which human and technological components work together to manage patient care. The emphasis on the services produced and the value of those services to the staff and patients is particularly beneficial in the health care setting. This will ensure that the implementation is focused on the end user and the ultimate goal of the system, rather than a narrow focus such as system's use. We saw that the external review helped to identify all the elements of the ED work system and not only the eHealth system portion of the surge management system that can affect the success of the system implementation. The resulting interventions helped to ensure alignment among elements in the WST's static framework and to focus the attention throughout implementation on the ultimate goal of reducing ED wait times.

We saw how taking a more holistic approach for the implementation helped to alleviate some of the common barriers to eHealth system success. For example, rather than focusing on the technical components of the surge management system, something that is a significant contributor to the gap between prospect and reality [22,23] and which is only a component of the WST static framework, the implementation team also focused on the other elements of the static framework. Furthermore, they addressed resistance to change, which is another barrier to eHealth interventions [24,25], by involving staff in the implementation, through training and creation of a change committee. We saw how they considered the different perspectives of different staff and management involved in the surge management process by involving them in the implementation process, something, which if not done can be another barrier [26,27]. There is a widespread perception in human-computer interaction that recognizes the importance of user-centered design and participatory design approaches. This has also been shown to be important in the design of workflow associated with eHealth system implementations. On the basis of their review of the factors influencing the outcome of eHealth interventions, Granja et al [6] proposed that user involvement in the design of the workflow is the most important factor for the success of eHealth systems. Using the WST lens helped us to identify the different elements of the work system and discover how the interventions addressed any misalignment between the elements. We propose that applying the static framework will help implementation teams to ensure that they are including the essential elements of the work system and provide them with a systematic way to assess the alignment between the elements.

Viewing the implementation of the surge management system through the WSLC model created an approach that differs fundamentally from the traditional systems development life cycle. In the systems development life cycle methodology, the system is the technical artifact that is created, and it does not necessarily incorporate iterations. We found that viewing the surge management system as part of dynamic work processes with a series of changes that emerged through planned and unplanned events, and their interactions, highlighted the importance of continuously identifying the changes or areas of misalignment in the work system. It also helped to clarify what interventions were needed and how to manage such changes, while recognizing that changes to the work system can occur through planned initiatives or emerge over time. We propose that implementation teams can use the WSLC model to follow an iterative approach that will allow them to be open to unanticipated opportunities and recognize unanticipated adaptations during the implementation and give them a structured way of dealing with the planned and emergent changes.

Researchers have adapted and used WST in different areas over several decades [9,10]; however, it has had limited application in health care [28,29]. To the best of our knowledge, it has not yet been applied in the ED, nor has it focused on how it can be used during the implementation of an eHealth system. Therefore, we broaden the focus of WST by applying it in the health care field (specifically ED) and to its use in the implementation of eHealth systems. We propose that the static framework can be used by non-IT and IT professionals to analyze the work system to incorporate the sociotechnical aspects of implementing an eHealth system as part of a large work system. Then, the WSLC model can be used to structure the implementation and monitor and manage the system over time. Future studies can examine the implementation of an eHealth system with a team that is following the WST approach, to uncover any problems that may arise and identify opportunities for overcoming them.

Limitations

This study was conducted in 1 rural hospital ED. Experiences regarding the implementation of the same or other ED systems may be different in other types of ED settings, thus limiting generalizability. However, the practices of WST have been shown to improve implementation success in other complex settings. Another limitation is the small sample size of the survey; however, approximately half of the ED staff responded. Moreover, the observations were conducted only for 2 days; however, during observation, we asked if the shifts represented a typical day, and they indicated that they did. They also indicated that after 10 PM, the ED is generally not busy, and thus, surge management during nights is similar to a nonbusy time during the day. The ED wait time data before, during, and after implementation of the surge management system support the success of this implementation. The use of the mixed methods approach with surveys, interviews, observations, document review, and ED wait time data analysis should have helped to combat potential limitations.

Conclusions

This conceptualization of the surge management system implementation through the lens of WST gives us insight into how to structure the implementation of a surge management system to incorporate the broad work system. We captured how the static framework can be a useful tool to assess the elements of the broad work system that need to be changed and managed to successfully implement an ED surge management system, and we propose that the WSLC model can provide a structured way to manage the implementation of such changes. This study addresses the need for more studies on surge management systems and methodologies to implement eHealth systems that incorporate the broad work system. These findings can guide further studies and implementations of eHealth systems.

Acknowledgments

This study was funded in part by the Canadian Institutes of Health Research, provincial sponsors including the Newfoundland and Labrador Provincial Government, Newfoundland and Labrador Eastern Regional Health Authority, and Trinity Conception Placentia Health Foundation. The author thanks Sileola Akinboyeje, Masters in Management student at the Faculty of Business, for her assistance with part of the data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions, interview questions, and observation protocol.
[DOCX File , 29 KB - jmir_v24i8e37472_app1.docx]

References

1. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS One 2018;13(8):e0203316 [FREE Full text] [doi: 10.1371/journal.pone.0203316] [Medline: 30161242]
2. Saghaeiannejad-Isfahani S, Hazhir F, Jalali R. An assessment of emergency department information systems based on the HL7 functional profile. J Educ Health Promot 2019;8:26 [FREE Full text] [doi: 10.4103/jehp.jehp_64_18] [Medline: 30993119]
3. Callen J, Li L, Georgiou A, Paoloni R, Gibson K, Li J, et al. Does an integrated Emergency Department Information System change the sequence of clinical work? A mixed-method cross-site study. Int J Med Inform 2014 Dec;83(12):958-966. [doi: 10.1016/j.ijmedinf.2014.08.010] [Medline: 25241155]
4. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
5. 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]
6. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
7. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 02;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
8. Kruse CS, Regier V, Rheinboldt KT. Barriers over time to full implementation of health information exchange in the United States. JMIR Med Inform 2014 Sep 30;2(2):e26 [FREE Full text] [doi: 10.2196/medinform.3625] [Medline: 25600635]
9. Alter S. Work system theory: overview of core concepts, extensions, and challenges for the future. J Assoc Inform Syst 2013 Feb;14(2):72-121. [doi: 10.17705/1jais.00323]
10. Alter S. Work system theory as a platform: response to a research perspective article by Niederman and March. J Assoc Inf Syst 2015 Jun;16(6):485-514. [doi: 10.17705/1jais.00398]
11. Allaudeen N, Vashi A, Breckenridge JS, Haji-Sheikhi F, Wagner S, Posley KA, et al. Using lean management to reduce emergency department length of stay for medicine admissions. Qual Manag Health Care 2017;26(2):91-96. [doi: 10.1097/QMH.0000000000000132] [Medline: 28375955]
12. Arbune A, Wackerbarth S, Allison P, Conigliaro J. Improvement through small cycles of change: lessons from an academic medical center emergency department. J Healthc Qual 2017;39(5):259-269. [doi: 10.1111/jhq.12078] [Medline: 28858964]
13. Aksel G, Bildik F, Demircan A, Keles A, Kilicaslan I, Guler S, et al. Effects of fast-track in a university emergency department through the National Emergency Department Overcrowding Study. J Pak Med Assoc 2014 Jul;64(7):791-797 [FREE Full text] [Medline: 25255588]
14. System and use assessment survey. Canada Health Infoway. 2015 Feb 1. URL: https://tinyurl.com/48d92x8y [accessed 2018-08-05]
15. Caldwell SD, Herold DM, Fedor DB. Toward an understanding of the relationships among organizational change, individual differences, and changes in person-environment fit: a cross-level study. J Appl Psychol 2004 Oct;89(5):868-882. [doi: 10.1037/0021-9010.89.5.868] [Medline: 15506866]
16. FEDOR DB, CALDWELL S, HEROLD DM. The effects of organizational changes on employee commitment: a multilevel investigation. Pers Psychol 2006 Mar;59(1):1-29. [doi: 10.1111/j.1744-6570.2006.00852.x]
17. Herscovitch L, Meyer JP. Commitment to organizational change: extension of a three-component model. J Appl Psychol 2002 Jun;87(3):474-487. [doi: 10.1037/0021-9010.87.3.474] [Medline: 12090605]
18. William HD, Ephraim RM. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inform Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
19. Glaser BG, Strauss AL. The Discovery of Grounded Theory Strategies for Qualitative Research. Piscataway, New Jersey: Aldine Transaction; 1967.
20. King N. Using templates in the thematic analysis of text. In: Essential Guide to Qualitative Methods in Organizational Research. Thousand Oaks, California: SAGE Publications; 2004.
21. Patey C, Norman P, Araee M, Asghari S, Heeley T, Boyd S, et al. SurgeCon: priming a community emergency department for patient flow management. West J Emerg Med 2019 Jul;20(4):654-665 [FREE Full text] [doi: 10.5811/westjem.2019.5.42027] [Medline: 31316707]
22. Biggs K, Lowe P, Walsh J, Lagios K. Audit of a sexual health website email link for general practitioners. Int J STD AIDS 2010 Jan;21(1):60-62. [doi: 10.1258/ijsa.2009.009192] [Medline: 20029066]

23. Declerck G, Aimé X. Reasons (not) to spend a few billions more on EHRs: how human factors research can help. Yearb Med Inform 2014 Aug 15;9:90-96 [FREE Full text] [doi: 10.15265/IY-2014-0033] [Medline: 25123727]
24. LeTourneau B. Managing physician resistance to change. J Healthc Manag 2004;49(5):286-288. [doi: 10.1097/00115514-200409000-00003]
25. Narine L, Persaud DD. Gaining and maintaining commitment to large-scale change in healthcare organizations. Health Serv Manage Res 2003 Aug;16(3):179-187. [doi: 10.1258/095148403322167933] [Medline: 12908992]
26. Kapadia V, Ariani A, Li J, Ray PK. Emerging ICT implementation issues in aged care. Int J Med Inform 2015 Nov;84(11):892-900. [doi: 10.1016/j.ijmedinf.2015.07.002] [Medline: 26250987]
27. Frauenberger C, Good J, Fitzpatrick G, Iversen OS. In pursuit of rigour and accountability in participatory design. Int J Hum Comput Stud 2015 Feb;74:93-106 [FREE Full text] [doi: 10.1016/j.ijhcs.2014.09.004] [Medline: 26109833]
28. Dadgar M, Joshi K. The role of information and communication technology in self-management of chronic diseases: an empirical investigation through value sensitive design. J Assoc Inf Syst 2018 Feb;19(2):86-112. [doi: 10.17705/jais1.00485]
29. Johnsen HM, Fruhling A, Fossum M. An analysis of the work system framework for examining information exchange in a healthcare setting. Commun Assoc Inf Syst 2016;39:73-95. [doi: 10.17705/1cais.03905]

Abbreviations

ED: emergency department
LOSDep: length of stay for departed patients
PIA: provider initial assessment
WSLC: work system life cycle
WST: work system theory

Edited by G Eysenbach; submitted 22.02.22; peer-reviewed by A Schmitz, R Riessen; comments to author 11.05.22; revised version received 13.06.22; accepted 18.07.22; published 25.08.22.

Please cite as:
Jewer J
Investigating a Work System Approach to Implement an Emergency Department Surge Management System: Case Study
J Med Internet Res 2022;24(8):e37472
URL: https://www.jmir.org/2022/8/e37472
doi:10.2196/37472
PMID:36006684

Original Paper

Predictors of Dropout in a Digital Intervention for the Prevention and Treatment of Depression in Patients With Chronic Back Pain: Secondary Analysis of Two Randomized Controlled Trials

Isaac Moshe[1], MA; Yannik Terhorst[2], MSc; Sarah Paganini[3], PhD; Sandra Schlicker[4], PhD; Laura Pulkki-Råback[1], PhD; Harald Baumeister[2], PhD; Lasse B Sander[5], PhD; David Daniel Ebert[6], PhD

[1]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland
[2]Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, Ulm University, Ulm, Germany
[3]Department of Sport Psychology, Institute of Sports and Sport Science, Albert-Ludwigs-University of Freiburg, Freiburg, Germany
[4]Clinic for Psychiatry and Psychotherapy, Rhein-Erft-Kreis, Germany
[5]Medical Psychology and Medical Sociology, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[6]Department for Sport and Health Sciences, Chair for Psychology & Digital Mental Health Care, Technical University of Munich, Munich, Germany

Corresponding Author:
Isaac Moshe, MA
Department of Psychology and Logopedics
Faculty of Medicine
University of Helsinki
P.O. Box 63
Helsinki, 00014
Finland
Phone: 358 406324442
Email: isaac.moshe@helsinki.fi

Abstract

Background: Depression is a common comorbid condition in individuals with chronic back pain (CBP), leading to poorer treatment outcomes and increased medical complications. Digital interventions have demonstrated efficacy in the prevention and treatment of depression; however, high dropout rates are a major challenge, particularly in clinical settings.

Objective: This study aims to identify the predictors of dropout in a digital intervention for the treatment and prevention of depression in patients with comorbid CBP. We assessed which participant characteristics may be associated with dropout and whether intervention usage data could help improve the identification of individuals at risk of dropout early on in treatment.

Methods: Data were collected from 2 large-scale randomized controlled trials in which 253 patients with a diagnosis of CBP and major depressive disorder or subclinical depressive symptoms received a digital intervention for depression. In the first analysis, participants' baseline characteristics were examined as potential predictors of dropout. In the second analysis, we assessed the extent to which dropout could be predicted from a combination of participants' baseline characteristics and intervention usage variables following the completion of the first module. Dropout was defined as completing <6 modules. Analyses were conducted using logistic regression.

Results: From participants' baseline characteristics, lower level of education (odds ratio [OR] 3.33, 95% CI 1.51-7.32) and both lower and higher age (a quadratic effect; age: OR 0.62, 95% CI 0.47-0.82, and age^2: OR 1.55, 95% CI 1.18-2.04) were significantly associated with a higher risk of dropout. In the analysis that aimed to predict dropout following completion of the first module, lower and higher age (age: OR 0.60, 95% CI 0.42-0.85; age^2: OR 1.59, 95% CI 1.13-2.23), medium versus high social support (OR 3.03, 95% CI 1.25-7.33), and a higher number of days to module completion (OR 1.05, 95% CI 1.02-1.08) predicted a higher risk of dropout, whereas a self-reported negative event in the previous week was associated with a lower risk of dropout (OR 0.24, 95% CI 0.08-0.69). A model that combined baseline characteristics and intervention usage data generated the most accurate predictions (area under the receiver operating curve [AUC]=0.72) and was significantly more accurate than models based on baseline characteristics only (AUC=0.70) or intervention usage data only (AUC=0.61). We found no significant influence of pain, disability, or depression severity on dropout.

Conclusions: Dropout can be predicted by participant baseline variables, and the inclusion of intervention usage variables may improve the prediction of dropout early on in treatment. Being able to identify individuals at high risk of dropout from digital health interventions could provide intervention developers and supporting clinicians with the ability to intervene early and prevent dropout from occurring.

(J Med Internet Res 2022;24(8):e38261) doi:10.2196/38261

KEYWORDS

adherence; dropout; law of attrition; attrition; digital health; internet intervention; depression; back pain; comorbidity; mental health; eHealth; mobile phone

Introduction

Background

Chronic back pain (CBP) is a major global health concern with lifetime prevalence rates of 60%-70% [1]. CBP is the leading cause of activity limitation and work absenteeism and accounts for the highest number of disability-adjusted life years worldwide [2]. Approximately 1 in 5 adults with CBP are diagnosed with major depression and many more experience subclinical symptoms [3]. In patients with CBP, comorbid depression is often associated with lower treatment adherence, poorer treatment outcomes, increased medical complications, and higher health care use [4-6].

Psychological interventions have been demonstrated to be effective in the treatment of depression for individuals with CBP [7]. Accordingly, cognitive behavioral therapy is now recommended as the first-line treatment in most international guidelines, for example, the National Institute for Health and Care Excellence [8]. However, the ability to access psychotherapy is a significant barrier to treatment. On average, less than 1 in 5 people in high-income countries and less than 1 in 27 people in low- and middle-income countries receive appropriate treatment [9], giving rise to a *treatment gap* in mental health care [10].

Digital interventions, which deliver psychotherapeutic components via a web browser or smartphone app, have been proposed as a scalable, cost-effective way to meet the growing demand for psychological treatment and address many of the challenges associated with accessing traditional face-to-face therapy [11-15]. There is now a large body of evidence supporting the efficacy of digital interventions for the prevention and treatment of depression [16-20], with recent studies providing promising evidence for the treatment of depression in patients with comorbid physical conditions [21-25].

Despite the demonstrated efficacy, adherence to digital interventions remains a major challenge. A recent meta-analysis on digital interventions for depression identified that, on average, little more than half of the participants completed the full intervention and only 25% of the participants completed the full intervention when the intervention was delivered within routine health care settings [16]. The same meta-analysis also identified that completing the full intervention was the strongest predictor of outcomes, underscoring the importance of treatment adherence [26]. Similar rates of intervention dropout have been reported in other studies on digital health interventions [27-30], giving rise to what Eysenbach [31] has referred to as the “law of attrition,” the observation that a substantial proportion of users in eHealth apps will stop using the intervention before completing the full treatment protocol.

Several studies have assessed predictors of dropout in digital mental health interventions [27,32-34]. In a systematic review of internet-based interventions for anxiety and depression, Christensen et al [34] found that higher baseline depression severity and older age were associated with higher rates of intervention dropout. In a meta-analysis of individual patient data, Karyotaki et al [33] found that being male, having a lower education level, younger age, and comorbid anxiety symptoms significantly increased the risk of dropout from self-guided digital interventions for depression.

However, to the best of our knowledge, no research has been published to date that assesses which factors may predict dropout in a digital intervention for depression in patients with comorbid somatic illness. This question may be particularly pertinent for these individuals as chronic pain and depression are often associated with reduced motivation to initiate or complete goal-directed tasks [35,36]. As the treatment schedules of patients with multimorbidities can already be quite demanding [37,38], higher levels of pain disability—or lower confidence in performing activities while in pain (pain self-efficacy)—may therefore influence an individual’s ability to adhere to a digital intervention and thus experience the benefits [39].

Knowing which patient characteristics predict a higher likelihood of dropout may be valuable for identifying individuals in advance of treatment that might benefit from alternative care pathways [33,40]. However, it is unclear how useful the baseline predictors may be once a patient has already started treatment. In most digital health interventions, there is a steady attrition over time, with most users dropping out after completing the first 1 or 2 modules—the “attrition-phase” [31,41,42]. The ability to identify which individuals are at high risk of dropout early on in treatment could be valuable for supporting clinicians, especially within stepped-care models where rapid identification of nonresponders and the speed of providing alternative treatment can have a significant influence on outcomes [43,44].

Although the field is still nascent, there is a wealth of data generated from digital interventions that may aid the prediction of dropout once a patient has started treatment. In the same way that many digital apps outside of health care use *churn-prediction models* to identify if a user is likely to churn (ie, stop using the app as intended), similar models and principles of human-computer interaction may be valuable to predict dropout and maximize adherence within the field of digital health. For example, in a digital intervention for a chronic lifestyle disease, Pedersen et al [45] were able to predict dropout with 89% precision using a model that included the number of messages received from the health coach, 2 weeks of inactivity, and the provider of the intervention as the most significant predictors. In a study of a self-help digital intervention for the treatment of insomnia, Bremer et al [46] identified individuals at risk of dropout with an area under the receiver operating curve (AUC) of 0.719 using a combination of baseline characteristics (eg, self-reported stress levels) and intervention usage data (eg, number of days to complete each module). One of the most significant findings of the study was that the prediction of user dropout was possible early on in the intervention (after completion of the introductory module).

Objectives

As digital mental health interventions are being increasingly adopted by health care systems worldwide [13,47], the ability to identify patients at risk of dropout may provide valuable information to improve the adherence, and thus effectiveness, of digital interventions. This study aimed to identify the factors that may predict dropout in a digital intervention for depression in individuals with CBP. In particular, we sought to assess which participant characteristics may be associated with a higher risk of dropout and whether intervention usage data could help improve the identification of individuals at risk of dropout early on in treatment.

Drawing from 2 recent studies on a guided digital intervention for the treatment and prevention of depression in individuals with comorbid CBP, we asked the following research questions:

1. Can we predict intervention dropout from participant baseline characteristics? If so, which participant characteristics predict a higher likelihood of dropout?
2. Can we develop an “early warning system” that identifies participants at risk of dropout early on in the intervention? Specifically, how accurately can we predict which participants will drop out after completion of the first module and what is the most accurate model for classifying at-risk individuals?
3. Do predictors of dropout differ between patients clinically diagnosed with major depressive disorder (where the intervention is targeting treatment) and patients with subclinical depressive symptoms (where the intervention is targeting prevention)?

Methods

Study Design

This study was a secondary analysis of data from 2 trials that assessed the efficacy of a therapist-guided internet-based intervention for the treatment [48] and prevention [21] of depressive symptoms in patients with comorbid CBP. Both trials were observer-masked, multicenter, pragmatic, and randomized controlled trials with a parallel design. The trials were conducted simultaneously using the same intervention, procedures, and research setting but targeted individuals with different levels of depressive symptomatology (diagnosed depressive disorder of mild to moderate severity in the study by Baumeister et al [48] and subclinical but at least mild levels of depressive symptoms in the study by Sander et al [21]). For the purpose of this study, the trial data were combined. All the participants provided written informed consent.

Participants

All participants (N=253) assigned to the intervention arms of the primary studies were included in this analysis. The inclusion criteria of the primary studies were as follows: (1) age ≥18 years; (2) presence of depressive symptoms, either reported persistent subthreshold depressive symptoms in the past 3 months or meeting the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, criteria for a mild to moderate depressive episode or persistent depressive disorder; (3) diagnosed back pain chronicity of at least 6 months; (4) have German language skills; and (5) access to internet and PC. The exclusion criteria were as follows: (1) having ongoing or planned psychotherapy within the forthcoming 3 months, (2) being currently suicidal or having had suicidal attempts within the past 5 years, or (3) having had a severe depressive episode within the past 6 months. In the primary studies, participants were recruited during or following discharge from 1 of the 82 orthopedic clinics across Germany. They were recruited personally by a clinician or on the web using a flyer and information letters distributed by the clinic.

Intervention

The intervention is a guided internet- and mobile-based intervention for the treatment (eSano BackCare-D [49]) or prevention (eSano BackCare-DP [50]) of depression in patients with comorbid CBP. The content of the intervention is based on cognitive behavioral therapy for depression and includes elements of psychoeducation, social skills, problem-solving, behavioral activation, relaxation, motivation for physical exercises, and psychological pain intervention elements. Modules consist of information provided by text, video, audio, and interactive exercises and include a homework assignment. At the start of each module, participants reported their perceived stress level at the time and whether they had experienced any negative events in the previous 7 days. There are 6 regular modules and 3 optional modules focusing on sleep, partnership and sexuality, and work. Participants were advised to complete 1 session per week. During the intervention, participants were guided by trained and supervised psychologists (e-coaches) who provided written feedback within 48 hours of each completed module and by answering any queries.

Measures

Baseline Measures

In this study, 8 baseline characteristic variables were assessed as potential predictors of dropout. Variables were chosen on the basis of previous research pointing to demonstrated or hypothetical relationships between the predictor variables and intervention adherence or dropout [32-34,45,46,51]. Demographic characteristics included age, sex (male or female), education level (based on the International Standard Classification of Education by UNESCO [52], low: level 1-2, medium: level 3-4, and high: level 5+), marital status (single, in a relationship, or divorced or widowed), and social support (low, medium, or high). Clinical characteristics included depression, as measured by the Hamilton Depression Rating Scale (HAM-D; [53]); pain disability, as measured by the Oswestry Disability Index (ODI; [54]); and pain self-efficacy, as measured by the Pain Self-Efficacy Questionnaire (PSEQ). The process variables included internet affinity, as measured by the Internet Affinity Scale. Further details on all measurements are provided in the original study protocols [49,50].

Intervention Usage Measures

Intervention usage measures included both active and passive measures. The active measures were the stress level reported by the participant at the start of each module ("Burden") and the occurrence of any negative events experienced in the past 7 days, self-reported by the participant at the start of each module ("Negative Events"). Burden was assessed using a Likert scale from 0 to 10, where 0="not burdened at all" and 10="extremely burdened." Negative events were dummy coded as 0="no negative event in the past week" and 1="at least one negative event in the past week." For passive measures, we included the number of days taken to complete each module ("N Days to complete module") and the number of minutes spent on the web completing each module ("Time spent online completing module").

Dropout

Dropout was defined as completing <6 intervention modules, in accordance with the intervention developers [49,50]. It was operationalized as a binary outcome (dropped out or did not drop out).

Analytic Strategy

Predicting Dropout From Participant Baseline Characteristics

To assess whether participants' baseline characteristics could predict dropout, analyses were conducted using logistic regression in 3 steps. First, we conducted a series of bivariate analyses to assess the odds ratios (ORs) of each baseline variable (bivariate "bivariate model"). Second, we repeated the analyses with all baseline variables simultaneously entered into the binomial model (the "complete model"). Finally, we built a "parsimonious model" in which we excluded nonsignificant predictors with no incremental predictive power from the complete model in a stepwise procedure.

Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used as measures of model fit and for model comparison. For nested models, likelihood ratio tests were used to directly compare whether 2 models were significantly different from one another [55]. Collinearity was assessed using variance inflation factors and tolerance (1/variance inflation factors). The assumption of linearity of the logit (a linear relationship between the predictors and dropout) was assessed for all continuous predictor variables, and any variables found violating the assumption were transformed based on a visual inspection of the plot.

As this was an exploratory study, we did not adjust for multiple testing. The study was not powered for confirmatory analysis of the predictors, and alpha adjustment may have increased the likelihood of type II errors.

Predicting Dropout Early on in the Intervention

To assess whether we could identify people at risk of dropout early on in the intervention, we first created a subset of the data available up until the point of module 1 completion (ie, baseline assessment data and intervention data captured until participants had completed the first module). We then compared the performance of three separate logistic regression models using the constrained data set: (1) a model based on participant baseline characteristics only—the "baseline characteristics model," (2) a model based on intervention usage variables only—the "intervention usage model," and (3) a model combining all baseline characteristics and intervention usage variables—the "combined model." The quality of the models was assessed using the area under the receiver operating characteristic curve (AUROC) and related measures of sensitivity and specificity [55]. The optimal threshold for AUROC was determined using Youden J statistic [56].

Sensitivity analyses were conducted to assess whether predictors differed in the prevention and treatment studies. Here, study was included as a dummy-coded variable (0=PROD-BP for the prevention study and 1=WARD-BP for the treatment study) in all parsimonious models, first as an additional predictor to assess for a main effect of study type on dropout and then as an interaction term with other predictors in the model to assess whether the effect of a predictor differed across studies.

To assess whether the number of modules completed influenced the relative risk of dropout, we conducted sensitivity analyses using Cox proportional hazards regression [57]. In this study, we assessed whether significant predictors of dropout differed between the 2 methods. Analyses were conducted according to the procedures outlined by Eysenbach [31]. The number of completed modules was used as a proxy for time. Models were built using the same 3-step procedure outlined above for logistic regression.

Missingness occurred in 111 out of 3084 (3.6%) data points and was assumed to be missing at random, indicating that missingness depended on observed data [58]. To avoid bias introduced by missingness, missing data were imputed using multiple imputation by chained equations [59,60]. Predictors for missing values were selected based on (1) model-induced predictors, (2) predictors based on bivariate correlation, and (3) bivariate correlation with missingness according to the procedures outlined by van Buuren and Groothuis-Oudshoorn [60]. Predictive mean matching was used as the imputation method. The number of imputed data sets was set to 20, and the number of iterations was set to 10. Convergence was visually assessed and confirmed. Regression analysis was performed on each imputed data set, and the results were pooled according to the rules by Rubin [61]. Sensitivity analyses were conducted using observed (nonimputed) data to compare with the results of the complete models using imputed data.

All analyses were conducted in R using R Studio (RStudio, PBC; [62]). The pROC package was used to calculate the AUROC [63]. The Caret package (R Foundation for Statistical

Computing) was used to calculate the sensitivity and specificity [64]. The multiple imputation by chained equations (MICE) package was used for multiple imputation and likelihood ratio tests [60].

Ethics Approval

This study was a secondary of analysis of data from two RCTs—Sander et al [21] and Baumeister et al [48]. In the original studies, all the participants provided written informed consent. The trial in Sander et al [21] was registered at German Clinical Trials Register (DRKS00007960). The trial in Baumeister et al [48] was registered at the World Health Organization International Clinical Trials Registry (DRKS00009272). All procedures were approved by the ethics committee of the Albert Ludwigs University of Freiburg, Germany (REC No. 8022-6-BW-H-2015; No. 297/14_150513 for the WARD-BP trial, EK-297/14_150513 for the PROD-BP study).

Results

Descriptive Statistics

Among the 253 participants, 149 (58.9%) were female and 104 (41.1%) were male. The age of participants ranged from 24 to 78 years, with a mean age of 51.1 (SD 8.88) years. Of the 253 participants, 171 (67.6%) reported having a low level of education. Of the 253 participants, 34 (13.4%) were single, 180 (71.1%) were in a relationship or married, and 39 (15.4%) were divorced or separated. The mean depression severity at baseline was 10.3 (SD 5.93), as measured by the HAM-D, and 9.94 (SD 4.41), as measured by the Patient Health Questionnaire-9. The mean level of pain disability was 31.3 (SD 14.7), as measured by the ODI, and the mean level of pain self-efficacy was 34.9 (SD 13.0), as measured by the PSEQ. Table 1 provides a detailed summary of the demographic and clinical characteristics of the participants.

On average, the participants completed 4.65 out of the 6 regular and 3 optional modules (SD 3.48). The participants took an average of 17.64 (SD 19.55) days to complete each module, and the mean time on the web taken to complete a module was 80.26 (SD 136.96) minutes. The mean self-reported burden was 4.55 (SD 1.97), and the mean number of self-reported negative events across the intervention was 0.80 (SD 1.33). Table 2 shows that 45.1% (114/253) of the participants dropped out of the intervention before completing at least six modules. The table also shows that the number of participants completing the modules decreased steadily as the intervention progressed.

Table 1. Demographic and clinical characteristics of the participants (N=253).

Variable	Value[a]
Age (years)	
Mean (SD)	51.1 (8.88)
Median (range)	52 (24-78)
Sex, n (%)	
Male	104 (41.1)
Female	149 (58.9)
Education level, n (%)	
Low	171 (67.6)
Medium	45 (17.8)
High	37 (14.6)
Marital status, n (%)	
Single	34 (13.4)
In a relationship (including married)	180 (71.1)
Divorced or separated	39 (15.4)
Children, n (%)	
Yes	200 (79.1)
No	53 (20.9)
Social support, n (%)	
None	9 (3.6)
Low	67 (26.5)
Sufficient	81 (32)
High	73 (28.9)
Very high	23 (9.1)
Internet affinity (IAS[b])	
Mean (SD)	9.33 (4)
Median (range)	8.5 (5-25)
Missing, n (%)	1 (0.4)
HAM-D[c]	
Mean (SD)	10.3 (5.93)
Median (range)	9 (0-30)
Missing, n (%)	1 (0.4)
PHQ-9[d]	
Mean (SD)	9.94 (4.41)
Median (range)	10.0 (1-24)
Missing, n (%)	3 (1.2)
Pain disability (ODI[e])	
Mean (SD)	31.3 (14.7)
Median (range)	30.0 (0-72)
Missing, n (%)	1 (0.4)
Pain self-efficacy (PSEQ[f])	
Mean (SD)	34.9 (13)

Variable	Value[a]
Median (range)	36 (0-59)
Missing, n (%)	1 (0.4)
Dropout, n (%)	
No	139 (54.9)
Yes	114 (45.1)

[a] Values are based on observed data.

[b] IAS: Internet Affinity Scale.

[c] HAM-D: Hamilton Depression Rating Scale.

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

[e] ODI: Oswestry Disability Index.

[f] PSEQ: Pain Self-Efficacy Questionnaire.

Table 2. Intervention usage data.

Variable	Value
Modules completed, mean (SD)	4.65 (3.48)
Participants completing modules, n (%)	
Module 1	188 (74.31)
Module 2	174 (68.77)
Module 3	159 (62.85)
Module 4	148 (58.5)
Module 5	136 (53.75)
Module 6	128 (50.59)
Module 7	109 (43.08)
Module 8	71 (28.06)
Module 9	61 (24.11)
Days to module completion, mean (SD)	17.64 (19.55)
Time spent on the web completing module (minutes), mean (SD)	80.26 (136.96)
Burden, mean (SD)	4.55 (1.97)
Negative events, mean (SD)	0.81 (1.33)
Dropout, n (%)	114 (45.1)

Predicting Dropout Using Participant Baseline Characteristics

Table 3 displays the performance of the models used to predict dropout based on the participant baseline characteristics. As the Patient Health Questionnaire-9 and PSEQ scores were highly correlated with HAM-D and ODI (r=0.63 and r=−0.73, respectively) and were not significant in the bivariate analyses, they were not included in the multivariate predictor models to prevent collinearity. The results of the bivariate analysis indicated that a lower level of education was significantly associated with a higher risk of dropout (OR 2.43, 95% CI 1.19-4.97; P=.01, whereas higher age predicted a lower risk of dropout (OR 0.97, 95% CI 0.94-0.99; P=.02). None of the other potential predictors (sex, social support, internet affinity, baseline depression severity, and baseline pain intensity) were statistically significant at the level of P<.05 in the bivariate analysis.

In the complete model, being single was found to be an additional significant predictor of dropout (OR 2.54, 95% CI 1.09-5.90; P=.03). When age was added as a quadratic term (age^2) to the model to account for the nonlinear relationship between age and dropout, we found that both age (OR 0.63, 95% CI 0.47-0.84; P<.001) and age^2 (OR 1.55, 95% CI 1.17 to –2.05; P<.001) were significant predictors, such that both lower and higher age were associated with increased risk of dropout.

In the parsimonious model, where predictors were reduced stepwise to relevant predictors only, low education (OR 3.33, 95% CI 1.51-7.32; P<.001) and age (OR 0.62, 95% CI 0.47-0.82; P<.001 and age^2: OR 1.55, 95% CI 1.18-2.04; P<.001) remained significant predictors of dropout. Marital status, internet affinity, baseline depression severity, and baseline pain intensity were found to be nonsignificant after controlling for the other predictors.

Table 3. Predictors of dropout from participant baseline characteristics.

Predictors	Bivariate model[a]		Complete model[a]		Parsimonious model[a]	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age	0.97 (0.94-0.99)	.02	0.63 (0.47-0.84)	<.001	0.62 (0.47-0.82)	<.001
Age^2	N/A[c]	N/A	1.55 (1.17-2.05)	<.001	1.55 (1.18-2.04)	<.001
Sex (male)	1.60 (0.96-2.66)	.07	1.68 (0.96-2.94)	.07	N/A	N/A
Marital status						
Single vs in a relationship	1.97 (0.93-4.20)	.08	2.54 (1.09-5.90)	.03	N/A	N/A
Divorced or widowed vs in a relationship	0.54 (0.26-1.14)	.11	0.62 (0.27-1.42)	.25	N/A	N/A
Education						
Low vs medium	2.43 (1.19-4.97)	.01	3.77 (1.68-8.49)	<.001	3.33 (1.51-7.32)	<.001
High vs medium	1.88 (0.75-4.71)	.18	2.08 (0.74-5.83)	.16	2.08 (0.78-5.57)	.15
Social support						
Low vs high	0.83 (0.45-1.53)	.55	0.83 (0.41-1.69)	.60	N/A	N/A
Medium vs high	1.60 (0.88-2.90)	.13	1.64 (0.86-3.14)	.13	N/A	N/A
IAS[d]	1.02 (0.96-1.09)	.53	1.02 (0.95-1.10)	.58	N/A	N/A
HAM-D[e]	0.99 (0.95-1.03)	.67	0.98 (0.93-1.03)	.36	N/A	N/A
Pain disability	1.00 (0.98-1.02)	.85	1.00 (0.97-1.02)	.69	N/A	N/A
Pain self-efficacy (PSEQ[f])	1.00 (0.98-1.02)	.92	N/A	N/A	N/A	N/A
PHQ-9[g]	0.97 (0.91-1.02)	.24	N/A	N/A	N/A	N/A

[a] Models based on imputed data.

[b] OR: odds ratio.

[c] N/A: Not applicable.

[d] IAS: Internet Affinity Scale.

[e] HAM-D: Hamilton Depression Rating Scale.

[f] PSEQ: Pain Self-Efficacy Questionnaire.

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

Predicting Dropout Early on in the Intervention

Tables 4-6 provide a comparison of the models used to predict dropout following the completion of the first module. In the parsimonious model using only participant baseline characteristics, higher and lower age (OR 0.57, 95% CI 0.41-0.79; *P*=.001 and age^2: OR 1.68, 95% CI 1.22-2.31; *P*=.001) and low education (OR 2.98, 95% CI 1.04-8.56; *P*=.04) were significant predictors of dropout. The AUROC for the model was 0.70, the sensitivity was 68%, and the specificity was 62%.

In the parsimonious model using only intervention usage data, a higher number of days to module completion predicted a higher risk of dropout (OR 1.04, 95% CI 1.01-1.07; *P*=.005), whereas a self-reported negative event in the previous week was associated with a lower risk of dropout (OR 0.30, 95% CI 0.11-0.81; *P*=.02). The AUROC for the model was 0.61, the sensitivity was 56%, and the specificity was 54%.

In the parsimonious model that combined participant baseline characteristics and intervention usage variables as predictors, higher and lower age (OR 0.60, 95% CI 0.42-0.85; *P*=.004 and age^2: OR 1.59, 95% CI 1.13-2.23; *P*=.008), medium versus high social support (OR 3.03, 95% CI 1.25-7.33; *P*=.02), and a higher number of days to module completion (OR 1.05, 95% CI 1.02-1.08; *P*=.002) all predicted a higher risk of dropout, whereas a self-reported negative event in the previous week was associated with a lower risk of dropout (OR 0.24, 95% CI 0.08-0.69; *P*=.008). The AUROC for the model was 0.72, the sensitivity was 76%, and the specificity was 59%.

As shown in Table 7, a comparison of the parsimonious models based on participant baseline characteristics and intervention usage variables revealed that the model that combined baseline and intervention usage variables was the most accurate in predicting dropout (AIC=198.9; BIC=253.9) and significantly more accurate than the model using participant baseline characteristics only (AIC=212.6; BIC=254.7; χ^2_{181}=5.3; *P*=.006).

Table 4. Predictors of dropout following module 1 completion—participant baseline characteristics.

Predictors	Bivariate model[a]		Complete model[a]		Parsimonious model[a]	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age	0.96 (0.92-1.00)	.048	0.61 (0.43-0.87)	.006	0.57 (0.41-0.79)	.001
Age^2	0.97 (0.93-1.01)	.12	1.58 (1.12-2.24)	.01	1.68 (1.22-2.31)	.001
Sex (male)	1.45 (0.75-2.82)	.27	1.37 (0.65-2.92)	.41	N/A[c]	N/A
Marital status						
Single vs in a relationship	1.79 (0.68-4.72)	.24	2.06 (0.67-6.27)	.20	N/A	N/A
Divorced or widowed vs in a relationship	0.75 (0.30-1.90)	.55	1.01 (0.35-2.93)	.99	N/A	N/A
Education						
Low vs medium	2.17 (0.83-5.68)	.11	3.60 (1.19-10.88)	.02	2.98 (1.04-8.56)	.04
High vs medium	2.03 (0.61-6.75)	.25	1.77 (0.46-6.80)	.40	2.18 (0.60-7.85)	.23
Social support						
Low vs high	0.84 (0.34-2.07)	.69	0.80 (0.29-2.22)	.66	N/A	N/A
Medium vs high	2.65 (1.22-5.80)	.01	2.32 (1.00-5.38)	.05	N/A	N/A
IAS[d]	1.02 (0.94-1.11)	.59	1.02 (0.92-1.13)	.72	N/A	N/A
HAM-D[e]	0.99 (0.94-1.05)	.79	1.00 (0.94-1.07)	.98	N/A	N/A
Pain disability	0.99 (0.97-1.01)	.38	0.98 (0.95-1.01)	.28	N/A	N/A
Pain self-efficacy (PSEQ[f])	1.00 (0.97-1.02)	.97	N/A	N/A	N/A	N/A
PHQ-9[g]	0.97 (0.90-1.04)	.41	N/A	N/A	N/A	N/A

[a] Models based on imputed data.

[b] OR: odds ratio.

[c] N/A: not applicable.

[d] IAS: Internet Affinity Scale.

[e] HAM-D: Hamilton Depression Rating Scale.

[f] PSEQ: Pain Self-Efficacy Questionnaire.

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

Table 5. Predictors of dropout following module 1 completion—intervention usage variables.

Predictors	Bivariate model[a]		Complete model[a]		Parsimonious model[a]	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Number of days to module 1 completion	1.04 (1.01-1.06)	.007	1.04 (1.01-1.07)	.005	1.04 (1.01-1.07)	.005
Negative events	0.34 (0.13-0.87)	.03	0.30 (0.11-0.81)	.02	0.30 (0.11-0.81)	.02
Burden	0.98 (0.84-1.15)	.83	1.01 (0.85-1.19)	.91	N/A[c]	N/A
Time spent on the web completing module 1	1.00 (0.99-1.01)	.81	1.00 (0.99-1.01)	.94	N/A	N/A

[a] Models based on imputed data.

[b] OR: odds ratio.

[c] N/A: not applicable.

Table 6. Predictors of dropout following module 1 completion—baseline and intervention usage variables.

Predictors	Complete model[a]		Parsimonious model[a]	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value
Age	0.54 (0.36-0.80)	.003	0.60 (0.42-0.85)	.004
Age^2	1.76 (1.19-2.61)	.005	1.59 (1.13-2.23)	.008
Sex (male)	1.53 (0.65-3.59)	.33	N/A[c]	N/A
Marital status				
Single vs in a relationship	1.88 (0.56-6.34)	.31	N/A	N/A
Divorced or widowed vs in a relationship	1.21 (0.38-3.78)	.75	N/A	N/A
Education				
Low vs medium	3.21 (0.91-11.33)	.07	N/A	N/A
High vs medium	1.17 (0.26-5.23)	.84	N/A	N/A
Social support				
Low vs high	0.83 (0.27-2.56)	.75	0.92 (0.34-2.51)	.88
Medium vs high	3.40 (1.33-8.64)	.01	3.03 (1.25-7.33)	.02
IAS[d]	0.99 (0.88-1.11)	.84	N/A	N/A
HAM-D[e]	0.99 (0.92-1.06)	.78	N/A	N/A
Pain disability	0.97 (0.94-1.00)	.08	N/A	N/A
Number of days to module 1 completion	1.05 (1.02-1.08)	.004	1.05 (1.02-1.08)	.002
Negative events	0.22 (0.07-0.68)	.009	0.24 (0.08-0.69)	.008
Burden	0.96 (0.80-1.17)	.71	N/A	N/A
Time spent on the web completing module 1	1.01 (0.99-1.02)	.28	N/A	N/A

[a] Models based on imputed data.

[b] OR: odds ratio.

[c] N/A: not applicable.

[d] IAS: Internet Affinity Scale.

[e] HAM-D: Hamilton Depression Rating Scale.

Table 7. Predictors of dropout following completion of module 1: model comparison (models based on imputed data).

Model	AIC[a]	BIC[b]	AUROC[c]	Sensitivity (%)	Specificity (%)
Model 1: baseline variables	212.6	254.7	0.70	68	62
Model 2: intervention variables	207.5	223.7	0.61	56	54
Model 3: baseline+intervention variables	198.9	253.9	0.72	76	59

[a] AIC: Akaike information criterion.

[b] BIC: Bayesian information criterion.

[c] AUROC: area under the receiver operating characteristic curve.

Sensitivity Analyses

Sensitivity analyses assessing whether findings differed between the treatment (WARD-BP) and prevention (PROD-BP) studies found no significant difference between the two, either in terms of main effect or interaction effects with other predictors. Sensitivity analyses assessing whether findings differed when using Cox proportional hazards regression versus logistic regression found no difference in the significant predictors. Multimedia Appendix 1 presents the results from the Cox regression analyses. Sensitivity analyses assessing whether the results differed between the models using observed data and those using imputed data revealed no difference in the predictors found to be significant. Results from the models using observed data are presented in Multimedia Appendix 2.

Discussion

Principal Findings

This study aimed to identify the predictors of treatment dropout in a digital intervention for the treatment and prevention of depression in patients with comorbid CBP. From the participants' baseline characteristics, we found that a lower education level and lower and higher age (a quadratic effect) predicted a higher risk of dropout. From the intervention usage variables, we found that a higher number of days to module completion predicted a higher risk of dropout, whereas the occurrence of a negative event in the previous week predicted a lower risk of dropout.

Participants with lower education levels were more likely to drop out of treatment, which is consistent with a large body of research on adherence to both digital interventions and face-to-face psychotherapy [33,51,65-67]. This may reflect the fact that these individuals find it harder to comprehend the intervention material or the digital format, and thus, they lose the motivation to continue [30,51]. It is worth noting that 67.5% (171/253) of the participants in this study were classified as having low levels of education. Lower levels of education have also been associated with longer duration or higher occurrence of back pain [68], underscoring the need for additional research on digital interventions for this particular patient group. The finding that both younger and older age predicted higher risk of dropout suggests that the relationship between age and dropout may be more complex than has been previously identified, either owing to the nonlinear relationship between the two or a possible interaction between age and other factors such as computer literacy [33,69]. More pertinently, it points to the challenges of predicting which individuals are likely to drop out of a digital intervention based on baseline characteristics alone [70]. Finally, the finding that neither pain disability nor depression severity levels were associated with an increased risk of dropout is important as it suggests that digital interventions targeting comorbid depression are acceptable for patients with varying levels of pain intensity and depression symptom severity. This is further supported by the fact that we found no significant difference in predictors when the intervention was aimed at prevention and when it was aimed at treatment. Taken together, these findings provide promising evidence that digital interventions may provide a scalable approach for integrating psychological treatment within pain management routines in health care settings.

This study also demonstrated the feasibility of predicting dropout early on in the intervention based on data restricted to the first module and participant baseline characteristics. Our finding that the number of days taken to complete the first module significantly predicted dropout is consistent with the study by Bremer et al [46] that identified the average number of days taken to complete each module as one of the strongest predictors of dropout in a digital intervention for insomnia. This may reflect a number of underlying causes, including challenges interacting with the intervention, low motivation, lack of time, or low perceived value [40]. The finding that a self-reported negative event in the previous week predicted a lower risk of dropout may be because of the fact that experiencing a negative event (or being asked to report on one) provided greater intrinsic motivation to complete the module and is consistent with research demonstrating that some people drop out from an intervention because they no longer feel they need it [30,31,40]. However, it is worth highlighting that the CIs for the predictor were wide, so the results should be interpreted with caution and examined in future studies using larger sample sizes to determine whether the findings replicate. Notwithstanding, this is the first study to identify that a simple 1-item self-report questionnaire may be used to aid the prediction of dropout during a digital intervention, thus highlighting the potential of incorporating such assessments within digital interventions in the future. Interestingly, the relationship between participant education level and dropout was no longer significant in the models that combined baseline characteristics and intervention usage variables (Table 6). This suggests that a patient's education level may be less important at predicting dropout when including variables that reflect how the patient interacts with the intervention, such as how long it takes them to complete a specific module.

Indeed, a comparison of models using baseline characteristics and intervention usage variables revealed that a model that combined baseline characteristics and intervention usage data generated the most accurate predictions and was significantly more accurate than models based on baseline characteristics only or intervention usage data only. Moreover, in terms of clinical utility, the AUROC of 0.72 and sensitivity of 76% exceeded the accuracy threshold of 65%-70%, at which clinicians reportedly become willing to act on predictions [71]. Implemented within an intervention, dropout risk models such as this could be used to alert supporting clinicians and health care workers when an individual is at high risk of dropout, so that they are able to intervene early and ideally prevent it.

Notwithstanding the above, there is still significant room for improving model performance. In particular, the development of models that are able to predict dropout *before* completion of the first module would be especially valuable as a significant proportion of individuals drop out before then [65,69]. Developing more accurate models will require intervention developers to capture more granular data related to engagement with the intervention, for example, the number and timing of log-ins, interaction with specific components of the modules (eg, homework; [72]), data specifically related to the intervention target (eg, sleep data for an insomnia intervention; [46]), and additional self-report data such as early measures of therapeutic alliance with the coach [40,73]. Armed with comprehensive intervention usage data such as this, researchers will be better positioned to engineer both handcrafted (theory-driven) and automated features and assess their impact on predictive accuracy. Exploring the role of nonlinear machine learning models in improving model performance is also an area that holds potential, as has been demonstrated in several studies comparing the classification performance of machine learning algorithms with logistic regression in the prediction of dropout [45,46].

Finally, as the findings of this study were specific to one intervention, future research would also benefit from assessing whether the predictors found herein are valuable for predicting dropout in other digital interventions, alternative disorders, and different populations. If it is consistently found that there are a set of variables such as "number of days to module completion" that are associated with higher risk of dropout, these predictors may then be used to inform the basis of models for other interventions in the future. In the same way that outcome feedback technology that identifies individuals at risk of deterioration during treatment has been shown to improve eventual treatment outcomes [43,44], dropout warning systems could be used to alert the supporting clinician, guide care pathways (eg, in stepped-care models), or personalize the intervention itself in the case of self-help interventions. Given the high dropout rates found in real-world settings [16], this will become increasingly important as interventions are implemented within public and private health care systems to meet the growing demand for psychotherapy [47].

Strengths and Limitations

To the best of our knowledge, this is the first study to examine the predictors of dropout in a digital intervention for depression in individuals with a comorbid somatic illness. This is also the first study to compare whether predictors of dropout differ when the intervention is aimed at prevention (in a subclinical population) versus treatment (in a clinically diagnosed population) and using study samples with clinically verified diagnoses at baseline. Finally, in contrast to most studies conducted to date, which have been based on efficacy trials with small sample sizes and convenient samples [33], this analysis was based on data from 2 large-scale effectiveness trials. These trials were conducted within routine health care settings, where dropout rates are typically significantly higher [16], thereby providing high ecological validity.

Despite these strengths, we acknowledge several limitations of this study. First, the analyses were based on data specific to a prevention and treatment version of one intervention and one population, namely, individuals with depressive symptoms and CBP. As such, the predictors we found to be significant and the subsequent accuracy of the classification models may not generalize to other interventions or other populations. For example, dropout has been found to be significantly greater in unguided interventions than in guided interventions, and the mechanisms underlying dropout may differ between the two [33,51]. Future research would, therefore, benefit from assessing whether the predictors found to be significant in this study generalize to other interventions, populations, and settings. Second, as this was an exploratory study, we did not just adjust for multiple testing as alpha adjustment may have increased the likelihood of type II errors. Future research aimed at replicating the current findings in studies that are sufficiently powered for a confirmatory analysis would be valuable. Third, we had only a limited set of data from the intervention available for analysis. Several studies have demonstrated that a number of other variables derived from intervention usage are valuable in the prediction of both adherence and outcomes, including in-depth measures of engagement such as the frequency of log-ins [45] and interactions with specific content formats [72]. Accordingly, there may be other variables with further explanatory power that were not included in our models. The same applies to baseline characteristics, where studies have shown that data obtained from electronic medical records may be used to identify those at risk of dropout during face-to-face therapy [74,75]. Finally, although dropout in this study was operationalized according to the usage intended by the clinicians who developed the intervention [76], it is important to highlight that it is not always necessary for patients to complete the full per-protocol treatment to benefit clinically [30]. In other words, dropout is not always representative of a negative experience [31,40].

Conclusions

The high dropout rates associated with digital health interventions remain one of the biggest challenges to their successful implementation in real-world health care settings. Being able to identify individuals at high risk of dropout early on in treatment may provide clinicians and intervention developers with a valuable opportunity to intervene early and prevent dropout from occurring. Using a combined set of predictors from patient baseline characteristics and intervention usage data, we were able to identify individuals at risk of dropout early on in a digital intervention for depression in patients with comorbid CBP. Future research should explore ways of improving model accuracy and investigate the feasibility and efficacy of using these models directly within the interventions themselves to improve adherence.

Acknowledgments

IM was supported by grants from the Finnish Foundation for Psychiatric Research and the Finnish Cultural Foundation. Open access was funded by the Helsinki University Library.

Data Availability

The data that support the findings of this study are available on reasonable request from HB (harald.baumeister@uni-ulm.de). Data requesters will need to sign a data access agreement. Provision of data is subject to data security regulations. Investigator support depends on the available resources. The statistical code may be made available to an investigator upon reasonable request to the corresponding author.

Authors' Contributions

IM, YT, HB, LBS, and DDE contributed to the study conception and design. IM, YT, HB, and DDE contributed to the methods and analysis. IM prepared the original draft. YT, SP, SS, LPR, HB, LBS, and DDE critically reviewed and edited the manuscript. All authors read and approved the final manuscript and account for all aspects of this work.

Conflicts of Interest

SS has received fees for lectures or workshops from training institutes for psychotherapists in the context of digital health research. HB has received consultancy fees, fees for lectures or workshops from chambers of psychotherapists and training institutes for psychotherapists in the context of digital health research and license fees for an Internet-intervention. LBS has received consultancy fees, fees for lectures and workshops from chambers of psychotherapists and training institutes for psychotherapists in the context of digital health research. DDE has received consultancy fees/served in the scientific advisory board from several companies such as Minddistrict, Lantern, Novartis, Sanofi, Schoen Kliniken, Ideamed, German health insurance companies (BARMER, Techniker Krankenkasse) and a number of federal chambers for psychotherapy. He is a stakeholder of the Institute for health training online (GET.ON), which aims to implement scientific findings related to digital health interventions into routine care. The authors have no further conflicts to declare.

Multimedia Appendix 1
Survival analyses assessing risk of dropout using Cox proportional hazards regression.
[DOCX File , 40 KB - jmir_v24i8e38261_app1.docx]

Multimedia Appendix 2
Results of the models using observed (nonimputed) data.
[DOCX File , 34 KB - jmir_v24i8e38261_app2.docx]

References

1. Wu A, March L, Zheng X, Huang J, Wang X, Zhao J, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020 Mar;8(6):299 [FREE Full text] [doi: 10.21037/atm.2020.02.175] [Medline: 32355743]
2. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
3. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003 Nov 10;163(20):2433-2445. [doi: 10.1001/archinte.163.20.2433] [Medline: 14609780]
4. Nicholas MK, Coulston CM, Asghari A, Malhi GS. Depressive symptoms in patients with chronic pain. Med J Aust 2009 Apr 06;190(S7):S66-S70. [doi: 10.5694/j.1326-5377.2009.tb02473.x] [Medline: 19351296]
5. Miles CL, Pincus T, Carnes D, Homer KE, Taylor SJ, Bremner SA, et al. Can we identify how programmes aimed at promoting self-management in musculoskeletal pain work and who benefits? A systematic review of sub-group analysis within RCTs. Eur J Pain 2011 Sep;15(8):775.e1-775.11. [doi: 10.1016/j.ejpain.2011.01.016] [Medline: 21354838]
6. Costa LC, Maher CG, McAuley JH, Hancock MJ, Herbert RD, Refshauge KM, et al. Prognosis for patients with chronic low back pain: inception cohort study. BMJ 2009 Oct 06;339:b3829 [FREE Full text] [doi: 10.1136/bmj.b3829] [Medline: 19808766]
7. Williams AC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012 Nov 14;11(11):CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
8. Depression in adults with a chronic physical health problem: recognition and management. National Institute for Health and Care Excellence. 2009 Oct 28. URL: https://www.nice.org.uk/guidance/cg91 [accessed 2020-10-27]
9. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Feb;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
10. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
11. Campion J, Javed A, Sartorius N, Marmot M. Addressing the public mental health challenge of COVID-19. Lancet Psychiatry 2020 Aug;7(8):657-659 [FREE Full text] [doi: 10.1016/S2215-0366(20)30240-6] [Medline: 32531299]
12. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of Internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. J Affect Disord 2018 Jan 01;225:733-755. [doi: 10.1016/j.jad.2017.07.018] [Medline: 28922737]

13. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020 Jun 15;3:85 [FREE Full text] [doi: 10.1038/s41746-020-0293-8] [Medline: 32566763]
14. Singla DR, Raviola G, Patel V. Scaling up psychological treatments for common mental disorders: a call to action. World Psychiatry 2018 Jun;17(2):226-227 [FREE Full text] [doi: 10.1002/wps.20532] [Medline: 29856556]
15. 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]
16. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
17. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
18. Wright JH, Owen JJ, Richards D, Eells TD, Richardson T, Brown GK, et al. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. J Clin Psychiatry 2019 Mar 19;80(2):18r12188 [FREE Full text] [doi: 10.4088/JCP.18r12188] [Medline: 30900849]
19. Sander L, Rausch L, Baumeister H. Effectiveness of Internet- and mobile-based psychological interventions for the prevention of mental disorders: a systematic review and meta-analysis protocol. Syst Rev 2016 Feb 16;5:30 [FREE Full text] [doi: 10.1186/s13643-016-0209-5] [Medline: 26880167]
20. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health: a report of the EFPA E-Health Taskforce. Eur Psychol 2018 May;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
21. Sander LB, Paganini S, Terhorst Y, Schlicker S, Lin J, Spanhel K, et al. Effectiveness of a guided Web-based self-help intervention to prevent depression in patients with persistent back pain: the PROD-BP randomized clinical trial. JAMA Psychiatry 2020 Oct 01;77(10):1001-1011 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1021] [Medline: 32459348]
22. O'moore KA, Newby JM, Andrews G, Hunter DJ, Bennell K, Smith J, et al. Internet cognitive-behavioral therapy for depression in older adults with knee osteoarthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2018 Jan;70(1):61-70. [doi: 10.1002/acr.23257] [Medline: 28426917]
23. Meyer B, Weiss M, Holtkamp M, Arnold S, Brückner K, Schröder J, et al. Effects of an epilepsy-specific Internet intervention (Emyna) on depression: results of the ENCODE randomized controlled trial. Epilepsia 2019 Apr;60(4):656-668. [doi: 10.1111/epi.14673] [Medline: 30802941]
24. Ebert DD, Nobis S, Lehr D, Baumeister H, Riper H, Auerbach RP, et al. The 6-month effectiveness of Internet-based guided self-help for depression in adults with type 1 and 2 diabetes mellitus. Diabet Med 2017 Jan;34(1):99-107. [doi: 10.1111/dme.13173] [Medline: 27334444]
25. Schlicker S, Baumeister H, Buntrock C, Sander L, Paganini S, Lin J, et al. A Web- and mobile-based intervention for comorbid, recurrent depression in patients with chronic back pain on sick leave (Get.Back): pilot randomized controlled trial on feasibility, user satisfaction, and effectiveness. JMIR Ment Health 2020 Apr 15;7(4):e16398 [FREE Full text] [doi: 10.2196/16398] [Medline: 32293577]
26. Musiat P, Johnson C, Atkinson M, Wilksch S, Wade T. Impact of guidance on intervention adherence in computerised interventions for mental health problems: a meta-analysis. Psychol Med 2022 Jan;52(2):229-240. [doi: 10.1017/S0033291721004621] [Medline: 34802474]
27. 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]
28. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
29. Wells MJ, Owen JJ, McCray LW, Bishop LB, Eells TD, Brown GK, et al. Computer-assisted cognitive-behavior therapy for depression in primary care: systematic review and meta-analysis. Prim Care Companion CNS Disord 2018 Mar 01;20(2):17r02196 [FREE Full text] [doi: 10.4088/PCC.17r02196] [Medline: 29570963]
30. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
31. 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]
32. Kok RN, Beekman AT, Cuijpers P, van Straten A. Adherence to a Web-based pre-treatment for phobias in outpatient clinics. Internet Interv 2017 Sep;9:38-45 [FREE Full text] [doi: 10.1016/j.invent.2017.05.004] [Medline: 30135835]

33. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided Web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
34. Christensen H, Griffiths KM, Farrer L. Adherence in Internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
35. Jonsson T, Christrup LL, Højsted J, Villesen HH, Albjerg TH, Ravn-Nielsen LV, et al. Symptoms and side effects in chronic non-cancer pain: patient report vs. systematic assessment. Acta Anaesthesiol Scand 2011 Jan;55(1):69-74. [doi: 10.1111/j.1399-6576.2010.02329.x] [Medline: 21039361]
36. Felger JC, Treadway MT. Inflammation effects on motivation and motor activity: role of dopamine. Neuropsychopharmacology 2017 Jan;42(1):216-241 [FREE Full text] [doi: 10.1038/npp.2016.143] [Medline: 27480574]
37. Knowles S, Hays R, Senra H, Bower P, Locock L, Protheroe J, et al. Empowering people to help speak up about safety in primary care: using codesign to involve patients and professionals in developing new interventions for patients with multimorbidity. Health Expect 2018 Apr;21(2):539-548 [FREE Full text] [doi: 10.1111/hex.12648] [Medline: 29266797]
38. Wallace E, Salisbury C, Guthrie B, Lewis C, Fahey T, Smith SM. Managing patients with multimorbidity in primary care. BMJ 2015 Jan 20;350:h176. [doi: 10.1136/bmj.h176] [Medline: 25646760]
39. Thompson EL, Broadbent J, Bertino MD, Staiger PK. Do pain-related beliefs influence adherence to multidisciplinary rehabilitation?: a systematic review. Clin J Pain 2016 Feb;32(2):164-178. [doi: 10.1097/AJP.0000000000000235] [Medline: 26735864]
40. Knowles SE, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. BMJ Open 2015 Nov 30;5(11):e008581 [FREE Full text] [doi: 10.1136/bmjopen-2015-008581] [Medline: 26621513]
41. Norlund F, Wallin E, Olsson EM, Wallert J, Burell G, von Essen L, et al. Internet-based cognitive behavioral therapy for symptoms of depression and anxiety among patients with a recent myocardial infarction: the U-CARE heart randomized Controlled trial. J Med Internet Res 2018 Mar 08;20(3):e88 [FREE Full text] [doi: 10.2196/jmir.9710] [Medline: 29519777]
42. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005 Mar 26;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]
43. Delgadillo J, de Jong K, Lucock M, Lutz W, Rubel J, Gilbody S, et al. Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial. Lancet Psychiatry 2018 Jul;5(7):564-572. [doi: 10.1016/S2215-0366(18)30162-7] [Medline: 29937396]
44. Shimokawa K, Lambert MJ, Smart DW. Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. J Consult Clin Psychol 2010 Jun;78(3):298-311. [doi: 10.1037/a0019247] [Medline: 20515206]
45. Pedersen DH, Mansourvar M, Sortsø C, Schmidt T. Predicting dropouts from an electronic health platform for lifestyle interventions: analysis of methods and predictors. J Med Internet Res 2019 Sep 04;21(9):e13617 [FREE Full text] [doi: 10.2196/13617] [Medline: 31486409]
46. Bremer V, Chow PI, Funk B, Thorndike FP, Ritterband LM. Developing a process for the analysis of user journeys and the prediction of dropout in digital health interventions: machine learning approach. J Med Internet Res 2020 Oct 28;22(10):e17738 [FREE Full text] [doi: 10.2196/17738] [Medline: 33112241]
47. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
48. Baumeister H, Paganini S, Sander LB, Lin J, Schlicker S, Terhorst Y, et al. Effectiveness of a guided Internet- and mobile-based intervention for patients with chronic back pain and depression (WARD-BP): a multicenter, pragmatic randomized controlled trial. Psychother Psychosom 2021;90(4):255-268. [doi: 10.1159/000511881] [Medline: 33321501]
49. Lin J, Sander L, Paganini S, Schlicker S, Ebert D, Berking M, et al. Effectiveness and cost-effectiveness of a guided Internet- and mobile-based depression intervention for individuals with chronic back pain: protocol of a multi-centre randomised controlled trial. BMJ Open 2017 Dec 28;7(12):e015226 [FREE Full text] [doi: 10.1136/bmjopen-2016-015226] [Medline: 29288172]
50. Sander L, Paganini S, Lin J, Schlicker S, Ebert DD, Buntrock C, et al. Effectiveness and cost-effectiveness of a guided Internet- and mobile-based intervention for the indicated prevention of major depression in patients with chronic back pain-study protocol of the PROD-BP multicenter pragmatic RCT. BMC Psychiatry 2017 Jan 21;17(1):36 [FREE Full text] [doi: 10.1186/s12888-017-1193-6] [Medline: 28109247]
51. Schmidt ID, Forand NR, Strunk DR. Predictors of dropout in Internet-based cognitive behavioral therapy for depression. Cognit Ther Res 2019 Jun;43(3):620-630 [FREE Full text] [doi: 10.1007/s10608-018-9979-5] [Medline: 32879540]
52. International Standard Classification of Education (ISCED). Institute for Statistics, The United Nations Educational, Scientific and Cultural Organization. 2017. URL: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf [accessed 2021-12-20]

53. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23(1):56-62 [FREE Full text] [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
54. Mannion AF, Junge A, Grob D, Dvorak J, Fairbank JC. Development of a German version of the Oswestry disability index. part 2: sensitivity to change after spinal surgery. Eur Spine J 2006 Jan;15(1):66-73 [FREE Full text] [doi: 10.1007/s00586-004-0816-z] [Medline: 15856340]
55. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd edition. Hoboken, NJ, USA: Wiley Online Library; 2000.
56. Youden WJ. Index for rating diagnostic tests. Cancer 1950 Jan;3(1):32-35. [doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3] [Medline: 15405679]
57. Cox DR. Regression models and life-tables. J R Stat Soc Series B Stat Methodol 1972 Jan;34(2):187-202. [doi: 10.1111/j.2517-6161.1972.tb00899.x]
58. Enders CK. Applied Missing Data Analysis. New York, NY, USA: The Guilford Press; 2010.
59. van Buuren S. Flexible Imputation of Missing Data. 2nd edition. Boca Raton, FL, USA: CRC Press; 2018.
60. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Soft 2011 Dec 12;45(3):1-67. [doi: 10.18637/jss.v045.i03]
61. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996 Jun;91(434):473-489. [doi: 10.1080/01621459.1996.10476908]
62. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. URL: https://www.r-project.org/about.html [accessed 2022-08-17]
63. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011 Mar 17;12:77 [FREE Full text] [doi: 10.1186/1471-2105-12-77] [Medline: 21414208]
64. Kuhn M. Building predictive models in R using the caret package. J Stat Soft 2008 Nov 10;28(5):1-26. [doi: 10.18637/jss.v028.i05]
65. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 02;2:97-105 [FREE Full text] [Medline: 19920949]
66. Al-Asadi AM, Klein B, Meyer D. Pretreatment attrition and formal withdrawal during treatment and their predictors: an exploratory study of the anxiety online data. J Med Internet Res 2014 Jun 17;16(6):e152 [FREE Full text] [doi: 10.2196/jmir.2989] [Medline: 24938311]
67. Jarrett RB, Minhajuddin A, Kangas JL, Friedman ES, Callan JA, Thase ME. Acute phase cognitive therapy for recurrent major depressive disorder: who drops out and how much do patient skills influence response? Behav Res Ther 2013 May;51(4-5):221-230 [FREE Full text] [doi: 10.1016/j.brat.2013.01.006] [Medline: 23485420]
68. Dionne CE, Von Korff M, Koepsell TD, Deyo RA, Barlow WE, Checkoway H. Formal education and back pain: a review. J Epidemiol Community Health 2001 Jul;55(7):455-468 [FREE Full text] [doi: 10.1136/jech.55.7.455] [Medline: 11413174]
69. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an Internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2019 Aug 01;21(8):e12775 [FREE Full text] [doi: 10.2196/12775] [Medline: 31373272]
70. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
71. Eisenberg JM, Hershey JC. Derived thresholds. Determining the diagnostic probabilities at which clinicians initiate testing and treatment. Med Decis Making 1983;3(2):155-168. [doi: 10.1177/0272989X8300300203] [Medline: 6415358]
72. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with Internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 01;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
73. Lawler K, Earley C, Timulak L, Enrique A, Richards D. Dropout from an Internet-delivered cognitive behavioral therapy intervention for adults with depression and anxiety: qualitative study. JMIR Form Res 2021 Nov 12;5(11):e26221 [FREE Full text] [doi: 10.2196/26221] [Medline: 34766909]
74. Davis A, Smith T, Talbot J, Eldridge C, Betts D. Predicting patient engagement in IAPT services: a statistical analysis of electronic health records. Evid Based Ment Health 2020 Feb;23(1):8-14 [FREE Full text] [doi: 10.1136/ebmental-2019-300133] [Medline: 32046987]
75. Chekroud AM, Foster D, Zheutlin AB, Gerhard DM, Roy B, Koutsouleris N, et al. Predicting barriers to treatment for depression in a U.S. national sample: a cross-sectional, proof-of-concept study. Psychiatr Serv 2018 Aug 01;69(8):927-934 [FREE Full text] [doi: 10.1176/appi.ps.201800094] [Medline: 29962307]
76. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]

Abbreviations

AIC: Akaike information criterion

AUC: area under the receiver operating curve
AUROC: area under the receiver operating characteristic curve
BIC: Bayesian information criterion
CBP: chronic back pain
HAM-D: Hamilton Depression Rating Scale
MICE: multiple imputation by chained equations
ODI: Oswestry Disability Index
OR: odds ratio
PSEQ: Pain Self-Efficacy Questionnaire

Edited by R Kukafka; submitted 25.03.22; peer-reviewed by AL Vuorinen, N Khalili-Mahani; comments to author 14.04.22; revised version received 03.06.22; accepted 15.07.22; published 30.08.22.

Please cite as:
Moshe I, Terhorst Y, Paganini S, Schlicker S, Pulkki-Råback L, Baumeister H, Sander LB, Ebert DD
Predictors of Dropout in a Digital Intervention for the Prevention and Treatment of Depression in Patients With Chronic Back Pain: Secondary Analysis of Two Randomized Controlled Trials
J Med Internet Res 2022;24(8):e38261
URL: https://www.jmir.org/2022/8/e38261
doi:10.2196/38261
PMID:36040780

Original Paper

Characterizing User Engagement With a Digital Intervention for Pain Self-management Among Youth With Sickle Cell Disease and Their Caregivers: Subanalysis of a Randomized Controlled Trial

Chitra Lalloo[1,2], PhD; Fareha Nishat[1], MPH; William Zempsky[3], MD; Nitya Bakshi[4,5], MBBS, MS; Sherif Badawy[6], MD; Yeon Joo Ko[7], BSc; Carlton Dampier[4,5*], MD; Jennifer Stinson[1,8*], RN, PhD; Tonya M Palermo[7*], PhD

[1]Department of Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Division of Pain & Palliative Medicine, Connecticut Children's Medical Center, Hartford, CT, United States

[4]Division of Pediatric Hematology/Oncology/Blood and Marrow Transplant, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, United States

[5]Aflac Cancer and Blood Disorders Center, Children's Healthcare of Atlanta, Atlanta, GA, United States

[6]Division of Hematology, Oncology and Stem Cell Transplantation, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States

[7]Child Health Behavior & Development, Seattle Children’s Hospital, Seattle, WA, United States

[8]Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Chitra Lalloo, PhD
Department of Child Health Evaluative Sciences
The Hospital for Sick Children
686 Bay Street
Toronto, ON, M5G 0A4
Canada
Phone: 1 416 813 2332 ext 302332
Email: chitra.lalloo@sickkids.ca

Abstract

Background: Sickle cell disease (SCD) is characterized by severe acute pain episodes as well as risk for chronic pain. Digital delivery of SCD pain self-management support may enhance pain self-management skills and accessibility for youth. However, little is known about how youth with SCD and their caregivers engage with digital health programs. iCanCope with pain is a digital pain self-management platform adapted for youth with SCD and caregivers through a user-centered design approach. The program was delivered via a website (separate versions for youth and caregiver) and mobile app (youth only).

Objective: We aimed to characterize patterns of user engagement with the iCanCope with SCD program among youth with SCD and their caregivers.

Methods: A randomized controlled trial was completed across multiple North American SCD clinics. Eligible youth were aged 12-18 years, diagnosed with SCD, English-speaking, and experiencing moderate-to-severe pain interference. Eligible caregivers were English-speaking with a child enrolled in the study. Dyads were randomized to receive the iCanCope intervention or attention-control education for 8-12 weeks. This report focused on engagement among dyads who received the intervention. User-level analytics were captured. Individual interviews were conducted with 20% of dyads. Descriptive statistics characterized quantitative engagement. Content analysis summarized qualitative interview data. Exploratory analysis tested the hypothesis that caregiver engagement would be positively associated with child engagement.

Results: The cohort included primarily female (60% [34/57] of youth; 91% [49/56] of caregivers) and Black (>90% of youth [53/57] and caregivers [50/56]) participants. Among 56 dyads given program access, differential usage patterns were observed: both the youth and caregiver engaged (16/56, 29%), only the youth engaged (24/56, 43%), only the caregiver engaged (1/56, 2%), and neither individual engaged (16/56, 29%). While most youth engaged with the program (40/57, 70%), most caregivers did not (39/56, 70%). Youth were more likely to engage with the app than the website (85% [34/57] versus 68% [23/57]), and the most popular content categories were goal setting, program introduction, and symptom history. Among caregivers, program introduction, behavioral plans, and goal setting were the most popular content areas. As hypothesized, there was a moderate positive association between caregiver and child engagement (χ^2_1=6.6; P=.01; φ=0.34). Interviews revealed that most dyads would continue to use the program (11/12, 92%) and recommend it to others (10/12, 83%). The reasons for app versus website preference among youth were ease of use, acceptable time commitment, and interactivity. Barriers to caregiver engagement included high time burden and limited perceived relevance of content.

Conclusions: This is one of the first studies to apply digital health analytics to characterize patterns of engagement with SCD self-management among youth and caregivers. The findings will be used to optimize the iCanCope with SCD program prior to release.

Trial Registration: ClinicalTrials.gov NCT03201874; https://clinicaltrials.gov/ct2/show/NCT03201874

(J Med Internet Res 2022;24(8):e40096) doi:10.2196/40096

KEYWORDS

engagement; adolescents; caregivers; sickle cell; pain; mHealth; self-management; digital health analytics; mixed methods; youth; management; disease; acute pain; chronic pain; coping; North America; intervention; child; digital health; program

Introduction

Sickle cell disease (SCD) is the most common genetic blood condition in North America and primarily affects people of African descent [1]. The hallmark feature of SCD is recurrent episodes of acute severe pain due to vaso-occlusive crisis [2]. Vaso-occlusive crisis occurs when sickled red blood cells block blood flow, resulting in tissues becoming deprived of oxygen and causing pain. SCD pain is reportedly worse than postoperative pain and as intense as terminal cancer pain, and it can negatively impact all aspects of health-related quality of life [3-5]. Youth tend to experience increased SCD pain burden as they age into adolescence and young adulthood, with increased health care utilization [4-7]. In addition to acute pain from vaso-occlusive crisis, many youth with SCD also experience daily chronic pain [8,9]. The negative consequences of acute and chronic SCD pain can include depression and anxiety, academic underachievement related to missing school, limited opportunities for social interaction with peers, impaired physical activity, poor sleep, and high stress [9]. The vast majority of SCD pain episodes are treated in the home setting [10]. Unfortunately, many of these episodes may not be optimally managed with pharmacological approaches alone [11-13].

Self-management has been defined as “the individual’s ability to manage the symptoms, treatment, physical and psychological consequences, and lifestyle changes inherent to living with a chronic illness” [14]. The most successful self-management interventions are rooted in the principles of cognitive behavioral therapy [15-18]. Cognitive behavioral therapy involves normalization of the patient’s experience through education, training in strategies for managing disease-related symptoms and other stressors, enhancing self-efficacy, and guidance on developing and maintaining a long-term self-management plan. Gaining skills in monitoring and managing SCD symptoms independently is particularly critical to achieve early in the disease trajectory, as many youth experience worsened pain and disability in adulthood [19].

Barriers to implementing cognitive behavioral therapy in SCD populations include the challenge of delivering pain self-management interventions in traditional office-based visits, costly resources required to provide such support, and maintaining patient engagement [12,20]. Ubiquitous consumer technologies (eg, smartphones) offer opportunities to enhance the accessibility and interactivity of self-management support for youth with SCD [21,22].

iCanCope with pain is a digital pain self-management platform originally developed for youth with chronic pain [23-25]. The platform was adapted for youth with SCD through a user-centered design approach, including a qualitative needs assessment and design sessions [12,26]. The iCanCope with SCD program is currently undergoing efficacy evaluation through a multi-centered randomized clinical trial (NCT03201874) [26]. In the present report, we explore engagement with the digital intervention during the course of this trial.

Little is known about how youth with SCD and their caregivers engage with digital health interventions, and research to date has focused mostly on internet-delivered programs [27]. A scoping review focusing on user engagement with mobile health apps did not identify any studies specific to SCD [28]. Moreover, there is little standardization of engagement metrics due to the heterogeneity of available digital health programs. To add to this body of knowledge, our goal was to operationally define user engagement (tied to program content) with the iCanCope with SCD pain self-management program, to characterize patterns of user engagement among youth with SCD and their caregivers, and to identify individual youth and caregiver perspectives about the digital intervention. Lastly, in an exploratory fashion, we tested the *a priori* hypothesis that caregiver engagement would be positively associated with child engagement.

Methods

Study Design

A 2-arm randomized trial design was used. Dyads were randomized to 1 of 2 possible study conditions: (1) iCanCope with SCD intervention or (2) attention-control education. Details on the trial methodology are available in an open-access protocol paper [26], and the trial has been registered on ClinicalTrials.gov (NCT03201874). This paper will focus on program engagement data from dyads who received the intervention condition. Comparative efficacy results from the broader trial will be reported separately in a future paper.

Ethics Approval

The study was approved by the locally responsible institutional ethics boards (Seattle Children's Institutional Review Board, STUDY00001578; Emory University Institutional Review Board, 00092216; Connecticut Children's Institutional Review Board, 17-115-CCMC; and The Hospital for Sick Children Research Ethics Board, 1000053724).

Participants

The main recruitment sites were Connecticut Children's Medical Center, Children's Healthcare of Atlanta/Emory University, Seattle Children's Hospital, and The Hospital for Sick Children in Toronto. In addition, study referrals were accepted from University of Mississippi Medical Center, Boston Medical Center, and Northwestern University/Lurie Children's Hospital. The study was centrally managed at Seattle Children's Research Institute.

Youth were eligible if they (1) were aged between 12 and 18 years, (2) were diagnosed with any SCD genotype, (3) were able to speak and read English, (4) scored at least 4 (indicating moderate-to-severe pain interference over the past month) on the Sickle Cell Pain Burden Interview [6], and (5) were willing and able to complete online measures. Youth were excluded if they had significant cognitive limitations that would impair their ability to use and understand the iCanCope program, as per their health care provider or caregiver. Youth were also excluded if they had received more than 4 sessions of outpatient psychological therapy for pain management in the 6 months prior to the time of screening.

Caregivers were eligible if (1) they were able to speak and read English, (2) they were willing and able to complete online measures, and (3) their child was a study participant. Youth were permitted to enroll in the study even if their caregiver chose not to participate.

Procedure

Dyads assigned to the intervention condition were given user-authenticated access to the iCanCope with SCD program for a period of 8-12 weeks. The program content, based on the principles of cognitive behavioral therapy, was delivered via a modular website and complementary smartphone app as summarized in Tables 1 and 2. Standardized orientation materials were used to guide participants on how to login and independently use the program. Separate websites with unique content were provided for youth and caregiver participants, while a smartphone app (iOS/Android) was provided for youth participants only. Website navigation can be found in Figures S1 and S2 in Multimedia Appendix 1. The youth and caregiver websites were each organized into 6 core modules in addition to 2 optional youth modules on insomnia and negative mood. Features of the smartphone app included a daily symptom tracker (ie, pain intensity, pain impact, mood, and sleep quality), a calendar to view historical symptom data, goal setting in relevant domains (eg, sleep and mood), a library of SCD education and pain coping strategies, and a community forum to interact with other users (Figure S3 in Multimedia Appendix 1).

Table 1. iCanCope with sickle cell disease program content and delivery for youth.

iCanCope with SCD[a] program		Content delivery	
Content category	Example content	Website	App
Introduction to the program	General information about the iCanCope with SCD program and what to expect	Yes	Yes
About pain management	3Ps of pain management (psychological, physical, and pharmacological strategies)	Yes	Yes
About SCD and treatment	Genetics, potential impact of SCD on daily life, and treatment options	Yes	Yes
Goal setting	Guidance on how to formulate specific, measurable, achievable, relevant, and time-bound goals; a feature within the smartphone app where users could set and track personalized goals to improve activity, mood, and sleep	Yes	Yes
Symptom tracking	A feature within the smartphone app where users could report daily pain intensity, pain impact, mood, and sleep quality and view their symptom trends	N/A[b]	Yes
History	A feature within the smartphone app where users could view previously reported symptoms	N/A	Yes
Community support	A forum feature within the smartphone app where users could post responses to community questions and "favorite" the posts of other app users	N/A	Yes
Stress, relaxation, and negative emotions	Reducing negative thoughts, thought stopping, deep breathing exercises, muscle relaxation, imagery, scheduling pleasant activities, and finding the positives	Yes	Yes
Sleep and insomnia	Pain and sleep, healthy sleep habits, ways to fall and stay asleep, and how to think differently about sleep	Yes	Yes
Communication and self-advocacy	Communication skills, talking with the health care team, and talking with the school	Yes	Yes
Healthy lifestyle and looking ahead	Pacing, graded activity, hydration, considerations for the future, and transition to adult health care	Yes	Yes

[a] SCD: sickle cell disease.

[b] N/A: not applicable.

Table 2. iCanCope with sickle cell disease program content and delivery for caregivers.

iCanCope with SCD[a] program		Content delivery	
Content category	Example content	Website	App
Introduction to the program	What teens are learning in the program; specific, measurable, achievable, relevant, and timebound goals to support teens	Yes	N/A[b]
Behavioral plans	How to create behavioral plans to increase teen adaptive behaviors	Yes	N/A
Problem solving	Learn how to approach problems positively, identify problems effectively, and generate and implement solutions	Yes	N/A
Communication	Strategies to help communicate with teens, health care providers, and school staff	Yes	N/A
Wrap-up	Review of key concepts	Yes	N/A

[a] SCD: sickle cell disease.

[b] N/A: not applicable.

Outcome Measurement

Baseline Characteristics

At baseline, caregivers completed a background questionnaire to capture sociodemographic data (eg, age, race, and ethnicity), and both caregivers and youth completed information about their technology access and usage. Youth were screened into the study using the Sickle Cell Pain Burden Interview to assess SCD pain burden in the past month. The Sickle Cell Pain Burden Interview yielded a total score ranging from 0 (no pain burden) to 28 (severe pain burden). Dyads additionally completed several other clinical effectiveness outcomes detailed in an open-access protocol paper [26].

Primary Outcome

Quantitative

User-level analytics were captured for the app and website components of the iCanCope with SCD program. App engagement was characterized using APEEE (Analytics Platform to Evaluate Effective Engagement) [29], while website engagement data were captured using Google Analytics.

Qualitative

To gather perceptions of the treatment program from youth and caregivers, individual semistructured interviews were conducted with a convenience subset of dyads representing approximately 20% of the intervention group. All dyads were invited to complete an interview following their posttreatment assessment and prior to their 6-month follow-up assessment. Recruitment for the interview continued until 12 caregivers and youth completed the interviews. Separate interviews were conducted with youth and caregivers. The interview guide was designed to capture perspectives on the experience of using the iCanCope with SCD program, including likes, dislikes, value, and areas for improvement. Interviews were 15-20 minutes in duration, conducted over the telephone, and audio recorded for later transcription.

Data Analysis

Quantitative data were analyzed using STATA version 15.1 (Stata Statistical Software). Descriptive statistics were used to summarize background characteristics of the sample and program engagement data. An engagement interaction with the website was defined as a unique content page view. For the app, the following engagement interactions were captured: symptom check-in completed; goal created; library article viewed; and interaction within the community forum. For the youth-specific intervention, program content was distributed across the website and app such that an individual user might access topic-specific information (eg, “healthy sleep habits”) by viewing a website page or reading an app library article. Given this overlap in program content delivery, engagement data for youth participants were also mapped by content category (Table 1).

To address our exploratory aim regarding the relationship between parent engagement and child engagement, we conducted a chi-square test (significance was set at $\alpha<.05$) and assessed the strength of the association using the phi correlation. The phi correlation can range from 0 to 1 and was interpreted as follows: no correlation or very weak correlation (0-0.19), weak correlation (0.20-0.29), moderate correlation (0.30-0.49), strong correlation (0.50-0.69), and very strong correlation (0.70-1.00) [30].

Audio recordings from the qualitative interviews were transcribed verbatim and analyzed by 2 team members (CL and FN) using Dedoose Version 9.0.17 (SocioCultural Research Consultants, LLC). Simple content analysis, a dynamic process that summarizes the informational content of data, was used [31,32]. Specifically, data for all participants were coded and organized into categories that reflected the emerging themes. The raw data were revisited on a regular basis throughout the analytic process to ensure that the codes were grounded in the data. Any disagreements were resolved through consensus.

Participant recruitment spanned periods before and after onset of the COVID-19 pandemic. For analysis purposes, March 2020 was used to differentiate between *prepandemic* and *pandemic* recruitment periods in North America.

Results

Baseline Characteristics

Recruitment was carried out between January 1, 2018, and September 30, 2021. A total of 57 youth and 56 caregivers (56 dyads and 1 youth-only participant) were randomized to the iCanCope with SCD intervention condition and received instructions on how to access the program. The sample included dyads who enrolled in the study before (43/56, 75%) and after (14/56, 25%) pandemic onset. Demographic characteristics of the sample are provided in Tables 3 and 4.

Nearly all youth participants (55/57, 96%) were in middle school or high school. Most youth (47/57, 82%) reported using a smartphone multiple times per day. Caregiver participants (n=56) reported their highest completed education level as high school or less (11/56, 20%), vocational or trade school or some college/university (15/56, 27%), college or university (18/56, 32%), or graduate degree or professional school (11/56, 20%). Most dyad households (30/56, 54%) included 1 or 2 children under 18 years of age (range of 1 to 6 children). With regard to the total annual household income before taxes, participants reported less than US $24,999 (11/56, 20%), US $25,000-49,999 (16/56, 29%), US $50,000-74,999 (9/56, 16%), US $75,000-99,999 (9/56, 16%), and US $100,000 or more (5/56, 9%).

Table 3. Characteristics of youth participants.

Characteristic	Youth value (N=57)
Age (years), mean (SD)	14.8 (2.0)
Sex, n (%)	
Female	34 (60)
Male	23 (40)
Gender identity[a], n (%)	
Female	34 (60)
Male	23 (40)
Race, n (%)	
American Indian	1 (2)
Black	53 (93)
Latino/Hispanic	1 (2)
Asian	2 (4)
Ethnicity[b], n (%)	
Hispanic	5 (9)
Non-Hispanic	46 (84)
Unknown	4 (7)
SCD[c] genotype[d], n (%)	
Hemoglobin SS	35 (66)
Hemoglobin SC	12 (23)
Hemoglobin S beta thal plus	3 (6)
Hemoglobin S beta thal zero	2 (4)
Unknown	1 (2)
SCD Pain Burden Interview score, mean (SD)	10.9 (5.8)
Currently taking hydroxyurea, n (%)	40 (70)
Receiving regular blood transfusions, n (%)	8 (14)

[a] Response options were available for additional gender identities including transgender.

[b] Missing for ethnicity (n=2).

[c] SCD: sickle cell disease.

[d] Missing for SCD genotype (n=4).

Table 4. Characteristics of caregiver participants.

Characteristic	Caregiver value (N=56)
Relationship with the youth participant[a], n (%)	
Biological mother	49 (89)
Biological father	3 (5)
Brother	2 (4)
Step-mother	1 (2)
Marital status[b], n (%)	
Common law or married	23 (44)
Separated or widowed	8 (15)
Single	21 (40)
Race, n (%)	
Black	51 (90)
White	1 (2)
Other[c]	2 (4)
Mixed	2 (4)
Ethnicity[d], n (%)	
Hispanic	3 (5)
Non-Hispanic	48 (87)
Unknown	4 (7)

[a] Missing for relationship to youth participant (n=1).

[b] Missing for marital status (n=4).

[c] Reported races were Dominican and Greek.

[d] Missing for ethnicity (n=1).

Quantitative Results

Dyad-Level Program Engagement

Of the 56 dyads who were given access to the iCanCope with SCD program, differential usage patterns were found. Engagement by both the youth and caregiver occurred for 29% (16/56) of the dyads, whereas it was more likely for only youth to engage in the program (24/56, 43%). In 1 dyad out of 56 (2%), only the caregiver engaged with the intervention, and neither individual engaged for 29% (16/56) of dyads. The 1 youth participant who joined the study without a corresponding caregiver did not engage with the program.

Differential engagement of youth and caregivers with components of the iCanCope with SCD intervention is visualized in Figure 1. Content-specific engagement among youth and caregivers is summarized in Figure 2 and Figure 3, respectively. Among the content categories, the most popular for youth were goal setting, introduction to the program, and history. For caregivers, the most popular content categories were introduction to the program, behavioral plans, and goal setting.

Figure 1. Engagement of youth and caregivers with components of the iCanCope with sickle cell disease (SCD) intervention.

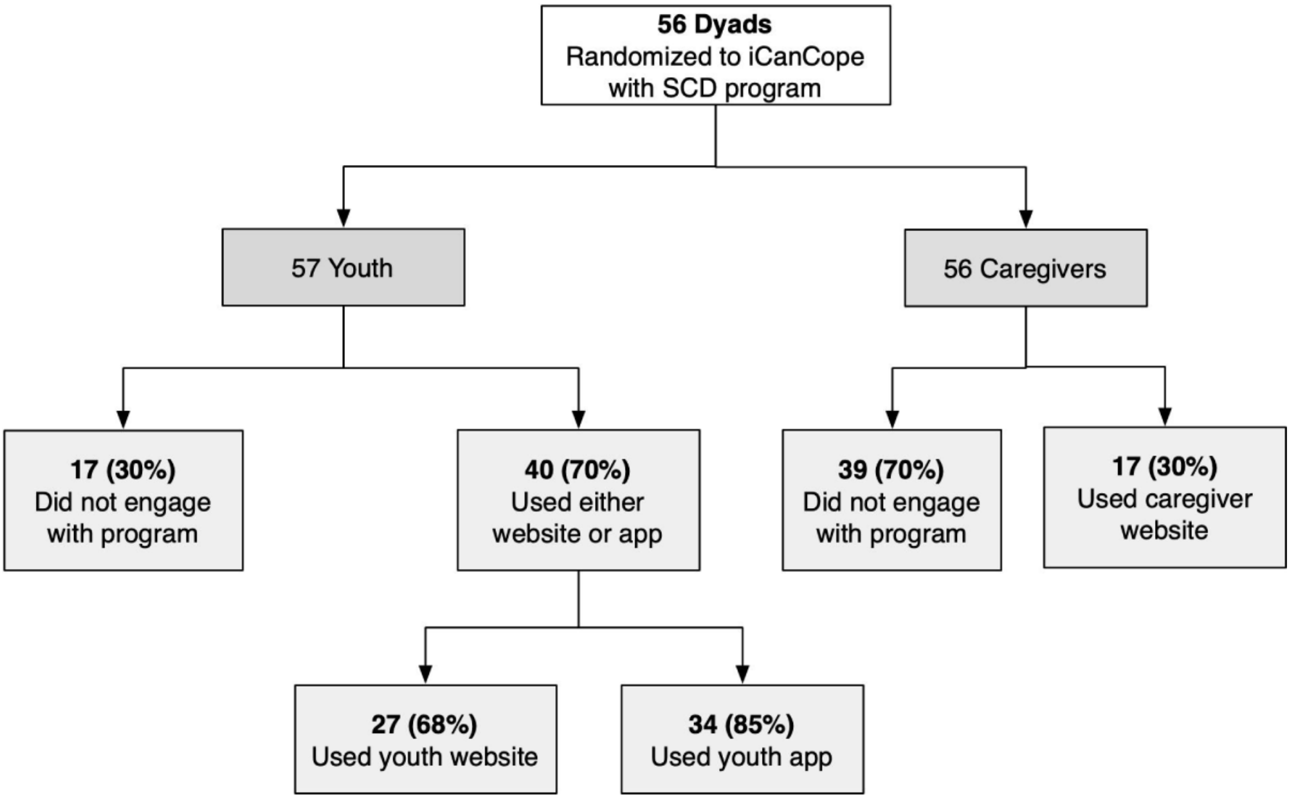


Figure 2. Differential content engagement among youth who used the iCanCope with sickle cell disease program (N=40). SCD: sickle cell disease.

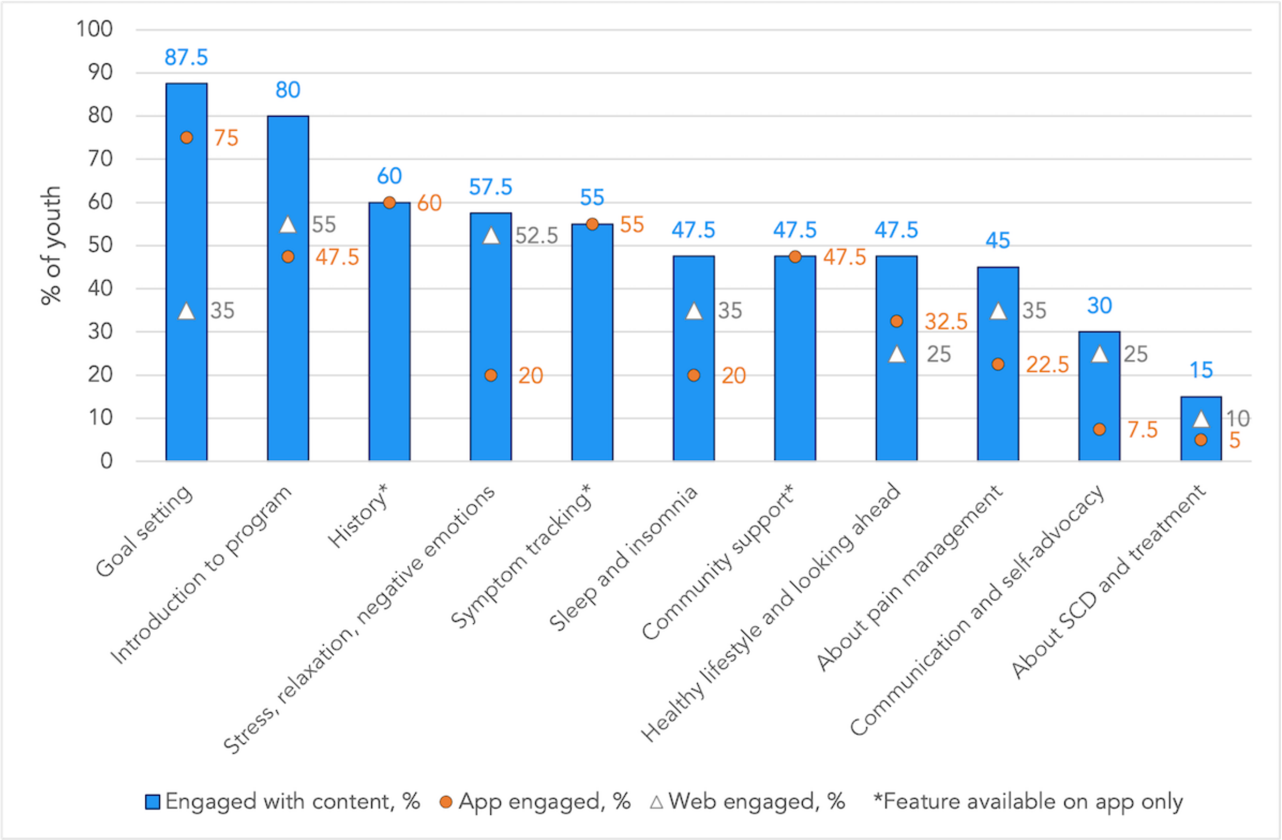
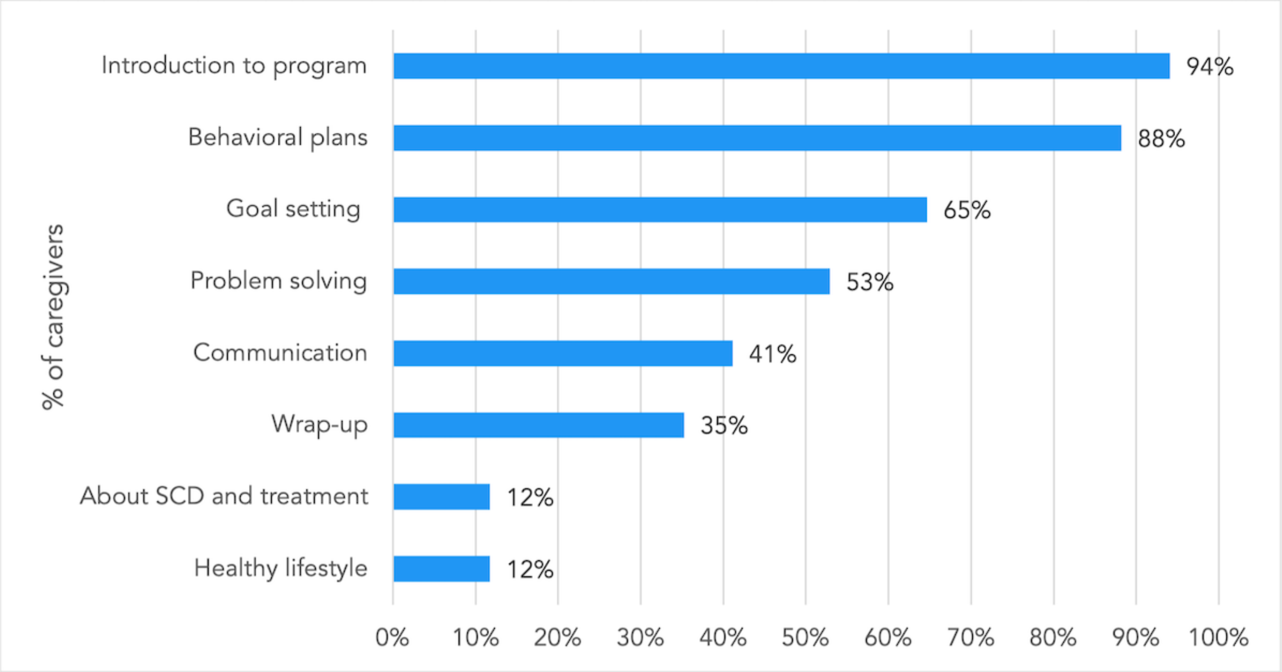


Figure 3. Differential content engagement among caregivers who used the iCanCope with sickle cell disease program (N=17). SCD: sickle cell disease.



Exploratory Predictor of Treatment Engagement

As hypothesized, there was a moderate positive association between caregiver engagement and child engagement with the intervention (χ^2_1=6.6; *P*=.01; φ=0.34).

Qualitative Results

Interview Participants

Individual interviews were completed with 12 dyads between September 2019 and October 2021, representing participants who completed the study either before (5/12, 42%) or after (7/12, 58%) onset of the COVID-19 pandemic. Youth interviewees ranged in age from 12 to 18 years, and most were female (10/12, 83%). All interviewed caregivers were mothers. The interview sample included dyads where both the youth and caregiver engaged with the program (5/12, 42%), only the youth engaged (6/12, 50%), and only the caregiver engaged (1/12, 8%). Among youth interviewees, 11 of 12 (92%) used either the website or app, 7 of 12 (58%) used the website, and 10 of 12 (83%) used the app. Among caregiver interviewees, 6 of 12 (50%) viewed the website content during the study period and the remainder did not engage with the program.

Motivation for Program Use Among Youth

Nearly all youth interviewees (11/12, 92%) indicated that they intended to continue using the iCanCope with SCD program in the future. Moreover, youth who used the app were asked how they would hypothetically rate it on the iOS or Android mobile store between 0 and 5 stars. Among the 10 youth who used the app, 6 (60%) gave a rating of 5 stars and 4 (40%) gave a rating of 4 stars. As rationale for their ratings, 1 participant shared,

> *…I think this is the first app I've heard about sickle cell and I think it's really informative and I like it.* [Youth #126]

Cited motivating factors for wishing to use the program after study completion are summarized in Figure 4.

Figure 4. Motivation for continued use of the iCanCope with sickle cell disease program among youth interviewees.

Tailoring for age group

"…[it] made me feel like it was **made for people my age** because of the **animations** and then you had the **informational side** and you could comment on stuff which was **really fun**… it made me **really comfortable**… it didn't make me feel like I had to do too much. Every time I go on the app, it would be **really calming.**" [Youth 538; age 13; female]

Resource in moments of pain

"…especially if I need it…**if I'm in pain**, I'll probably look at it and see if there's anything **I could refer to**." [Youth 511; age 12; female]

Caregiver encouragement

"The thing that is motivating me to continue using this program is **my mom**." [Youth 337; age 13; female]

Motivating factors for continued program use

Appealing design

"It's really just like a **peaceful** app." [Youth 538; age 13; female]

"I like that I got to make a little character – that was **fun**." [Youth 336; age 16; female]

"They can **heart your posts** [in the community]…it was fun." [Youth 124; age 12; male]

Ease of finding information

"…[the website] had different sections and it had different columns. For example, if I were looking for one specific article, I could go to that specific column, and **I can find it**, so I really like [that]." [Youth 126; age 14; female]

Ability to track symptoms over time

"…when I would set my pain to like a zero, it would either tell me **I'm on track or I'm not** and I really like that about the app." [Youth 126; age 14; female]

Rationale for Program Recommendation Among Caregivers

Most caregiver interviewees (10/12, 83%) indicated that they would recommend the program to a friend or family member living with SCD. One caregiver shared,

> *I have a host of family members that deal with sickle cell, so I would also like to be able to visit the website and provide them with different coping mechanisms or different kind of information that I have found useful.* [Caregiver #337]

Acknowledging the ups and downs of SCD, another caregiver stated,

> *…let me help you cope with this craziness that's going on in your life every day.* [Caregiver #336]

Several caregivers noted that they would be more likely to recommend the program to youth who are newly diagnosed with SCD or have frequent pain crises. For instance, 1 caregiver commented as follows:

> *…for someone that doesn't have a lot of crises, that wouldn't be helpful...for someone that [has] a lot of crises, this would be very helpful.* [Caregiver #537]

Similarly, a caregiver who characterized the program as a good refresher made the following statement:

> *I like the way it's laid out in terms of the content being very layman terms. You know, there's not a lot of – you're not overwhelmed with a lot of medical things that people are saying 'oh, well what's this, and what's that, and what's this?'. So, in terms of that content, I think it's fantastic. So, I would recommend it.* [Caregiver #119]

Exploring Reasons for Low Engagement

Among the 6 of 12 (50%) caregiver interviewees who did not use the program during the study period, the reasons for low engagement included a perceived lack of relevance to their circumstances.

> *…I found that some of the questions on it, things they would ask me - the exercises the [program] would ask me to do was sort of not relevant to me…I felt like it didn't really apply to [my child] or to me.* [Caregiver #127]

This caregiver also noted that another reason for lack of engagement was the large amount of content.

> *…there's so many sections within one unit itself. So, when I thought I was almost done, I still had more to do. I'm able to do only one section at a time or not even able to finish the whole unit because of timing.*
>
> *…when I get home, I'm exhausted. I'm a nurse and I work in ICU so we're very busy at work especially with the COVID…I live really far from work, so I leave home early, I get home really late, and it's just busy that's all but nothing else.*

Among caregivers who did not use the program themselves, some noted that it was helpful for their children.

> *I think it made her more aware and more focused and…process…how she can cope with it and mostly like she's not alone in this… that there are people out there that care. I think it really helped.* [Caregiver #126]

Others noted that the program helped them understand how to better support their children.

> *…it actually shed a lot of light into a couple of things especially when it gets to her mood swings and why she does certain things and also helping her manage her pain.* [Caregiver #112]

One youth interviewee who did not engage with the program did not provide a specific reason why but did indicate interest in using it after the study was complete.

> *…because I feel like it's a good…way to…learn more about it and I could learn and study more myself.* [Youth #520]

Differential Engagement With Program Components

As shown in Figure 1, youth who engaged with the program were more likely to use the app (34/40, 85%) compared with the website (27/40, 68%). Reasons for this preference described by interviewees included perceived ease of use and acceptable time commitment.

> *I liked the app more because it was easier to use than the website…the app was simple, and it only took a couple of minutes.* [Youth #537]

The interactivity of the app feature set was also cited as a reason for use.

> *I like the app more than I like the website because of the daily diaries and connecting to others…and there [is] a library inside the app where you can read and how you can set goals.* [Youth #126]

This user continued with the following statement:

> *I think I have some downsides with the website because the website you only have to read, but the app – it's fun.*

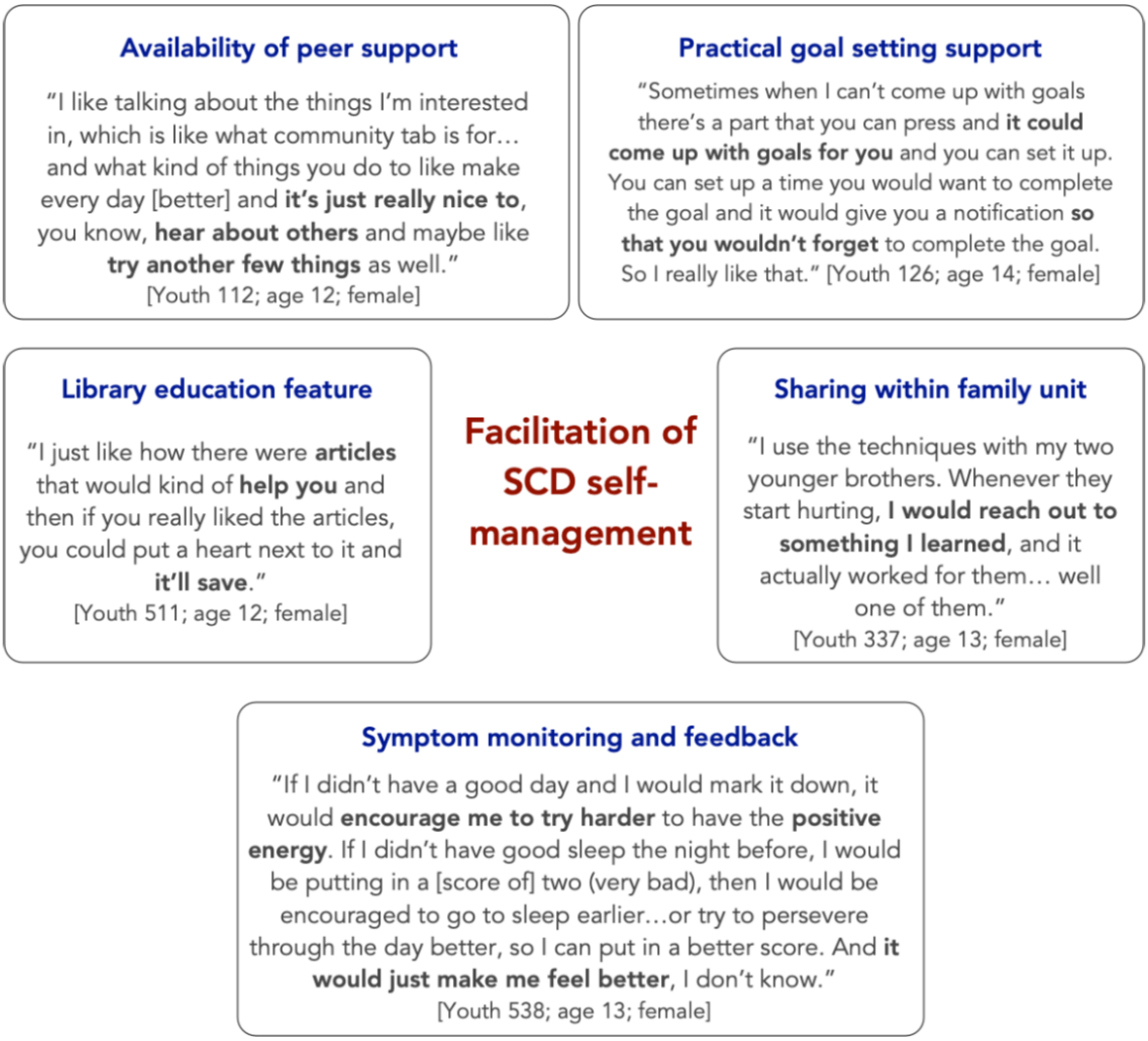
While technical bugs were rare and minimally disruptive (eg, broken external resource link within the youth website and temporary app diary error), technical limitations of the website caused frustration among some participants and may have impacted engagement with that program component. For instance, a caregiver who had low program engagement noted the following:

> *…the website can't save what you write, so every time you come back it does get deleted.* [Caregiver #537]

Promotion of Disease Self-management Behaviors

When asked about how they used the program in their day-to-day life, youth interviewees described several concrete ways that the program supported them to engage in SCD self-management, as summarized in Figure 5.

Figure 5. Examples of how the program facilitated disease self-management. SCD: sickle cell disease.



Availability of peer support

"I like talking about the things I'm interested in, which is like what community tab is for… and what kind of things you do to like make every day [better] and **it's just really nice to**, you know, **hear about others** and maybe like **try another few things** as well." [Youth 112; age 12; female]

Practical goal setting support

"Sometimes when I can't come up with goals there's a part that you can press and **it could come up with goals for you** and you can set it up. You can set up a time you would want to complete the goal and it would give you a notification **so that you wouldn't forget** to complete the goal. So I really like that." [Youth 126; age 14; female]

Library education feature

"I just like how there were **articles** that would kind of **help you** and then if you really liked the articles, you could put a heart next to it and **it'll save**." [Youth 511; age 12; female]

Facilitation of SCD self-management

Sharing within family unit

"I use the techniques with my two younger brothers. Whenever they start hurting, **I would reach out to something I learned**, and it actually worked for them… well one of them." [Youth 337; age 13; female]

Symptom monitoring and feedback

"If I didn't have a good day and I would mark it down, it would **encourage me to try harder** to have the **positive energy**. If I didn't have good sleep the night before, I would be putting in a [score of] two (very bad), then I would be encouraged to go to sleep earlier…or try to persevere through the day better, so I can put in a better score. And **it would just make me feel better**, I don't know." [Youth 538; age 13; female]

Influence of the Pandemic on Self-management

Youth who completed the study during the pandemic described that the program helped to mitigate some disruptions to their self-management routines.

> *I found the app very helpful because for me it can be very hard especially because I'm in high school…now that we're online…I like do my work and…I just stay in my room and I don't do the stuff that I'm supposed to do to ensure that I'm preventing crises. So, having the app, especially the goal chapter, was very helpful because I was able to put what goals I wanted. For example, if I wanted to drink more water or I wanted to get more sleep that week then I was able to do that. That motivated me to want to achieve the goal.* [Youth #127]

Discussion

Principal Findings

This is one of the first studies to characterize patterns of engagement with a web- and app-based pain self-management program among youth with SCD and their caregivers. The data demonstrate that most youth with SCD engaged with the program (40/57, 70%) while most caregivers did not (39/56, 70%). The youth program was delivered via a mobile app and a website. Youth were more likely to engage with the app than the website, and the most popular content categories were goal setting, introduction to the program, and symptom history. Among caregivers who engaged with the website, introduction to the program, behavioral plans, and goal setting were the most popular content areas. Exploratory analysis revealed a significant moderate association between caregiver and youth engagement with the intervention. The key insights from qualitative interviews were as follows: (1) nearly all youth intended to continue using the program; (2) caregivers would recommend the program to others living with SCD who were newly diagnosed or who frequently experienced pain crises; (3) caregivers cited the large amount of content and low relevance to their specific circumstance as reasons for low engagement; (4) youth preferred the app over the website because of perceived greater ease of use, acceptable time commitment, and interactivity; and (5) use of the youth program helped to facilitate SCD self-management behaviors such as goal setting, applying strategies for pain management, symptom trend monitoring, and learning through peer support.

Comparison With Prior Work

There are several app- and web-based interventions designed for children and youth with SCD [33-36] that focus on enhancing SCD knowledge, medication adherence, quality of life, transition to adult care, and pain management. However, engagement or adherence metrics for these interventions were reported in only a few studies. Among 4 recent studies [34-36] that reported some engagement data, operational definitions varied, making direct comparisons difficult. First, Palermo et al reported high engagement with web-based management of adolescent pain (WebMAP), an internet-delivered cognitive-behavioral therapy intervention tailored for SCD [27]. Engagement was defined as participants completing at least one module in the WebMAP program. They found that 14 of 15 (93%) participants met this threshold, which was higher than the proportion of youth (40/57, 70%) who engaged with any iCanCope content. Second, Phillips et al reported engagement with an app-based program called Voice Crisis Alert V2 over a 12-week intervention period [35]. They characterized engagement based on the use of individual app components, similar to the content categorization approach used in our study. Among 60 dyads, symptom tracking and history were the most popular features (used by 90% of dyads), while a clinician messaging feature was least popular (total of 5 message threads generated during the study). In our study, we similarly found that symptom tracking and history were among the most popular content areas. Third, Saulsberry et al offered the Sickle Cell Transition E-Learning Program to 183 youth, of whom, 53 (29%) completed at least one of the 6 available modules [36]. In our study, a higher proportion of youth (70%) who were offered the iCanCope program engaged with the content. Fourth, Leonard et al reported engagement with a medication adherence app as percentage of days that each participant logged medication administration [34]. In a pilot sample of 11 patients, average engagement was 80%. Our study found a wide range of youth engagement with various program components, with the most popular feature being goal setting (35/40, 88%) and the least popular being education about SCD and treatment (6/40, 15%).

The iCanCope program has also been assessed in youth with juvenile idiopathic arthritis and chronic pain [24,25]. Usage levels were similar in the SCD population, although engagement was higher with the symptom history function in the juvenile idiopathic arthritis population [25] at 80% versus 60% among SCD users. A systematic review focusing on engagement and adherence to mHealth interventions in children and youth in clinical and nonclinical populations reported average adherence of 78% [37].

Several contextual factors may explain the differences in engagement patterns seen across studies in the SCD population, among users of other iterations of the iCanCope program, and generally among the pediatric population. These factors include differences in the intervention scope (eg, pain management versus transition readiness), distinct definitions of engagement, different study designs and clinical populations, and technology variation (eg, wearable technology vs app based) [38]. The literature suggests that treatment adherence among youth with SCD may be impacted by forgetfulness, side effects of pharmacological interventions (eg, nausea, heart racing, and taste aversion), questioning efficacy of treatment, and a desire to be “normal,” and some of these factors may similarly impact adherence to digital health interventions [39]. Previous literature also suggests that sociodemographic factors can impact engagement with digital interventions, highlighting lower rates of engagement among racial or ethnic minorities, and those with low socioeconomic status [40-42].

There is limited research on caregiver engagement with digital health interventions for SCD. Many interventions designed for youth do not have a caregiver component. However, 3 recent digital health studies have reported parent engagement in pediatric SCD, chronic pain, and cancer populations, respectively [43,44]. The WebMAP program was designed for pediatric chronic pain patients and adapted for the SCD population. In the chronic pain population, caregiver engagement was higher compared with youth (82% vs 74% completed at least one module) [43]. Similarly, caregiver engagement was also higher (100% vs 93% completed at least one module) than youth engagement in the SCD population [27]. In contrast, the Teens Taking Charge Cancer web-based self-management program [44] reported lower caregiver engagement compared with youth (8% vs 28% completed at least one module), which is consistent with our results. One study on the psychology of eHealth use among caregivers with children having complex health conditions found that factors, such as poor caregiver psychosocial health, high eHealth literacy, and high acceptability of eHealth, were associated with increased engagement [45]. Furthermore, caregiver engagement with nondigital or analog health interventions that report high levels of engagement tend to include methods, such as motivational interviewing and teaching-learning processes (eg, role play and interactive problem solving) [46]. Our program was self-guided, which may have contributed to lower levels of engagement from caregivers.

Strengths and Limitations

Our study had several strengths. First, a phased evidence-based approach was used to develop the iCanCope with SCD program. A qualitative needs assessment involving youth with SCD, caregivers, and health care providers was completed to determine essential components of the program [12]. Next, design sessions were completed with patient partners with SCD to adapt the platform and content for a population with SCD [26]. The changes that were implemented (eg, content tailoring and app design) were cited by participants and their caregivers as positive aspects of the iCanCope program. Second, this study used both qualitative and quantitative methods, which enhanced our ability to understand and contextualize the diverse set of engagement experiences. Furthermore, by including youth and caregiver perspectives, we were able to distinguish their unique experiences and understand how the family unit engaged with the program. Third, this study was a multi-center randomized controlled trial, and this increases the generalizability of the findings to SCD patients in North America. Moreover, the study sample was inclusive of families from a variety of socioeconomic backgrounds.

These findings should be considered in the context of a few limitations. First, caregivers were not heavily included in the design of the iCanCope program, which may partially explain their reduced engagement with the content. As per the user design approach, the creation of iCanCope was focused on meeting the needs of youth, as the primary users were envisioned to be youth, rather than caregivers. However, our study findings indicate that parent engagement with SCD program content is important as it is positively associated with child engagement. Unfortunately, poststudy interviews with caregivers revealed that the program was too content heavy and time-consuming. One strategy to improve caregiver engagement would be to account for the individual need for cognition, which refers to the enjoyment of deep thinking [47]. For those with a low need for cognition, information could be provided in short easily accessible excerpts, such as videos or infographics. In contrast, long-form content would be more suited for individuals with a high need for cognition. Notably, once the program is publicly released, caregivers will be able to review content at their own pace rather than within the time constraints of a research study. In addition, consultations with parents may be completed to fine tune the program to meet their needs prior to public release. Second, differential use of the app versus website among youth suggests a need for changes to the website to increase engagement with the overall program. A specific technical limitation of the website was that it was unable to store the responses of participants, making it difficult for them to gauge their progress. Finally, we were only able to interview 1 youth participant who did not engage with the program at all.

Future Directions

Prior to public release of the program, there are opportunities for refinement using the elements of persuasive design [48]. Wen et al found that more frequent prompts were associated with higher engagement in studies with clinical pediatric populations [37]. Although the iCanCope app did incorporate reminders to prompt participants to engage (eg, push notifications to complete check-ins or set a goal), the frequency and timing of the reminders could be customized. However, it is important to balance the number of reminders while minimizing interruptions to the user. As such, identifying ideal opportunities to use the iCanCope program individualized to each participant is important. A novel approach includes using algorithms to determine these opportunities, which have been developed for adult populations [45,49]. In tandem, providing praise (eg, congratulatory messaging) and rewards when participants complete a check-in, read an article from the library, or complete a goal may influence engagement positively [48]. Rewards can include elements of gamification where participants gain points or virtual rewards for completing certain tasks, which they can use to promote engagement with self-management content. Finally, real-world use of the iCanCope program can also be studied in the future, after public release.

In terms of program dissemination, using strategies to target the family unit, rather than individual members, may increase engagement as SCD typically impacts multiple members of a family. This approach is supported by youth participants who reported *ad hoc* sharing of program content with their siblings and the finding that parent and child engagement was positively associated. Furthermore, by partnering with SCD clinics, the program could be introduced to youth and families when the SCD pain becomes more frequent to help cope with the corresponding increase in pain burden.

Conclusions

This is one of the first studies to apply digital health analytics to characterize patterns of engagement with SCD self-management among youth and caregivers. Differential engagement patterns were found, with more youth engaging in the program than their caregivers; however, youth were more likely to engage with the app than the website. These findings will be used to optimize the iCanCope with SCD program prior to release, with the potential to improve the accessibility and acceptability of pain self-management support for families affected by SCD.

Acknowledgments

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R01HD086978. This project was also partially supported by a grant (K23HL150232, principal investigator: SB) from the National Heart, Lung, and Blood Institute of the National Institutes of Health only for the protected research time that allowed the author to contribute to this work. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

iCanCope with sickle cell disease program screenshots.

[DOCX File , 697 KB - jmir_v24i8e40096_app1.docx]

References

1. Hassell KL. Population estimates of sickle cell disease in the U.S. Am J Prev Med 2010 Apr;38(4 Suppl):S512-S521. [doi: 10.1016/j.amepre.2009.12.022] [Medline: 20331952]
2. Dampier C, Palermo TM, Darbari DS, Hassell K, Smith W, Zempsky W. AAPT diagnostic criteria for chronic sickle cell disease pain. J Pain 2017 May;18(5):490-498 [FREE Full text] [doi: 10.1016/j.jpain.2016.12.016] [Medline: 28065813]
3. Bonner MJ. Health related quality of life in sickle cell disease: just scratching the surface. Pediatr Blood Cancer 2010 Jan;54(1):1-2. [doi: 10.1002/pbc.22301] [Medline: 19785024]
4. Keenan ME, Loew M, Berlin KS, Hodges J, Alberts NM, Hankins JS, et al. Empirically derived profiles of health-related quality of life in youth and young adults with sickle cell disease. J Pediatr Psychol 2021 Mar 18;46(3):293-303 [FREE Full text] [doi: 10.1093/jpepsy/jsaa104] [Medline: 33249456]
5. Badawy SM, Thompson AA, Lai J, Penedo FJ, Rychlik K, Liem RI. Health-related quality of life and adherence to hydroxyurea in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2017 Jun;64(6):26369. [doi: 10.1002/pbc.26369] [Medline: 27896936]
6. Zempsky WT, O'Hara EA, Santanelli JP, Palermo TM, New T, Smith-Whitley K, et al. Validation of the sickle cell disease pain burden interview-youth. J Pain 2013 Sep;14(9):975-982 [FREE Full text] [doi: 10.1016/j.jpain.2013.03.007] [Medline: 23701707]
7. Brousseau DC, Owens PL, Mosso AL, Panepinto JA, Steiner CA. Acute care utilization and rehospitalizations for sickle cell disease. JAMA 2010 Apr 07;303(13):1288-1294. [doi: 10.1001/jama.2010.378] [Medline: 20371788]
8. Dampier C, Ely B, Brodecki D, O'Neal P. Characteristics of pain managed at home in children and adolescents with sickle cell disease by using diary self-reports. J Pain 2002 Dec;3(6):461-470. [doi: 10.1054/jpai.2002.128064] [Medline: 14622732]
9. Alberts NM, Kang G, Li C, Richardson PA, Hodges J, Hankins JS, et al. Pain in youth with sickle cell disease: A report from the sickle cell clinical research and intervention program. Clin J Pain 2021 Jan;37(1):43-50. [doi: 10.1097/AJP.0000000000000889] [Medline: 33093339]
10. Smith WR, Scherer M. Sickle-cell pain: advances in epidemiology and etiology. Hematology Am Soc Hematol Educ Program 2010;2010:409-415. [doi: 10.1182/asheducation-2010.1.409] [Medline: 21239827]
11. Dampier C, Ely E, Brodecki D, O'Neal P. Home management of pain in sickle cell disease: a daily diary study in children and adolescents. J Pediatr Hematol Oncol 2002 Nov;24(8):643-647. [doi: 10.1097/00043426-200211000-00008] [Medline: 12439036]
12. Kulandaivelu Y, Lalloo C, Ward R, Zempsky WT, Kirby-Allen M, Breakey VR, et al. Exploring the needs of adolescents with sickle cell disease to inform a digital self-management and transitional care program: Qualitative study. JMIR Pediatr Parent 2018 Sep 25;1(2):e11058 [FREE Full text] [doi: 10.2196/11058] [Medline: 31518307]
13. Lanzkron S, Carroll CP, Haywood C. The burden of emergency department use for sickle-cell disease: an analysis of the national emergency department sample database. Am J Hematol 2010 Oct;85(10):797-799 [FREE Full text] [doi: 10.1002/ajh.21807] [Medline: 20730795]
14. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
15. Fisher E, Heathcote L, Palermo TM, de C Williams AC, Lau J, Eccleston C. Systematic review and meta-analysis of psychological therapies for children with chronic pain. J Pediatr Psychol 2014 Sep;39(8):763-782 [FREE Full text] [doi: 10.1093/jpepsy/jsu008] [Medline: 24602890]
16. Fisher E, Law E, Dudeney J, Palermo T, Stewart G, Eccleston C. Psychological therapies for the management of chronic and recurrent pain in children and adolescents. Cochrane Database Syst Rev 2018 Sep 29;9(9):CD003968 [FREE Full text] [doi: 10.1002/14651858.CD003968.pub5] [Medline: 30270423]
17. Palermo T, Eccleston C, Lewandowski A, de C Williams AC, Morley S. Randomized controlled trials of psychological therapies for management of chronic pain in children and adolescents: an updated meta-analytic review. Pain 2010 Mar;148(3):387-397 [FREE Full text] [doi: 10.1016/j.pain.2009.10.004] [Medline: 19910118]
18. Chen E, Cole SW, Kato PM. A review of empirically supported psychosocial interventions for pain and adherence outcomes in sickle cell disease. J Pediatr Psychol 2004;29(3):197-209. [doi: 10.1093/jpepsy/jsh021] [Medline: 15131137]
19. Kayle M, Docherty SL, Sloane R, Tanabe P, Maslow G, Pan W, et al. Transition to adult care in sickle cell disease: A longitudinal study of clinical characteristics and disease severity. Pediatr Blood Cancer 2019 Jan;66(1):e27463 [FREE Full text] [doi: 10.1002/pbc.27463] [Medline: 30251318]
20. Barakat LP, Schwartz LA, Salamon KS, Radcliffe J. A family-based randomized controlled trial of pain intervention for adolescents with sickle cell disease. J Pediatr Hematol Oncol 2010 Oct;32(7):540-547 [FREE Full text] [doi: 10.1097/MPH.0b013e3181e793f9] [Medline: 20686425]
21. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a pain app for that": Review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
22. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]

23. Stinson JN, Lalloo C, Harris L, Isaac L, Campbell F, Brown S, et al. iCanCope with Pain™: User-centred design of a web- and mobile-based self-management program for youth with chronic pain based on identified health care needs. Pain Res Manag 2014;19(5):257-265 [FREE Full text] [doi: 10.1155/2014/935278] [Medline: 25000507]
24. Lalloo C, Hundert A, Harris L, Pham Q, Campbell F, Chorney J, et al. Capturing daily disease experiences of adolescents with chronic pain: mHealth-mediated symptom tracking. JMIR Mhealth Uhealth 2019 Jan 17;7(1):e11838 [FREE Full text] [doi: 10.2196/11838] [Medline: 30664472]
25. Lalloo C, Harris LR, Hundert AS, Berard R, Cafazzo J, Connelly M, et al. The iCanCope pain self-management application for adolescents with juvenile idiopathic arthritis: a pilot randomized controlled trial. Rheumatology (Oxford) 2021 Jan 05;60(1):196-206. [doi: 10.1093/rheumatology/keaa178] [Medline: 32613229]
26. Palermo TM, Zempsky WT, Dampier CD, Lalloo C, Hundert AS, Murphy LK, et al. iCanCope with Sickle Cell Pain: Design of a randomized controlled trial of a smartphone and web-based pain self-management program for youth with sickle cell disease. Contemp Clin Trials 2018 Nov;74:88-96 [FREE Full text] [doi: 10.1016/j.cct.2018.10.006] [Medline: 30316999]
27. Palermo TM, Dudeney J, Santanelli JP, Carletti A, Zempsky WT. Feasibility and acceptability of internet-delivered cognitive behavioral therapy for chronic pain in adolescents with sickle cell disease and their parents. J Pediatr Hematol Oncol 2018 Mar;40(2):122-127 [FREE Full text] [doi: 10.1097/MPH.0000000000001018] [Medline: 29176462]
28. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: Scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
29. Pham Q, Graham G, Lalloo C, Morita PP, Seto E, Stinson JN, et al. An analytics platform to evaluate effective engagement with pediatric mobile health apps: Design, development, and formative evaluation. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11447 [FREE Full text] [doi: 10.2196/11447] [Medline: 30578179]
30. McHugh M. Phi correlation coefficient. In: Frey BB, editor. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation. Thousand Oaks, CA: SAGE Publications, Inc; 2018:1252-1253.
31. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
32. Creswell JW, Poth CN. Qualitative Inquiry and Research Design Choosing Among Five Approaches. Thousand Oaks, CA: SAGE Publications, Inc; 2012.
33. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered eHealth interventions for children, adolescents, and adults with sickle cell disease: Systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
34. Leonard S, Anderson LM, Jonassaint J, Jonassaint C, Shah N. Utilizing a novel mobile health "selfie" application to improve compliance to iron chelation in pediatric patients receiving chronic transfusions. J Pediatr Hematol Oncol 2017 Apr;39(3):223-229. [doi: 10.1097/MPH.0000000000000743] [Medline: 28099398]
35. Phillips S, Kanter J, Mueller M, Gulledge A, Ruggiero K, Johnson M, et al. Feasibility of an mHealth self-management intervention for children and adolescents with sickle cell disease and their families. Transl Behav Med 2021 Apr 07;11(3):724-732 [FREE Full text] [doi: 10.1093/tbm/ibaa132] [Medline: 33410488]
36. Saulsberry AC, Hodges JR, Cole A, Porter JS, Hankins J. Web-based technology to improve disease knowledge among adolescents with sickle cell disease: Pilot study. JMIR Pediatr Parent 2020 Jan 07;3(1):e15093 [FREE Full text] [doi: 10.2196/15093] [Medline: 31909718]
37. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
38. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
39. Modi AC, Crosby LE, Guilfoyle SM, Lemanek KL, Witherspoon D, Mitchell MJ. Barriers to treatment adherence for pediatric patients with sickle cell disease and their families. Children's Health Care 2009 Apr 10;38(2):107-122. [doi: 10.1080/02739610902813245]
40. Senecal C, Widmer RJ, Bailey K, Lerman LO, Lerman A. Usage of a digital health workplace intervention based on socioeconomic environment and race: Retrospective secondary cross-sectional study. J Med Internet Res 2018 Apr 23;20(4):e145 [FREE Full text] [doi: 10.2196/jmir.8819] [Medline: 29685862]
41. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample Of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
42. Grossman L, Masterson Creber RM, Benda N, Wright D, Vawdrey D, Ancker J. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]

43. Palermo TM, de la Vega R, Murray C, Law E, Zhou C. A digital health psychological intervention (WebMAP Mobile) for children and adolescents with chronic pain: results of a hybrid effectiveness-implementation stepped-wedge cluster randomized trial. Pain 2020 Dec 10;161(12):2763-2774 [FREE Full text] [doi: 10.1097/j.pain.0000000000001994] [Medline: 32658147]
44. Breakey V, Gupta A, Johnston D, Portwine C, Laverdiere C, May S, et al. A pilot randomized control trial of teens taking charge: A web-based self-management program for adolescents with cancer. J Pediatr Hematol Oncol Nurs 2022 Jan 21:27527530211068778. [doi: 10.1177/27527530211068778] [Medline: 35759365]
45. Sarkar M, Sanders LM, Kelleher KJ, Chisolm DJ. Psychosocial health, e-Health literacy, and perceptions of e-Health as predictors and moderators of e-Health use among caregivers of children with special healthcare needs. Telemed J E Health 2016 Feb;22(2):123-131 [FREE Full text] [doi: 10.1089/tmj.2015.0028] [Medline: 26295487]
46. Dishion T, Forgatch M, Chamberlain P, Pelham WE. The Oregon model of behavior family therapy: From intervention design to promoting large-scale system change. Behav Ther 2016 Nov;47(6):812-837 [FREE Full text] [doi: 10.1016/j.beth.2016.02.002] [Medline: 27993335]
47. Short C, Rebar A, Plotnikoff R, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. The European Health Psychologist 2015;17(1):32-38 [FREE Full text]
48. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. 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]
49. Hayes TL, Cobbinah K, Dishongh T, Kaye JA, Kimel J, Labhard M, et al. A study of medication-taking and unobtrusive, intelligent reminding. Telemed J E Health 2009 Oct;15(8):770-776 [FREE Full text] [doi: 10.1089/tmj.2009.0033] [Medline: 19780692]

Abbreviations

SCD: sickle cell disease
WebMAP: web-based management of adolescent pain

Edited by T Leung; submitted 06.06.22; peer-reviewed by S Stones, S Easton; comments to author 30.06.22; revised version received 19.07.22; accepted 31.07.22; published 30.08.22.

Please cite as:
Lalloo C, Nishat F, Zempsky W, Bakshi N, Badawy S, Ko YJ, Dampier C, Stinson J, Palermo TM
Characterizing User Engagement With a Digital Intervention for Pain Self-management Among Youth With Sickle Cell Disease and Their Caregivers: Subanalysis of a Randomized Controlled Trial
J Med Internet Res 2022;24(8):e40096
URL: https://www.jmir.org/2022/8/e40096
doi:10.2196/40096
PMID:36040789

Original Paper

The Interplay of Work, Digital Health Usage, and the Perceived Effects of Digitalization on Physicians' Work: Network Analysis Approach

Petra Saukkonen[1], MSc; Marko Elovainio[1,2], PhD; Lotta Virtanen[1], BSc; Anu-Marja Kaihlanen[1], PhD; Janna Nadav[1], MSc; Tinja Lääveri[3,4], MD, PhD; Jukka Vänskä[5], MSocSc; Johanna Viitanen[4], DSc; Jarmo Reponen[6,7], MD, PhD; Tarja Heponiemi[1], PhD

[1]Finnish Institute for Health and Welfare, Helsinki, Finland
[2]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland
[3]Infectious Diseases and Meilahti Vaccine Research Center MeVac, Inflammation Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[4]Department of Computer Science, Aalto University, Espoo, Finland
[5]Finnish Medical Association, Helsinki, Finland
[6]Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland
[7]Medical Research Centre Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland

Corresponding Author:
Petra Saukkonen, MSc
Finnish Institute for Health and Welfare
Mannerheimintie 166
Helsinki, PO Box 30
Finland
Phone: 358 29 524 8289
Email: petra.saukkonen@thl.fi

Abstract

Background: In health care, the benefits of digitalization need to outweigh the risks, but there is limited knowledge about the factors affecting this balance in the work environment of physicians. To achieve the benefits of digitalization, a more comprehensive understanding of this complex phenomenon related to the digitalization of physicians' work is needed.

Objective: The aim of this study was to examine physicians' perceptions of the effects of health care digitalization on their work and to analyze how these perceptions are associated with multiple factors related to work and digital health usage.

Methods: A representative sample of 4630 (response rate 24.46%) Finnish physicians (2960/4617, 64.11% women) was used. Statements measuring the perceived effects of digitalization on work included the patients' active role, preventive work, interprofessional cooperation, decision support, access to patient information, and faster consultations. Network analysis of the perceived effects of digitalization and factors related to work and digital health usage was conducted using mixed graphical modeling. Adjusted and standardized regression coefficients are denoted by b. Centrality statistics were examined to evaluate the relative influence of each variable in terms of node strength.

Results: Nearly half of physicians considered that digitalization has promoted an active role for patients in their own care (2104/4537, 46.37%) and easier access to patient information (1986/4551, 43.64%), but only 1 in 10 (445/4529, 9.82%) felt that the impact has been positive on consultation times with patients. Almost half of the respondents estimated that digitalization has neither increased nor decreased the possibilities for preventive work (2036/4506, 45.18%) and supportiveness of clinical decision support systems (1941/4458, 43.54%). When all variables were integrated into the network, the most influential variables were purpose of using health information systems, employment sector, and specialization status. However, the grade given to the electronic health record (EHR) system that was primarily used had the strongest direct links to faster consultations (b=0.32) and facilitated access to patient information (b=0.28). At least 6 months of use of the main EHR was associated with facilitated access to patient information (b=0.18).

Conclusions: The results highlight the complex interdependence of multiple factors associated with the perceived effects of digitalization on physicians' work. It seems that a high-quality EHR system is critical for promoting smooth clinical practice. In addition, work-related factors may influence other factors that affect digital health success. These factors should be considered when developing and implementing new digital health technologies or services for physicians' work. The adoption of digital health is not just a technological project but a project that changes existing work practices.

(J Med Internet Res 2022;24(8):e38714) doi:10.2196/38714

KEYWORDS

network analysis; mixed graphical model; physicians; health care digitalization; digitalization of work; work in transformation; digital health

Introduction

Background

Digital transformation is rapidly changing the health care sector, and the COVID-19 pandemic has accelerated this transition to digital solutions [1]. The digital transformation of health care is expected to enhance health outcomes by improving person-centered care and self-management, data-driven treatment decisions, and medical diagnoses as well as creating more evidence-based knowledge, skills, and competencies for professionals to support health care delivery [2]. Physicians are one of the most important stakeholders in health care, and they have the potential to shape this change for the benefit of clinical care [3].

The digital transformation of workplace can be defined as a phenomenon in which new technologies significantly change the way employees perform tasks and processes, their social relationships, and subsequently their overall workplace experience [4]. Indeed, physicians have seen digital health as a dynamic facet of new ways of working [5]. Digital health is the field of knowledge and practice related to the development and use of digital technologies to improve health [2]. The broad scope of digital health includes categories such as health information systems (HISs; including, eg, electronic health record [EHR] systems and clinical decision support systems [CDSSs]) as well as telemedicine, wearable devices, mobile health, and personalized medicine [6].

Digital health can provide additional work processes next to existing ones or completely replace current processes [7]. It also changes the culture toward shared decision-making and the democratization of care [8]. Patients are suggested to no longer be just customers but active participants in their own care processes [5,8]. Digital health can empower patients to advocate for themselves, take control of their care, and make better-informed decisions about their health [9-11]. In a variety of settings, digital interventions can be effective in both preventing and treating disease [12-17]. In addition, digital health appears to impact interprofessional cooperation [18-23]. EHRs influence cooperation by facilitating access to patient information and data sharing between different stakeholders and hospitals [21,24,25]. There is also a need for effective cooperation between information technology professionals and physicians to improve the quality and implementation of HISs [26]. Participation in development may also increase one's sense of control over work [27]. Technology such as CDSSs can support physicians by minimizing errors [28], improving the accuracy of physician diagnoses [29] and outcomes [28,30,31], and increasing efficiency [29,32]. Physicians have described the greatest benefits of digitalization in terms of care quality, readability, and ease of access to patient data [33,34]. In addition, digitalization has been shown to support collaboration, decision-making, and continuous learning [5], and it has been associated with improved job satisfaction and work-life balance [35].

However, the digital transformation of health care is a highly multifaceted issue [7]. Physicians have expressed concerns about the impact of digitalization on information overload and ambiguity, interaction with patients, privacy issues, disruptions to workflows, and increasing workloads [5,34,36-39]. The digitalization of work has also been found to be associated with the stress levels of physicians [40,41]. Dissatisfaction has been particularly associated with the implementation of new EHRs [42-45] and the subsequent transition period [46].

Reports on the impact of digitalization on work performance have been inconclusive [34]. Digitalization potentially affects and is affected not only by the characteristics of the digital health used but also by the physical and psychosocial work environment. It is well known that work characteristics have an important impact on employee attitudes toward the digital workplace transformation [4]. Moreover, the various effects of digitalization are likely to be interconnected. Although physicians' perceived benefits of digitalization appear to outweigh the risks [5], there is limited knowledge about the factors affecting this balance. To achieve the benefits of digitalization, a more comprehensive understanding of this complex phenomenon related to the digitalization of physicians' work is required.

To date, conceptual and statistical tools to analyze and illustrate such complexities have been lacking. However, the recently introduced psychological network approach offers a promising methodology to address the interplay between multiple factors in multiple areas [47]. To understand complex phenomena, it is often insufficient to only focus on how the individual components of a system function. Instead, one must also study the organization of the system's components, which can be represented in a network. In this field of research, psychosocial, organizational, and behavioral entities are conceptualized as an interplay of social, psychological, and other components that interact in a network consisting of nodes representing observed variables and connected by edges representing statistical relationships [47-51].

Goal of This Study

The aim of this study was to examine physicians' perceptions of the effects of health care digitalization on their work and to analyze how these perceptions are associated with multiple factors related to work and digital health usage. With this information, it is possible to further develop digital health to meet the needs of clinical practice. In addition, the information can be used to improve the understanding of the changing nature of clinicians' work and enable organizations to develop physicians' orientation, promote staff empowerment and well-being, and improve the quality of care services. The research questions (RQs) were as follows:

- RQ1: How has the digitalization of health care affected the work of physicians from their perspective?
- RQ2: How are (1) the effects of digitalization (patients' active role, preventive work, interprofessional cooperation, decision support, access to patient information, and faster consultations), (2) factors related to work (purpose of using the HIS, employment sector, and specialization status), and (3) factors related to digital health usage (EHR experience, EHR grade, participation in HIS development, and telemedicine) connected in the network structure?

Methods

Study Sample

The nationwide survey *EHR systems as a tool for physicians 2021* was conducted in Finland as part of the national STePS 3.0 project [52,53]. The survey method and questionnaire have been described in detail elsewhere [54]. The data were collected between January and March 2021. An invitation to participate in the web-based survey was sent by email to all physicians of working age (<65 years) who had provided their email addresses to the Finnish Medical Association (n=19,142). The register represents 90.51% (19,142/21,148 of all working age physicians who live in Finland [55]. We received responses from 4683 physicians (response rate 4683/19,142, 24.46%). A total of 53 responses were removed from the data because the respondents reported that they did not use health care information systems at all (n=43) or they did not respond to this question (n=10). The final sample included 4630 physicians (2960/4617, 64.11% women) who worked directly with patients, in administration, or both. Information on the status of clinical work was no longer included in the registry, so the number of clinically active physicians is an estimate based on several data sources and the expertise of researchers at the Finnish Medical Association. According to their analyses, the respondents were representative of the population. However, older physicians responded slightly more often than younger physicians, as did specialists. The hospital sector was also slightly overrepresented [56].

Ethics Approval

According to Finnish legislation, a statement from the ethics committee is not required to conduct surveys on respondents' opinions [57]. Participation in the survey on the EHR systems as a tool for physicians in 2021 was voluntary. All participants provided informed consent by choosing to participate actively in the study by answering the questionnaire.

Context

Finland is one of the leading countries in digitalization, ranking first in a comparison of digitalization levels across European Union member states [58]. The public sector has the primary responsibility for health services, which are complemented by private sector services [59]. Private service providers deliver a quarter of all social and health services [60]. Almost half (49%) of Finnish physicians work in hospitals, a quarter (25%) in health centers, and 16% in the public sector [55].

EHRs are widely deployed across Finland [61]. As early as 2007, EHR coverage in the public sector in Finland reached 100%, and in 2017, all hospital districts achieved 100% usage intensity in conservative, operative, and psychiatric care. In 2020, 91% of private actors reported usage intensities above 90% [61]. All EHRs are integrated with national health information exchange services (Kanta services), which comprise My Kanta Pages, Prescription service, Pharmaceutical database, Patient Data Repository, and archiving of old patient data [62]. Participation in Kanta services is mandatory for all public health care providers. Private providers are also required to join Kanta services if the organization archives its clients and medical records in an electronic form [62]. In addition, there are several ancillary systems, for example, radiology and laboratory information systems, and HISs for operating rooms, intensive care units, labor and delivery, and emergency departments [63,64]. However, due to suboptimal integration solutions within the organizations, insufficient data structures in Kanta services, and barriers set by legislation, patient information is not always readily usable across different sectors, organizations, or facilities [61]. Comprehensive digital services are also made available to patients by solutions such as patient portals, access to their own data in the Kanta eArchive via My Kanta Pages, digital symptom checkers, and digital self-management guides [65]. Telemedicine with patients via video visits or chat messages, as well as self-recorded health data and monitoring services, already increased before the COVID-19 pandemic [61].

Measurements

The Effects of Digitalization on Work

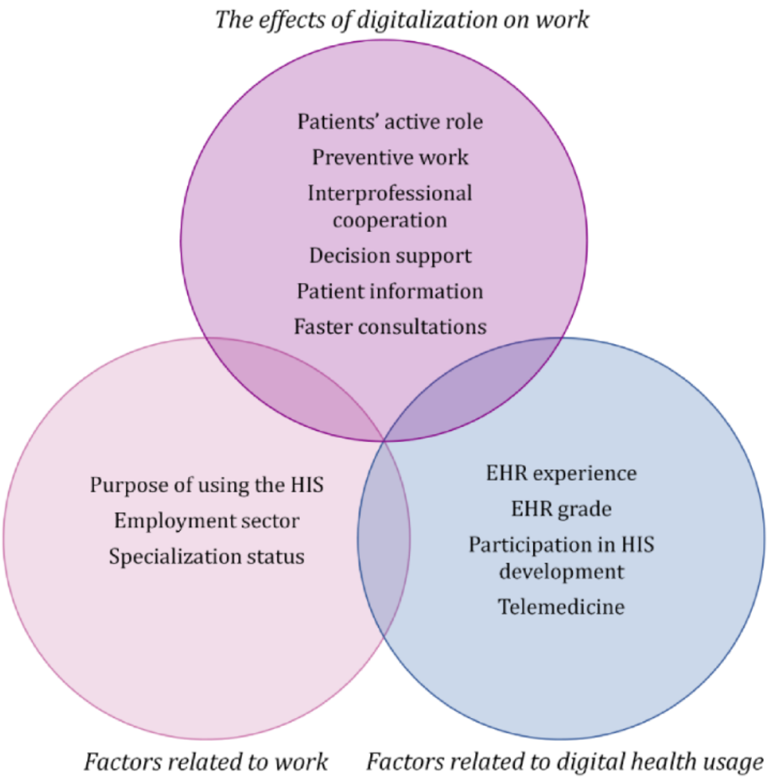
All variables used are presented in Figure 1 and Multimedia Appendix 1. The measurement of the perceived effects of digitalization on work was based on the strategic focus areas of the Finnish eHealth and eSocial Strategy 2020 [66]. The survey questionnaire included 6 statements related to physicians' experiences with digitalization. Respondents were asked how the digitalization of health care has affected their work and asked to assess the change during the last 3 years.

The 6 statements were as follows:

1. Patients have assumed a more active role in their treatment (patients' active role).
2. Possibilities for preventive work have improved (*preventive work*).
3. Interprofessional cooperation has progressed (*interprofessional cooperation*).
4. Intelligent decision support systems support a physician's work (*decision support*).
5. It has become easier to obtain information on patients (*patient information*).
6. Consultations with patients have become faster (*faster consultations*).

Response options were rated on a 5-point Likert scale ranging from 1 (fully agree) to 5 (fully disagree). For the analysis, the responses were reversed from 1 (fully disagree) to 5 (fully agree). Variables were recoded for the descriptive results, as (1) somewhat or fully disagree (response options 1-2), (2) neither agree nor disagree (response option 3), and (3) somewhat or fully agree (response options 4-5). In the network analysis, each statement was used separately as a continuous variable.

Figure 1. Variables used in this study. EHR: electronic health record; HIS: health information system.



Work-Related Factors

Work-related factors included purpose of using the HIS, employment sector, and specialization status.

Purpose of using the HIS was assessed by asking whether respondents used the HIS (1) for patient work, (2) for administrative work, (3) for both, or (4) not at all. The variable was coded as a dichotomous variable: 0=uses not only for work with patients (response options 2-3) and 1=uses only for work with patients (response option 1). Response option 4 was coded as missing.

Employment sector was assessed with the following response options: (1) municipality, (2) state, (3) private (including The Social Insurance Institution of Finland [Kela]), (4) university, and (5) I am not employed. The variable was coded as a dichotomous variable: 0=public (response options 1-2) and 1=private (response options 3-4).

Specialization status was assessed using response options (1) not specialized, (2) in specialist training, and (3) specialized and coded as a dichotomous variable: 0=not specialized (response options 1 and 2) and 1=specialized.

Factors Related to Digital Health Usage

The factors related to digital health usage included EHR experience, EHR grade, participation in HIS development, and telemedicine.

EHR experience was assessed by asking respondents how long they had used the EHR system that they mainly use in their employment. The response options were (1) less than 6 months, (2) 6 months—less than a year, (3) 1 to 3 years, (4) 4 to 6 years, and (5) more than 6 years. The variable was coded as a dichotomous variable: 0=less than 6 months and 1=6 months or more.

The EHR grade was assessed by asking, "On a scale of 4 to 10 (with 4 being the lowest score and 10 being the highest score), how would you rate the EHR you mainly use?" Response options ranged from 4 to 10, including a response option of "I am not able to give a grade, or I do not wish to answer." The variable was coded as a dichotomous variable: 0=4 to 7 (low grade) and 1=8 to 10 (high grade). The response option "I am not able to give a grade, or I do not wish to answer" was coded as missing.

Participation in HIS development was assessed by asking whether the respondent had participated in HIS development activities. The response options were (1) yes, some of my working time has been allocated for such development work; (2) yes, in addition to my work; and (3) no. The variable was recoded as a dichotomous variable: 0=no and 1=yes (response options 1-2).

Telemedicine was assessed by asking whether the respondent's main employment involved telemedicine with patients (remote treatment by phone, chat, video contact, and other electronic contact). The response options were (1) very much, (2) much, (3) some, (4) a little, and (5) not at all. The variable was recoded as a dichotomous variable: 0=little or not at all (response options 4-5) and 1=somewhat to very much (response options 1-3).

Statistical Analysis

Descriptive analyses were performed using SPSS (IBM SPSS Statistics 27) to characterize the sample characteristics and the variables used. Owing to nonresponse in some items, the number of observations varied in the descriptive analyses. Subsequently, network analysis was performed using mixed graphical models [67] in R Statistical Software (version 4.1.1; R Foundation for Statistical Computing 2020) to estimate the associations between the perceived effects of digitalization, work-related factors, and factors related to digital health usage. The scale-based effects of digitalization were modeled as continuous variables: patients' active role, preventive work, interprofessional cooperation, decision support, patient information, and faster consultations. The skewed distributions of these variables were normalized using the nonparanormal transformation (huge.npn function) [68]. The following binary work-related factors were modeled as 2-level categorical variables: purpose of using the HIS, employment sector, and specialization status. The following binary digital health usage–related factors were also modeled as 2-level categorical variables: EHR experience, EHR grade, participation in HIS development, and telemedicine.

As the data contained both continuous and binary variables, we estimated the main network with mixed graphical models using the mgm package (version 1.2.12) [67]. The package estimates a network model by running regularized generalized regressions on each variable and estimating the edges associated with that variable. We provided the data, removed the missing values, and specified the type and number of levels for each variable. The regularization parameter λ was selected by 10-fold cross-validation, and the parameter k was set to 2 only to estimate the pairwise relationships. The computed relationships were represented in undirected graphical models [69] and visualized using the qgraph package [70]. Each variable was represented as a node in the network and pairwise connections between variables were represented as edges. The adjusted and standardized regression coefficients are denoted by b in the text. We added the strength of the dependencies by the width of the edges and information about the sign of the edges: green and blue edges indicate positive relationships and red edges indicate negative relationships. Two different colors indicating positive edges (green and blue) were used for illustration purposes.

Centrality statistics for the networks were examined to assess the relative influence of each factor in the network in terms of standardized node strength (the sum of edge weights associated with a given node) [47,71]. The predict function in mgm was used to obtain estimates of predictability for each factor. Predictability refers to the extent to which the variable can be explained by other variables included in the network [72]. Estimates are reported on a scale from 0 to 1, with 1 reflecting complete predictability. To evaluate the stability and accuracy test of the main network, the bootnet package [71] with the mgm specification was used to compute nonparametric bootstrap intervals around the estimated network edges and significance tests for edge differences using 1000 bootstrap samples.

Results

Characteristics of the Study Population

The characteristics of the sample are shown in Table 1. The majority (3668/4630, 79.22%) used HISs exclusively to work with patients. Most participants (3676/4614, 79.67%) worked in the public sector, and more than two-third (3134/4630, 67.69%) were specialists. For nearly half of the respondents (2177/4625, 47.07%), their work involved at least some telemedicine with patients. In total, 15.77% (727/4611) of the respondents had less than 6 months of experience with their current EHR. Nearly two-third (2974/4610, 64.51%) gave the EHR they primarily used a low rating.

Table 1. Characteristics of the study sample of Finnish physicians (N=4630).

Characteristic	Participants, n (%)
Age group (years) (n=4591)	
<35	948 (20.65)
35-44	1211 (26.38)
45-54	1148 (25)
55-64	1284 (27.97)
Gender (n=4617)	
Women	2960 (64.11)
Men	1626 (35.22)
Other or did not want to respond	31 (0.67)
Purpose of using the HIS[a] (n=4630)	
Not only for work with patients	962 (20.78)
Only for work with patients	3668 (79.22)
Employment sector (n=4614)	
Public	3676 (79.67)
Private	938 (20.33)
Specialization status (n=4630)	
Not specialized	1496 (32.31)
Specialized	3134 (67.69)
EHR[b]experience (months) (n=4611)	
<6	727 (15.77)
≥6	3884 (84.23)
EHR grade (4-10) (n=4610)	
Low grade	2974 (64.51)
High grade	1636 (35.49)
Participation in HIS development (n=4601)	
No	3495 (75.96)
Yes	1106 (24.04)
Telemedicine with patients (n=4625)	
Little or not at all	2448 (52.93)
Somewhat to very much	2177 (47.07)

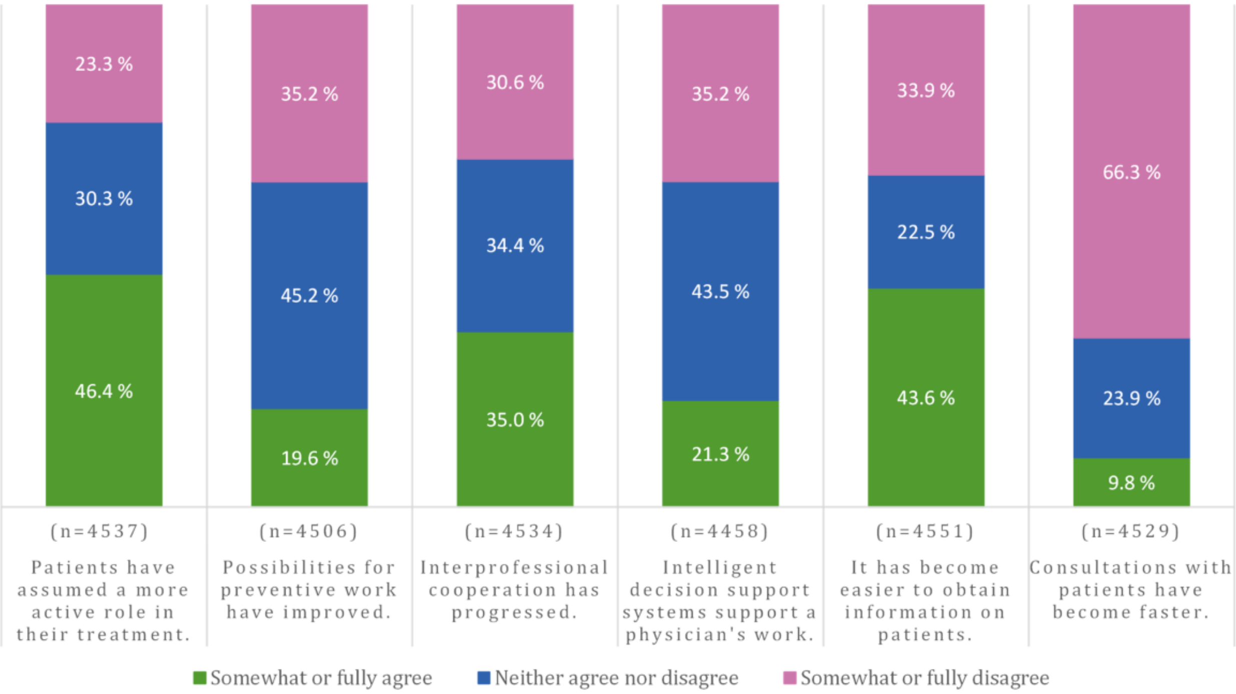
[a] HIS: health information system.

[b] EHR: electronic health record.

Perceived Effects of Digitalization on Work

The descriptive statistics of the perceived effects of digitalization are shown in Figure 2. Nearly half (2104/4537, 46.37%) of the respondents agreed that digitalization of health care has had a positive effect by giving patients a more active role in their treatment (mean 3.24, SD 0.99). This statement was rated as the most positive of all the estimated effects of digitalization on work. In addition, nearly half (1986/4551, 43.64%) of the respondents indicated that obtaining information about patients has become easier (mean 3.06, SD 1.18). The weakest positive effect of digitalization was found in consultation times (mean 2.12, SD 1.03). Only one-tenth (445/4529, 9.82%) found that consultations with patients have become faster. Almost half (2036/4506, 45.18%) of the respondents estimated that digitalization has neither increased nor decreased the possibilities for preventive work (mean 2.78, SD 0.91). Almost half (1941/4458, 43.54%) of the respondents could not state whether the CDSSs have supported their work (mean 2.77, SD 0.99). Assessments of the progress of interprofessional cooperation were evenly distributed (mean 3.01, SD 1.01).

Figure 2. Physicians' perceptions about the effects of digitalization on work. The scale ranged from 1 (fully disagree) to 5 (fully agree). Somewhat or fully disagree included response options 1 to 2, and somewhat or fully agree included response options 4 to 5.



Network Analyses

The resulting network (Figure 3) shows the interconnections between the perceived effects of digitalization on work. In the estimated network, each node was connected to 3, 4, or 5 other nodes. The strongest associations were between the patients' active role and preventive work (b=0.36), patient information and faster consultations (b=0.32), interprofessional cooperation and preventive work (b=0.29), interprofessional cooperation and patient information (b=0.23), and decision support and patient information (b=0.19). The other associations ranged from 0.07 to 0.16, indicating weaker interconnections.

According to the centrality statistics (Figure 4), the most central perceived effect of digitalization in terms of strength (ie, how strongly a variable was connected to all other nodes) was preventive work, followed by interprofessional cooperation. Patients' active role had the lowest cumulative strength of connections to other variables.

Figure 3. A visualized network (n=4339) of the perceived effects of digitalization on the physicians' work. The strength of the dependency is reflected in the weight of the pairwise edge. Positive edges are shown in green.

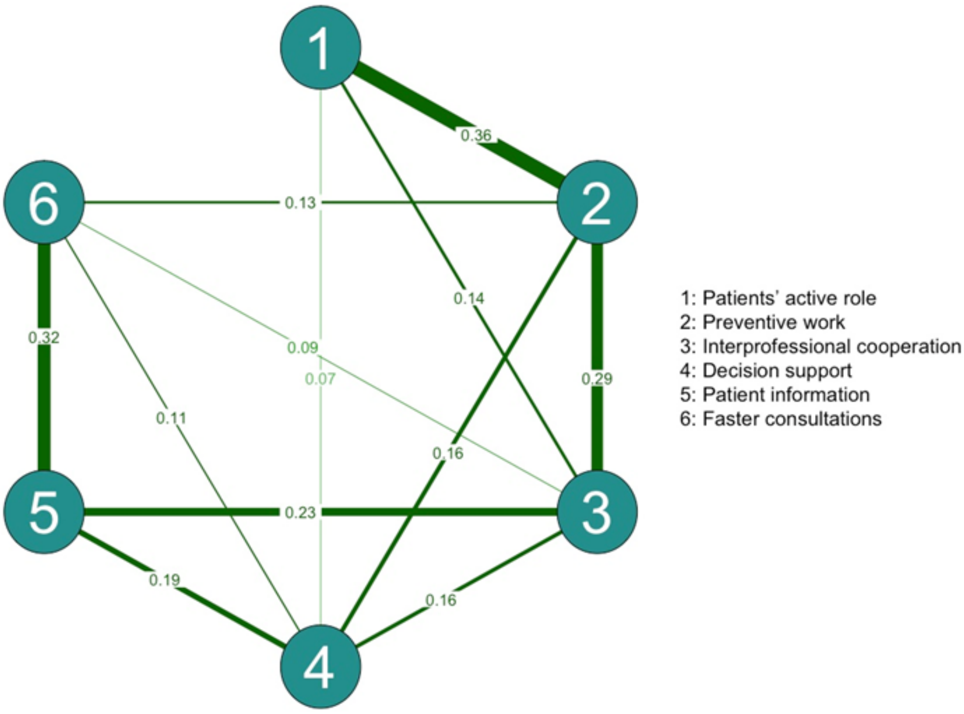
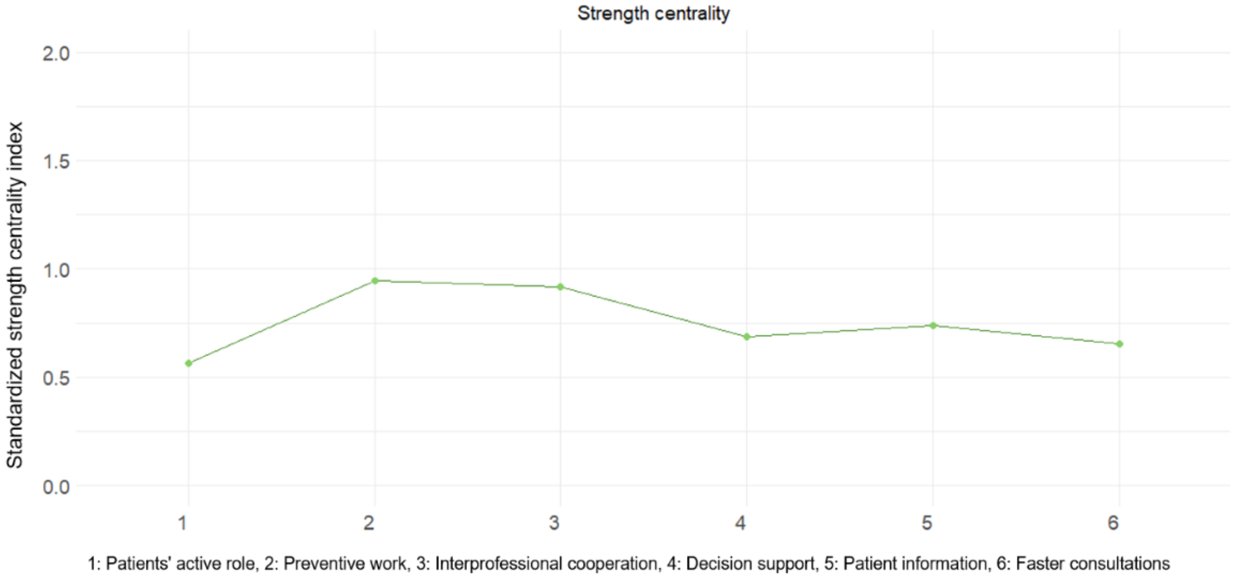


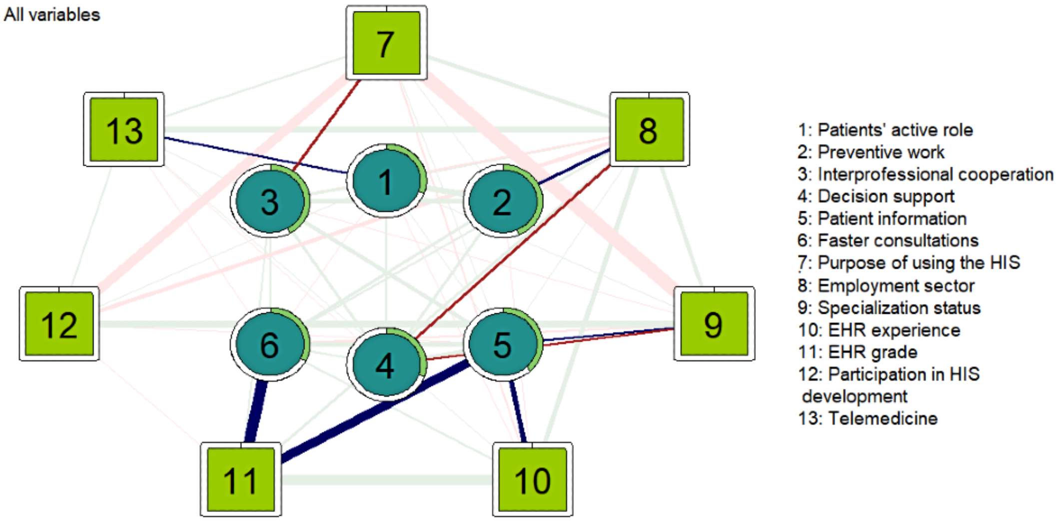
Figure 4. Standardized (ie, *z* scores) centrality indexes denoting node strength for perceived effects of digitalization.



The resulting main network (Figure 5) shows the connections between all 13 factors: perceived effects of digitalization on work (1-6) and factors related to work (7-9) and digital health usage (10-13). In the estimated network, the nodes were connected to 5 to 10 other nodes. The strongest direct links to effects of digitalization were with EHR grade. A higher EHR grade was associated with faster consultations (b=0.32) and facilitated access to patient information (b=0.28). At least 6 months of experience with the main EHR was associated with facilitated access to patient information (b=0.18). Using the HIS only for working with patients was negatively associated with progressed interprofessional cooperation (b=−0.16). The private sector was positively associated with improved possibilities for preventive work (b=0.16) and negatively associated with the supportiveness of CDSSs (b=−0.15). Specialization was negatively associated with the supportiveness of CDSSs (b=−0.13) and positively associated with facilitated access to patient information (b=0.13). A greater amount of telemedicine with patients was associated with a more active role of patients (b=0.12).

Pairwise connections for all visualized variables that reach a value above 0.20 are reported. There was a strong negative association between specialization status and the purpose of using the HIS (b=−0.85). A strong positive association was found between a longer experience with the main EHR and a higher EHR grade (b=0.67). There was also a strong negative association between the purpose of using the HIS and participation in HIS development (b=−0.59). In addition, being specialized was associated with participation in HIS development (b=0.38). The private sector was associated with a greater number of telemedicine services (b=0.34), longer experience with the main EHR (b=0.27), and using the HIS only for patient work (b=0.21). The predictability of each node (perceived effects of digitalization) ranged from 30.2% (decision support) to 43.6% (preventive work) of the explained variance in the continuous variables.

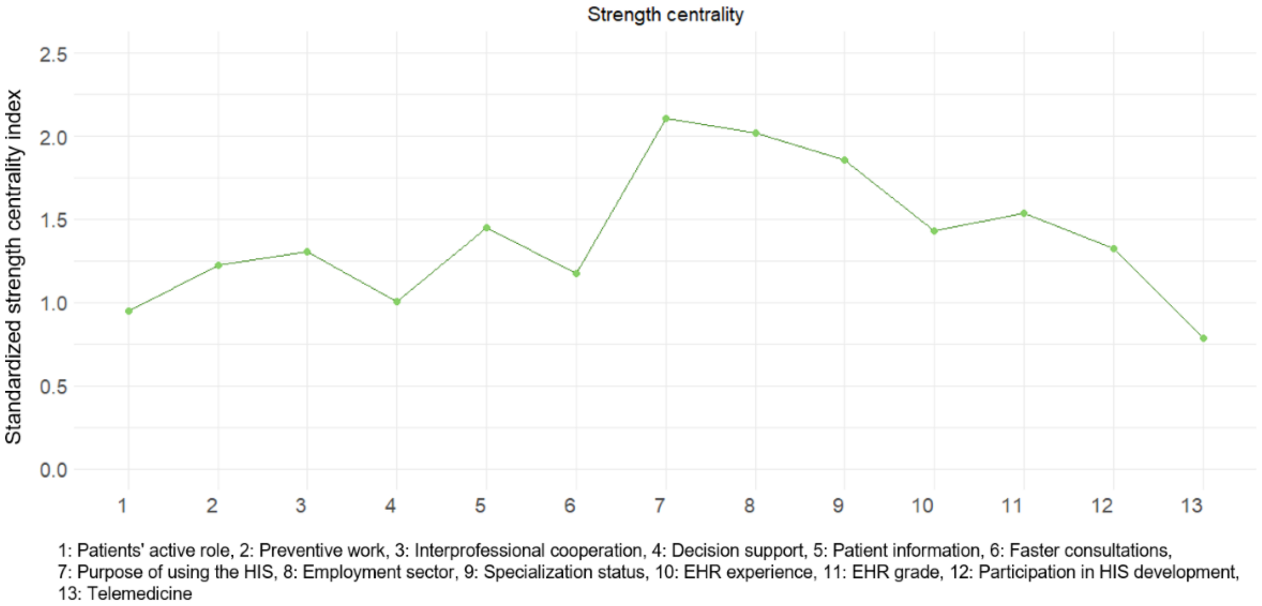
Figure 5. A visualized main network (n=4339) of the perceived effects of digitalization on the work and factors related to work and digital health usage. The strength of the dependency is reflected in the weight of the pairwise edge. Positive edges are shown in green and blue, and negative edges are shown in red. The green ring around each node represents its predictability. EHR: electronic health record; HIS: health information system.



The work-related factors were the most central in the main network (Figure 6). In terms of strength, reflecting the overall influence in the network, purpose of using the HIS had the highest cumulative strength of connections to other variables, followed by employment sector and specialization status. Among the factors related to digital health usage, EHR grade had the highest cumulative strength, while the level of telemedicine had the lowest. Of the perceived effects of digitalization, obtaining information had the highest strength and patients' role had the lowest.

Network stability analysis provided some large and overlapping bootstrapped CIs around the edge weights (Figure S1 in Multimedia Appendix 2). The generally large, bootstrapped CIs urge caution in interpreting the relative sizes of edges. However, correlation stability for strength centrality was 0.517, meaning that 51.7% of cases could be dropped to maintain a correlation with the original centrality greater than 0.7 with a 95% confidence. Values greater than 0.5 were considered stable. Figure S2 in Multimedia Appendix 2 shows the resulting strength-stability plot. Testing for significant differences revealed that all edges differed significantly from several other edges (Figure S3 in Multimedia Appendix 2). In addition, all node strengths differed significantly from those of at least half of the other nodes (Figure S4 in Multimedia Appendix 2).

Figure 6. Standardized (ie, *z* scores) centrality indexes denoting node strength for each factor. EHR: electronic health record; HIS: health information system.



Discussion

Principal Findings

On the basis of our results, nearly half of the physicians assessed that digitalization has promoted an active role for patients in their own care and easier access to patient information. However, only 1 in 10 participants felt that digitalization has had a positive impact on consultation times with patients. The network analysis highlighted the complex interdependence of several factors related to the perceived effects of digitalization on physicians' work. The most central factors in the main network were work related: purpose of using the HIS, employment sector, and specialization status. However, the strongest direct links to the perceived effects of digitalization was with how highly physicians rated the EHR they primarily used. A higher EHR grade was associated with perceptions of faster consultations and easier access to patient information. Overall, purpose of using the HIS, employment sector, specialization status, and EHR grade seemed to be key factors that influenced how positively the effects of digitalization on work were perceived.

The Perceived Effects of Digitalization on Work

Our results showed that nearly half of the respondents agreed that digitalization has helped patients take a more active role in their care. This finding supports the idea that digital health empowers patients to be active participants [9-11]. However, the changing role of patients may have required physicians to more actively encourage their patients to engage in digital health [73]. Patients also need a variety of skills, such as digital health literacy, to play an independent role [74]. Simultaneously, physicians are increasingly expected to assess whether patients have properly understood health information in relation to their own situation [75].

While our results show that the role of patients in their own care was seen as more active and it was associated with improved possibilities for preventive work, nearly half of the physicians estimated that digitalization had neither increased nor decreased opportunities for prevention. Although there is extensive research on the effectiveness of digital health interventions [12-17], not all digital health interventions are created equal and many lack evidence, and achieving outcomes depends on providing the right type of intervention to the right population [12]. From the point of view of the changing work of physicians, digital health prevention work may require a new kind of support for patients, which should be properly and differently targeted.

Almost half of the respondents agreed that digitalization had facilitated access to patient information. Similarly, previous studies have shown that EHRs support the bidirectional flow of patient information and facilitate information sharing among stakeholders [21]. Physicians have perceived obtaining real-time patient data as one of the benefits of EHRs [25]. Patient data sharing between hospitals has also increased [24]. According to our results, easier access to patient information was associated with faster consultation times. However, a significant majority (3003/4529, 66.31%) of the respondents disagreed that consultations with patients have become faster due to digitalization. This was clearly the lowest-rated area and indicates that the time advantage gained in obtaining the information may be lost elsewhere. Previous studies have reported, for example, increased documentation time and time spent on the computer during short-term follow-ups [18,76,77]. The initial transition to the new EHR appears to increase documentation time, but the workflow seems to improve as staff members become more familiar with the system [18]. Most physicians in this study had been using the same system for at least 6 months; thus, recent implementations were unlikely to significantly explain respondents' time use.

Interprofessional cooperation was also connected to improved possibilities for preventive work and for more easily obtaining information on patients. The connection to improved possibilities for preventive work was presumable as interprofessional context had previously been argued to help expand a narrow interpretation of one field and promote the contribution of each member of the team [78]. Health promotion generally involves professionals with different disciplinary backgrounds, who typically also work in different sectors [79]. In this study, over one-third of the respondents estimated that interprofessional cooperation has improved. However, almost as many felt that digitalization has not influenced cooperation. Multiple studies have suggested that the implementation of EHRs [18,19] and electronic prescribing systems [20] can have negative effects on interprofessional communication. However, Chao [21] showed an increased frequency of interprofessional communication while maintaining intraprofessional communication patterns after EHR implementation. Digital health use is also shown to support interdisciplinary cooperation [22], and there are encouraging results for the use of specialists in nonspecialist telemedicine [23]. We cannot draw direct conclusions about the efficiency of collaboration, but our results suggest that digitalization is clearly changing the way professionals interact.

Nearly half of the respondents felt that digitalization had neither increased nor decreased the support physicians received from the CDSSs. However, better support from CDSSs and easier access to patient information were associated. Documenting patient information in a structured, uniform, and simple manner has been previously described as essential for electronic decision support [80]. Previous research suggests that CDSSs can support physicians [28,29,32] and improve outcomes [28,30,31], but they continue to fall short of their full potential [31]. Ford et al [81] pointed out that many previous CDSSs were not developed with the end user, practice context, or clinical workflow in mind. In addition, previous studies show that user attitudes [82] and acceptance are central to the success of CDSSs [83,84].

The Main Network

In the estimated main network, the strongest direct links to the perceived effects of digitalization were with EHR ratings. Higher EHR grades were associated with the perceptions of faster consultations and easier access to patient information. This finding largely supports the work of other studies in this area that link EHR to accessible patient information [21,24,25]. Our network analysis also showed that longer experience with the EHR that was primarily used was associated with perceptions of facilitated access to patient information.

The network also revealed other factors directly associated with the perceived effects of digitalization on work, but these were all relatively weaker. Physicians who used the HIS only to work with patients perceived less progression of interprofessional cooperation than physicians who also used the HIS for other purposes. Physicians with administrative roles may use digital technologies and also collaborate in diverse ways. Leadership in digital health services is thought to require interprofessional and intersectoral collaboration [85]. It is also suggested that leaders in health care may generally view the effects of digitalization differently and more positively than professionals (Kaihlanen, unpublished data, June 2022).

Specialists experienced less support from CDSSs but better access to patient information compared with nonspecialists. Previous research has shown that specialists have also used CDSSs less frequently than general practitioners [86]. This could be related not only to expertise but also to the clinical work itself. General practitioners are known to find electronic medical records more useful than specialists because they are faced with a wider range of symptoms to diagnose, treat, or refer [87].

Physicians who worked in the private sector perceived better opportunities for preventive work but less support from CDSSs than physicians in the public sector. These associations may indicate the different natures of medical work in the private and public sectors. For example, specialized medical care and emergency care are provided in hospitals, and most hospitals in Finland are public sector hospitals [59]. Moreover, employers are responsible for the preventive health care of their employees, most of which is provided by the private sector [60]. The perceived lower level of support from the CDSSs may also be because the CDSSs in Finland are used less in the private sector than in the public sector [61].

Physicians who used more telemedicine with patients felt that digitalization has promoted the active role of patients more than those who barely used telemedicine or did not use it at all. Previous studies have also shown that telemedicine requires patient engagement but also encourages patients to take more responsibility for their own care [88,89]. In addition, physicians who have used telemedicine are more likely known to perceive the potential benefits of telemedicine than physicians who do not use telemedicine with patients [90].

Work-related factors (purpose of using the HIS, employment sector, and specialization status) were found to be the most central factors in the network. Our results suggest that physicians who use HISs primarily for patient work participate less in development work than physicians who also have administrative responsibilities. Participation is known to provide an important sense of control over one's work [27], and physician-initiated improvements to EHR systems have also been found to be useful [91]. The estimated network also showed that specialized physicians were more likely to participate in development work than physicians who were not specialized or were still in training. One possible explanation for this is that those with specialization are in demand for development work because their expertise is in a narrow area of medicine and they receive recognition for their esoteric skills and knowledge [92].

Working in the private sector was associated with a greater amount of telemedicine, longer experience with the main EHR, and the use of HISs only for patient work. An earlier study also found differences among practices, hospitals, and academic medical centers in the use of telemedicine with patients [93], and private hospitals have been more successful than public hospitals in adopting telemedicine [94]. Physicians in the private sector also had more experience with the primary EHR brand used than those in the public sector. Longer experience with the EHR was associated with higher EHR ratings. Shorter experience may indicate new staff members or recent EHR implementation. Several previous studies have shown that physicians are less satisfied after implementing a new EHR [43,44] primarily because of increased workflow disruptions [45]. Significantly more disruptions are noted during the transition period of approximately 6 months, after which the situation recovers [46]. Overall, the results suggest that the employer sector plays an important role in indirectly influencing how physicians view the effects of digitalization through factors related to telemedicine, experience with using EHRs, and the purpose of using HISs.

Limitations

This study has several limitations. First, the psychometric properties of the measure of the perceived effects of digitalization on physicians' work have not been tested previously. However, the statements are based on the Finnish eHealth and eSocial Strategy 2020 and describe the effects of digitalization on the work of physicians from the perspective of the focus areas included in the strategy. Moreover, the items were planned by a large team of experts who have been working on the digitalization of physicians' work for a long time. The content and wording of the measures were pilot tested and evaluated by 2 physicians. On the basis of their feedback, minor revisions were made before the survey was disseminated. Contrary to many previous studies that have focused on the negative ramifications of digitalization on physicians' work [40,95], our questions focused on aspects that can be considered to be positive in nature. Thus, the results may have been different if we would have focused on the negative ramifications of digitalization on the work of physicians.

Second, we used the Likert scale variables as continuous variables in the network, but this has been successfully practiced in other studies of mgm network modeling [eg, 96]. Third, as our data were cross-sectional, the directionality of the observed relationships was not established. Some factors may precede others but some may both contribute to and be influenced by other factors in the network. The network analyses suggest that incorporating factors measured over time, or studies that rely on longitudinal structural model tests, could provide the confirmatory evidence needed [47]. In addition, the response rate was relatively low (24.46%), indicating a higher likelihood of nonresponse bias. However, the sample size was large, and it was well representative of the target population [56]. Finally, Finland is one of the pioneers of the digitalization of health care, and tax-funded universal health care is available to all residents [97]. Therefore, caution should be exercised when generalizing our findings to countries with other types of health care systems or information communication technology infrastructure. However, digitalization is advancing at a rapid pace, and all physicians and health care organizations should be prepared for future changes.

Conclusions

Our results suggest that, from the physicians' perspective, digitalization has improved the active role of patients and the patient information flow, while consultation times with patients have not become faster. Further studies are needed to examine where the potential time benefit of accessible patient information is lost and how the potentially increased documentation time affects the quality of care. Physicians' work should be organized so that the time spent on the computer is not out of the time spent with patients during appointments.

The network highlights that several factors have a complex relationship with the perceived effects of digitalization on physicians' work. The EHR system used appears to be critical for easier access to patient information and faster consultations. Therefore, it seems that a high-quality EHR system is important for the promotion of smooth clinical practice. Thus, organizations could benefit by investing in a well-functioning EHR and the factors that influence its successful use.

In addition, the physician's work (patient or administrative), employer sector, and career stage may play an important background role and influence many other factors that affect the success of digital interventions and their implementation. Thus, it would be advisable for organizations to involve specialists and nonspecialists equally when developing new digital tools or processes in their work. Moreover, some benefits of digitalization seem to be sector specific. Therefore, the specific context and intended use should be considered when developing and implementing digital health. Sectors can also learn from each other; for example, when and how to use telemedicine with patients, how to use digital health in health promotion, and how to benefit from CDSSs in practice. Thus, the information flow and exchange between sectors should be improved.

Overall, a more comprehensive view is needed when assessing the impact of digitalization on specific work environments and work processes. Digitalization of work and related factors should also be considered when orienting physicians. When new digital health is introduced, training on changing work processes should be provided. This would be a priority in addition to technical training. The introduction of digital health is not only a technological project but also a project that changes existing work practices and the professionals' work environment.

Acknowledgments

This study was supported by the Strategic Research Council of the Academy of Finland (project 327145) and the Ministry of Social Affairs and Health (project 414919001). ME was supported by the Academy of Finland (project 339 390). The authors would like to express their gratitude to all physicians who responded to the questionnaire as well as the Finnish Medical Association for collaboration.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author PS upon reasonable request after approval by the study group and the Finnish Medical Association.

Conflicts of Interest

TL is employed by a publicly owned in-house health information system software vendor included in the study; the employer did not provide any support, financial or otherwise, for the study. The employer was not involved in the design of the study or the collection, analysis, or interpretation of the data.

Multimedia Appendix 1
Variables used in this study.
[DOCX File , 22 KB - jmir_v24i8e38714_app1.docx]

Multimedia Appendix 2
Stability analyses.
[DOCX File , 112 KB - jmir_v24i8e38714_app2.docx]

References

1. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
2. Global Strategy on Digital Health 2020-2025. Geneva: World Health Organization; 2021.
3. 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]
4. Meske C, Junglas I. Investigating the elicitation of employees' support towards digital workplace transformation. Behav Inform Technol 2020 Mar 16;40(11):1120-1136. [doi: 10.1080/0144929x.2020.1742382]
5. Vallo Hult H, Hansson A, Svensson L, Gellerstedt M. Flipped healthcare for better or worse. Health Informatics J 2019 Sep;25(3):587-597 [FREE Full text] [doi: 10.1177/1460458219833099] [Medline: 30887867]
6. What is digital health? U.S. Food & Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2022-03-11]
7. Ricciardi W. Assessing the impact of digital transformation of health services: opinion by the Expert Panel on Effective Ways of Investing in Health (EXPH). Eur J Public Health 2019;29(Supplement_4). [doi: 10.1093/eurpub/ckz185.769]
8. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
9. Awad A, Trenfield SJ, Pollard TD, Ong JJ, Elbadawi M, McCoubrey LE, et al. Connected healthcare: improving patient care using digital health technologies. Adv Drug Deliv Rev 2021 Nov;178:113958. [doi: 10.1016/j.addr.2021.113958] [Medline: 34478781]
10. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
11. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. J Family Med Prim Care 2020 May;9(5):2207-2213 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_105_20] [Medline: 32754475]
12. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]

13. Pouls BPH, Vriezekolk JE, Bekker CL, Linn AJ, van Onzenoort HAW, Vervloet M, et al. Effect of interactive eHealth interventions on improving medication adherence in adults with long-term medication: systematic review. J Med Internet Res 2021 Jan 08;23(1):e18901 [FREE Full text] [doi: 10.2196/18901] [Medline: 33416501]
14. Jeminiwa R, Hohmann L, Qian J, Garza K, Hansen R, Fox BI. Impact of eHealth on medication adherence among patients with asthma: a systematic review and meta-analysis. Respir Med 2019 Mar;149:59-68 [FREE Full text] [doi: 10.1016/j.rmed.2019.02.011] [Medline: 30803887]
15. Cavero-Redondo I, Saz-Lara A, Sequí-Dominguez I, Gómez-Guijarro MD, Ruiz-Grao MC, Martinez-Vizcaino V, et al. Comparative effect of eHealth interventions on hypertension management-related outcomes: a network meta-analysis. Int J Nurs Stud 2021 Dec;124:104085 [FREE Full text] [doi: 10.1016/j.ijnurstu.2021.104085] [Medline: 34601205]
16. 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]
17. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
18. Baumann LA, Baker J, Elshaug AG. The impact of electronic health record systems on clinical documentation times: a systematic review. Health Policy 2018 Aug;122(8):827-836. [doi: 10.1016/j.healthpol.2018.05.014] [Medline: 29895467]
19. Michel L. A failure to communicate? Doctors and nurses in American hospitals. J Health Polit Policy Law 2017 Aug;42(4):709-717. [doi: 10.1215/03616878-3856149] [Medline: 28483809]
20. Mohsin-Shaikh S, Furniss D, Blandford A, McLeod M, Ma T, Beykloo MY, et al. The impact of electronic prescribing systems on healthcare professionals' working practices in the hospital setting: a systematic review and narrative synthesis. BMC Health Serv Res 2019 Oct 22;19(1):742 [FREE Full text] [doi: 10.1186/s12913-019-4554-7] [Medline: 31640689]
21. Chao CA. The impact of electronic health records on collaborative work routines: a narrative network analysis. Int J Med Inform 2016 Oct;94:100-111. [doi: 10.1016/j.ijmedinf.2016.06.019] [Medline: 27573317]
22. Janssen A, Brunner M, Keep M, Hines M, Nagarajan SV, Kielly-Carroll C, et al. Interdisciplinary eHealth practice in cancer care: a review of the literature. Int J Environ Res Public Health 2017 Oct 25;14(11):1289 [FREE Full text] [doi: 10.3390/ijerph14111289] [Medline: 29068377]
23. Patil R, Shrivastava R, Juvekar S, McKinstry B, Fairhurst K. Specialist to non-specialist teleconsultations in chronic respiratory disease management: a systematic review. J Glob Health 2021 Mar 27;11:04019 [FREE Full text] [doi: 10.7189/jogh.11.04019] [Medline: 34326988]
24. Pylypchuk Y, Meyerhoefer CD, Encinosa W, Searcy T. The role of electronic health record developers in hospital patient sharing. J Am Med Inform Assoc 2022 Jan 29;29(3):435-442 [FREE Full text] [doi: 10.1093/jamia/ocab263] [Medline: 34871412]
25. Rathert C, Porter TH, Mittler JN, Fleig-Palmer M. Seven years after Meaningful Use: physicians' and nurses' experiences with electronic health records. Health Care Manage Rev 2019;44(1):30-40. [doi: 10.1097/HMR.0000000000000168] [Medline: 28614166]
26. Huang ME. It is from Mars and physicians from Venus: bridging the gap. PM R 2017 May 18;9(5S):S19-S25. [doi: 10.1016/j.pmrj.2017.02.011] [Medline: 28527499]
27. Vainiomäki S, Aalto AM, Lääveri T, Sinervo T, Elovainio M, Mäntyselkä P, et al. Better usability and technical stability could lead to better work-related well-being among physicians. Appl Clin Inform 2017 Dec 14;08(04):1057-1067. [doi: 10.4338/aci-2017-06-ra-0094]
28. Jia P, Zhang L, Chen J, Zhao P, Zhang M. The effects of clinical decision support systems on medication safety: an overview. PLoS One 2016;11(12):e0167683 [FREE Full text] [doi: 10.1371/journal.pone.0167683] [Medline: 27977697]
29. Tao L, Zhang C, Zeng L, Zhu S, Li N, Li W, et al. Accuracy and effects of clinical decision support systems integrated with BMJ best practice-aided diagnosis: interrupted time series study. JMIR Med Inform 2020 Jan 20;8(1):e16912 [FREE Full text] [doi: 10.2196/16912] [Medline: 31958069]
30. Berner ES, editor. Clinical Decision Support Systems: Theory and Practice. New York: Springer International Publishing AG; 2016.
31. Jaspers MW, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011 May 01;18(3):327-334 [FREE Full text] [doi: 10.1136/amiajnl-2011-000094] [Medline: 21422100]
32. Laing S, Mercer J. Improved preventive care clinical decision-making efficiency: leveraging a point-of-care clinical decision support system. BMC Med Inform Decis Mak 2021 Nov 11;21(1):315 [FREE Full text] [doi: 10.1186/s12911-021-01675-8] [Medline: 34763691]
33. Burmann A, Tischler M, Faßbach M, Schneitler S, Meister S. The role of physicians in digitalizing health care provision: web-based survey study. JMIR Med Inform 2021 Nov 11;9(11):e31527 [FREE Full text] [doi: 10.2196/31527] [Medline: 34545813]
34. Vuori V, Helander N, Okkonen J. Digitalization in knowledge work: the dream of enhanced performance. Cogn Tech Work 2018 Jun 27;21(2):237-252. [doi: 10.1007/s10111-018-0501-3]

35. Zaresani A, Scott A. Does digital health technology improve physicians' job satisfaction and work-life balance? A cross-sectional national survey and regression analysis using an instrumental variable. BMJ Open 2020 Dec 12;10(12):e041690 [FREE Full text] [doi: 10.1136/bmjopen-2020-041690] [Medline: 33310807]
36. Lluch M. Healthcare professionals' organisational barriers to health information technologies-a literature review. Int J Med Inform 2011 Dec;80(12):849-862. [doi: 10.1016/j.ijmedinf.2011.09.005] [Medline: 22000677]
37. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]
38. Müller MB, Knop M, Ressing C, Freude H, Oschinsky FM, Klein HC, et al. Constituting factors of a digitally influenced relationship between patients and primary care physicians in rural areas. ScholarSpace. 2020. [doi: 10.24251/HICSS.2020.447]
39. Fiske A, Buyx A, Prainsack B. The double-edged sword of digital self-care: physician perspectives from Northern Germany. Soc Sci Med 2020 Sep;260:113174 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113174] [Medline: 32659512]
40. Heponiemi T, Kujala S, Vainiomäki S, Vehko T, Lääveri T, Vänskä J, et al. Usability factors associated with physicians' distress and information system-related stress: cross-sectional survey. JMIR Med Inform 2019 Nov 05;7(4):e13466 [FREE Full text] [doi: 10.2196/13466] [Medline: 31687938]
41. Heponiemi T, Hyppönen H, Vehko T, Kujala S, Aalto AM, Vänskä J, et al. Finnish physicians’ stress related to information systems keeps increasing: a longitudinal three-wave survey study. BMC Med Inform Decis Mak 2017 Oct 17;17(1):147-147. [doi: 10.1186/s12911-017-0545-y]
42. Nadav J, Kaihlanen A, Kujala S, Laukka E, Hilama P, Koivisto J, et al. How to implement digital services in a way that they integrate into routine work: qualitative interview study among health and social care professionals. J Med Internet Res 2021 Dec 01;23(12):e31668 [FREE Full text] [doi: 10.2196/31668] [Medline: 34855610]
43. Meyerhoefer CD, Sherer SA, Deily ME, Chou SY, Guo X, Chen J, et al. Provider and patient satisfaction with the integration of ambulatory and hospital EHR systems. J Am Med Inform Assoc 2018 Aug 01;25(8):1054-1063 [FREE Full text] [doi: 10.1093/jamia/ocy048] [Medline: 29788287]
44. Vishwanath A, Singh SR, Winkelstein P. The impact of electronic medical record systems on outpatient workflows: a longitudinal evaluation of its workflow effects. Int J Med Inform 2010 Nov;79(11):778-791. [doi: 10.1016/j.ijmedinf.2010.09.006] [Medline: 20947415]
45. Nguyen L, Bellucci E, Nguyen LT. Electronic health records implementation: an evaluation of information system impact and contingency factors. Int J Med Inform 2014 Nov;83(11):779-796. [doi: 10.1016/j.ijmedinf.2014.06.011] [Medline: 25085286]
46. Jacobs M, Boersma LJ, Swart R, Mannens R, Reymen B, Körver F, et al. Electronic Health Record implementation in a large academic radiotherapy department: temporarily disruptions but long-term benefits. Int J Med Inform 2019 Sep;129:342-348. [doi: 10.1016/j.ijmedinf.2019.07.008] [Medline: 31445276]
47. Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. Nat Rev Methods Primers 2021 Aug 19;1(1):58. [doi: 10.1038/s43586-021-00055-w]
48. Barabási AL. The network takeover. Nature Phys 2011 Dec 22;8(1):14-16. [doi: 10.1038/nphys2188]
49. Jordan DG, Winer ES, Salem T. The current status of temporal network analysis for clinical science: considerations as the paradigm shifts? J Clin Psychol 2020 Sep 09;76(9):1591-1612 [FREE Full text] [doi: 10.1002/jclp.22957] [Medline: 32386334]
50. Newman M, Barabási AL, Watts D. The Structure and Dynamics of Networks. Princeton, New Jersey, United States: Princeton University Press; 2006.
51. Newman M. Networks: An Introduction. Oxford: University Press; 2010.
52. Information management in social welfare and health care. Finnish Institute for Health and Welfare. URL: https://thl.fi/en/web/information-management-in-social-welfare-and-health-care [accessed 2022-01-27]
53. Electronic health records as professionals' tools. Finnish Institute for Health and Welfare. URL: https://sampo.thl.fi/pivot/prod/en/steps2/laakari/summary_tiiviste6 [accessed 2022-01-27]
54. Viitanen J, Hyppönen H, Lääveri T, Vänskä J, Reponen J, Winblad I. National questionnaire study on clinical ICT systems proofs: physicians suffer from poor usability. Int J Med Inform 2011 Oct;80(10):708-725. [doi: 10.1016/j.ijmedinf.2011.06.010] [Medline: 21784701]
55. Physicians 2019. Finnish Medical Association. URL: https://www.laakariliitto.fi/site/assets/files/5256/sll_taskutilasto_en_220620.pdf [accessed 2022-01-08]
56. Potilastietojärjestelmät lääkärin työvälineenä 2021. Finnish Medical Association. URL: https://www.laakariliitto.fi/site/assets/files/5229/x_tiedotemateriaalit_polte_2021_final.pdf [accessed 2022-04-11]
57. The ethical principles of research with human participants and ethical review in the human sciences in Finland. Finnish National Board on Research. Integrity TENK guidelines 2019. Finnish National Board on Research Integrity TENK 3/2019. URL: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf [accessed 2022-03-31]
58. Finland in the Digital Economy and Society Index. European Commission. URL: https://digital-strategy.ec.europa.eu/en/policies/desi-finland [accessed 2022-01-11]

59. Social welfare and health care system in Finland, responsibilities. Ministry of Social Affairs and Health. URL: https://stm.fi/en/social-and-health-services/responsible-agencies [accessed 2021-12-13]
60. Private social and health services. Ministry of Social Affairs and Health. URL: https://stm.fi/en/private-health-care [accessed 2021-12-13]
61. Reponen J, Keränen N, Ruotanen R, Tuovinen T, Haverinen J, Kangas M. Tieto- ja viestintäteknologian käyttö terveydenhuollossa vuonna 2020: Tilanne ja kehityksen suunta. [Use of information and communications technology in Finnish health care in 2020. Current situation and trends.]. Helsinki: Finnish Institute for Health and Welfare. Report 11/2021; 2021. URL: https://urn.fi/URN:ISBN:978-952-343-771-5
62. Legislation. Kanta. URL: https://www.kanta.fi/en/professionals/legislation [accessed 2021-12-13]
63. Kenkimäki H, Keränen N, Haverinen J, Reponen J. EHR-connected specialty specific auxiliary systems in public specialized healthcare 2014–2020. FinJeHeW 2021;13(3). [doi: 10.23996/fjhw.107667]
64. Jansson M, Liisanantti J, Ala-Kokko T, Reponen J. The negative impact of interface design, customizability, inefficiency, malfunctions, and information retrieval on user experience: a national usability survey of ICU clinical information systems in Finland. Int J Med Inform 2022 Mar;159:104680 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104680] [Medline: 34990942]
65. Ruotanen R, Kangas M, Tuovinen T, Keränen N, Haverinen J, Reponen J. Finnish e-health services intended for citizens – national and regional development. FinJeHeW 2021 Oct 25;13(3):283-301. [doi: 10.23996/fjhw.109778]
66. Information to Support Well-Being and Service Renewal. eHealth and eSocial Strategy 2020. Helsinki: Finnish Government; 2020. URL: http://urn.fi/URN:ISBN:978-952-00-3575-4
67. Yang E, Baker Y, Ravikumar P, Allen G, Liu Z. Mixed graphical models via exponential families. In: Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS). 2014 Presented at: International Conference on Artificial Intelligence and Statistics (AISTATS); Apr 22- 25, 2014; Reykjavic, Iceland URL: https://proceedings.mlr.press/v33/yang14a.html
68. Liu H, Lafferty J, Wasserman L. The nonparanormal: semiparametric estimation of high dimensional undirected graphs. J Mach Learn Res 2009;10(80):2328-2328.
69. Haslbeck JMB, Waldorp LJ. mgm: estimating time-varying mixed graphical models in high-dimensional data. J Stat Softw 2020 Dec;93(8):1-46 [FREE Full text] [doi: 10.18637/jss.v093.i08]
70. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: network visualizations of relationships in psychometric data. J Stat Soft 2012;48(4):1-18. [doi: 10.18637/jss.v048.i04]
71. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. Behav Res Methods 2018 Feb 24;50(1):195-212 [FREE Full text] [doi: 10.3758/s13428-017-0862-1] [Medline: 28342071]
72. Haslbeck JMB, Waldorp LJ. How well do network models predict observations? On the importance of predictability in network models. Behav Res Methods 2018 Apr 17;50(2):853-861 [FREE Full text] [doi: 10.3758/s13428-017-0910-x] [Medline: 28718088]
73. Shimoga SV, Lu YZ. Role of provider encouragement on patient engagement via online portals. J Am Med Inform Assoc 2019 Oct 01;26(10):968-976 [FREE Full text] [doi: 10.1093/jamia/ocz026] [Medline: 30925585]
74. Xu RH, Zhou LM, Wong EL, Wang D. The association between patients' eHealth literacy and satisfaction with shared decision-making and well-being: multicenter cross-sectional study. J Med Internet Res 2021 Sep 24;23(9):e26721 [FREE Full text] [doi: 10.2196/26721] [Medline: 34559062]
75. Huisman M, Joye S, Biltereyst D. Searching for health: doctor Google and the shifting dynamics of the middle-aged and older adult patient-physician relationship and interaction. J Aging Health 2020 Oct;32(9):998-1007. [doi: 10.1177/0898264319873809] [Medline: 31517558]
76. Beckmann M, Dittmer K, Jaschke J, Karbach U, Köberlein-Neu J, Nocon M, et al. Electronic patient record and its effects on social aspects of interprofessional collaboration and clinical workflows in hospitals (eCoCo): a mixed methods study protocol. BMC Health Serv Res 2021 Apr 23;21(1):377 [FREE Full text] [doi: 10.1186/s12913-021-06377-5] [Medline: 33892703]
77. Pelland KD, Baier RR, Gardner RL. "It's like texting at the dinner table": a qualitative analysis of the impact of electronic health records on patient-physician interaction in hospitals. J Innov Health Inform 2017 Jun 30;24(2):894 [FREE Full text] [doi: 10.14236/jhi.v24i2.894] [Medline: 28749316]
78. Dean E, Moffat M, Skinner M, Dornelas de Andrade A, Myezwa H, Söderlund A. Toward core inter-professional health promotion competencies to address the non-communicable diseases and their risk factors through knowledge translation: curriculum content assessment. BMC Public Health 2014 Jul 14;14:717 [FREE Full text] [doi: 10.1186/1471-2458-14-717] [Medline: 25016946]
79. Ledderer L, Christensen M, Burau V. Interprofessional collaboration in health promotion: changing tasks as leverage for innovation. Eur J Public Health 2019;29(Supplement_4). [doi: 10.1093/eurpub/ckz185.317]
80. Koskela T, Sandström S, Mäkinen J, Liira H. User perspectives on an electronic decision-support tool performing comprehensive medication reviews - a focus group study with physicians and nurses. BMC Med Inform Decis Mak 2016 Jan 22;16(1):6 [FREE Full text] [doi: 10.1186/s12911-016-0245-z] [Medline: 26801630]

81. Ford E, Edelman N, Somers L, Shrewsbury D, Lopez Levy M, van Marwijk H, et al. Barriers and facilitators to the adoption of electronic clinical decision support systems: a qualitative interview study with UK general practitioners. BMC Med Inform Decis Mak 2021 Jun 21;21(1):193 [FREE Full text] [doi: 10.1186/s12911-021-01557-z] [Medline: 34154580]
82. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017 Sep 15;12(1):113 [FREE Full text] [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
83. Ji M, Genchev GZ, Huang H, Xu T, Lu H, Yu G. Evaluation framework for successful artificial intelligence-enabled clinical decision support systems: mixed methods study. J Med Internet Res 2021 Jun 02;23(6):e25929 [FREE Full text] [doi: 10.2196/25929] [Medline: 34076581]
84. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
85. Laukka E, Pölkki T, Heponiemi T, Kaihlanen AM, Kanste O. Leadership in digital health services: protocol for a concept analysis. JMIR Res Protoc 2021 Mar 04;10(2):e25495 [FREE Full text] [doi: 10.2196/25495] [Medline: 33538702]
86. Razmak J, Bélanger CH, Farhan W. Managing patients’ data with clinical decision support systems: a factual assessment. J Decision Syst 2018 Oct 14;27(3):123-145. [doi: 10.1080/12460125.2018.1533159]
87. Butler RJ, Johnson WG. Rating the digital help: electronic medical records, software providers, and physicians. Int J Health Econ Manag 2016 Sep 9;16(3):269-283. [doi: 10.1007/s10754-016-9190-8] [Medline: 27878677]
88. Pariser P, Pham TT, Brown JB, Stewart M, Charles J. Connecting people with multimorbidity to interprofessional teams using telemedicine. Ann Fam Med 2019 Aug 12;17(Suppl 1):S57-S62 [FREE Full text] [doi: 10.1370/afm.2379] [Medline: 31405877]
89. Alvandi M. Telemedicine and its role in revolutionizing healthcare delivery. Am J Accountable Care 2017;5(1).
90. Barton PL, Brega AG, Devore PA, Mueller K, Paulich MJ, Floersch NR, et al. Specialist physicians' knowledge and beliefs about telemedicine: a comparison of users and nonusers of the technology. Telemed J E Health 2007 Oct;13(5):487-499. [doi: 10.1089/tmj.2006.0091] [Medline: 17999611]
91. Guo U, Chen L, Mehta PH. Electronic health record innovations: helping physicians - One less click at a time. Health Inf Manag 2017 Sep;46(3):140-144. [doi: 10.1177/1833358316689481] [Medline: 28671038]
92. Mishra AN, Anderson C, Angst CM, Agarwal R. Electronic health records assimilation and physician identity evolution: an identity theory perspective. Inform Syst Res 2012 Sep;23(3-part-1):738-760. [doi: 10.1287/isre.1110.0407]
93. Pierce BS, Perrin PB, Dow AW, Dautovich ND, Rybarczyk BD, Mishra VK. Changes in physician telemedicine use during COVID-19: effects of practice setting, demographics, training, and organizational policies. Int J Environ Res Public Health 2021 Sep 22;18(19):9963 [FREE Full text] [doi: 10.3390/ijerph18199963] [Medline: 34639269]
94. Sood SP, Negash S, Mbarika VW, Kifle M, Prakash N. Differences in public and private sector adoption of telemedicine: Indian case study for sectoral adoption. Stud Health Technol Inform 2007;130:257-268. [Medline: 17917199]
95. Heponiemi T, Hyppönen H, Kujala S, Aalto A, Vehko T, Vänskä J, et al. Predictors of physicians' stress related to information systems: a nine-year follow-up survey study. BMC Health Serv Res 2018 Apr 13;18(1):284 [FREE Full text] [doi: 10.1186/s12913-018-3094-x] [Medline: 29653530]
96. Liu D, Epskamp S, Isvoranu AM, Chen C, Liu W, Hong X. Network analysis of physical and psychiatric symptoms of hospital discharged patients infected with COVID-19. J Affect Disord 2021 Nov 01;294:707-713 [FREE Full text] [doi: 10.1016/j.jad.2021.07.043] [Medline: 34343929]
97. Keskimäki I, Tynkkynen LK, Reissell E, Koivusalo M, Syrjä V, Vuorenkoski L, et al. Finland: health system review. World Health Organization. Regional Office for Europe. URL: https://apps.who.int/iris/handle/10665/327538 [accessed 2022-01-09]

Abbreviations

CDSS: clinical decision support system
EHR: electronic health record
HIS: health information system
RQ: research question

Edited by T Leung; submitted 14.04.22; peer-reviewed by G Yu, J Baker; comments to author 10.06.22; revised version received 17.06.22; accepted 24.06.22; published 17.08.22.

Please cite as:
Saukkonen P, Elovainio M, Virtanen L, Kaihlanen AM, Nadav J, Lääveri T, Vänskä J, Viitanen J, Reponen J, Heponiemi T
The Interplay of Work, Digital Health Usage, and the Perceived Effects of Digitalization on Physicians' Work: Network Analysis Approach
J Med Internet Res 2022;24(8):e38714
URL: https://www.jmir.org/2022/8/e38714
doi:10.2196/38714
PMID:35976692

Original Paper

Development and Evaluation of a Natural Language Processing Annotation Tool to Facilitate Phenotyping of Cognitive Status in Electronic Health Records: Diagnostic Study

Ayush Noori[1*]; Colin Magdamo[1*], BSc; Xiao Liu[1], MSc; Tanish Tyagi[1]; Zhaozhi Li[1], MSc; Akhil Kondepudi[1]; Haitham Alabsi[1,2], MD; Emily Rudmann[1,3], BSc; Douglas Wilcox[1,2], MD, PhD; Laura Brenner[2,4], MD; Gregory K Robbins[2,5], MD; Lidia Moura[1,2], MD, PhD; Sahar Zafar[1,2], MD; Nicole M Benson[2,6,7], MD, MBI; John Hsu[2,6], MD, PhD; John R Dickson[1,2], MD, PhD; Alberto Serrano-Pozo[1,2], MD, PhD; Bradley T Hyman[1,2], MD, PhD; Deborah Blacker[2,8], MD, PhD; M Brandon Westover[1,2], MD, PhD; Shibani S Mukerji[1,2,5], MD, PhD; Sudeshna Das[1,2], PhD

[1]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States
[2]Harvard Medical School, Boston, MA, United States
[3]Vaccine and Immunotherapy Center, Division of Infectious Disease, Boston, MA, United States
[4]Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, Boston, MA, United States
[5]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, United States
[6]Mongan Institute, Massachusetts General Hospital, Boston, MA, United States
[7]McLean Hospital, Belmont, MA, United States
[8]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Sudeshna Das, PhD
Department of Neurology
Massachusetts General Hospital
55 Fruit Street
Boston, MA, 02114
United States
Phone: 1 617 726 2000
Email: SDAS5@mgh.harvard.edu

Abstract

Background: Electronic health records (EHRs) with large sample sizes and rich information offer great potential for dementia research, but current methods of phenotyping cognitive status are not scalable.

Objective: The aim of this study was to evaluate whether natural language processing (NLP)–powered semiautomated annotation can improve the speed and interrater reliability of chart reviews for phenotyping cognitive status.

Methods: In this diagnostic study, we developed and evaluated a semiautomated NLP-powered annotation tool (NAT) to facilitate phenotyping of cognitive status. Clinical experts adjudicated the cognitive status of 627 patients at Mass General Brigham (MGB) health care, using NAT or traditional chart reviews. Patient charts contained EHR data from two data sets: (1) records from January 1, 2017, to December 31, 2018, for 100 Medicare beneficiaries from the MGB Accountable Care Organization and (2) records from 2 years prior to COVID-19 diagnosis to the date of COVID-19 diagnosis for 527 MGB patients. All EHR data from the relevant period were extracted; diagnosis codes, medications, and laboratory test values were processed and summarized; clinical notes were processed through an NLP pipeline; and a web tool was developed to present an integrated view of all data. Cognitive status was rated as cognitively normal, cognitively impaired, or undetermined. Assessment time and interrater agreement of NAT compared to manual chart reviews for cognitive status phenotyping was evaluated.

Results: NAT adjudication provided higher interrater agreement (Cohen κ=0.89 vs κ=0.80) and significant speed up (time difference mean 1.4, SD 1.3 minutes; P<.001; ratio median 2.2, min-max 0.4-20) over manual chart reviews. There was moderate agreement with manual chart reviews (Cohen κ=0.67). In the cases that exhibited disagreement with manual chart reviews, NAT adjudication was able to produce assessments that had broader clinical consensus due to its integrated view of highlighted relevant information and semiautomated NLP features.

Conclusions: NAT adjudication improves the speed and interrater reliability for phenotyping cognitive status compared to manual chart reviews. This study underscores the potential of an NLP-based clinically adjudicated method to build large-scale dementia research cohorts from EHRs.

(J Med Internet Res 2022;24(8):e40384) doi:10.2196/40384

KEYWORDS

chart review; cognition; cognitive status; dementia; diagnostic; electronic health record; health care; natural language processing; research cohort

Introduction

In recent years, electronic health records (EHRs) have become increasingly common in US health care facilities; they provide a wealth of information on patient demographics, medical history, clinical data, and health system interactions. EHRs offer an unprecedented opportunity to improve clinical care and examine a broad variety of scientific, health care utilization, and heath policy questions [1-3]. An important first step in conducting EHR research is accurately identifying patients with a certain health condition, event, or disease, which is known as phenotyping [1,4]. The identified patient sample is subsequently leveraged for a wide range of purposes, such as providing clinical decision support for health care delivery [5], conducting epidemiological research [4,6], and for the practice of precision medicine [7].

Phenotyping cognitive status (ie, distinguishing between normal cognition and any stage of cognitive impairment) in EHR is a major challenge since dementia is underrecognized, underdiagnosed, and underreported in claims data [8-12], leading to inaccurate identification of dementia cases in many studies based on claims or EHR data [13-15]. Informative missingness, errors, and biases in EHR may further exacerbate the challenges of defining dementia outcomes [16]. Yet another challenge of phenotyping arises from complex, subjective, loosely-defined diagnostic criteria as well as the format—that is, structured (eg, diagnosis codes and medications) versus unstructured (eg, clinical notes and images)—in which the information is stored [4]. Previous studies have demonstrated that information on cognitive status is often found only in free text [17-19]. Clinicians may chart symptoms of cognitive problems in clinical notes but may not make a formal diagnosis, refer to a specialist, or prescribe medication for multiple reasons including clinical role, lack of time or expertise, patient resistance, or limited treatment options [20-22]. Thus, accurately phenotyping cognitive status requires the combined use of both structured data, such as diagnosis codes, medications, and laboratory test results, as well as unstructured clinical notes.

Several algorithms have been developed for phenotyping cognitive status; some studies used structured data, such as diagnosis codes, missed appointments, or health care utilization patterns [15,23], whereas others have applied natural language processing (NLP) to unstructured notes [18,19,24]. None of these prior efforts combined both structured and unstructured input modalities, and manual annotation by clinical experts is limited by the lack of available tools to facilitate efficient chart review [25]. Thus, we hypothesized that the best approach for phenotyping cognitive status is a semiautomated one in which automated NLP is applied to clinical notes and presented in an integrated view to the clinical expert for final manual adjudication of cognitive status.

We developed NAT, a semiautomated NLP-powered annotation tool, to facilitate adjudication of cognitive status. The tool extracts and processes data from EHRs and then ranks clinical notes based on a deep learning NLP algorithm (Macro F_1=0.92) that classifies whether a note indicates normal cognition, cognitive impairment, or has no pertinent information [26]. It highlights key information and presents a summarized view to the annotator. We evaluated NAT in two EHR data sets: (1) Medicare beneficiaries from the Mass General Brigham (MGB) Accountable Care Organization (ACO) who were labeled in another study using manual chart reviews [15] and (2) MGB patients with laboratory confirmed SARS-CoV-2 (a case-control study to investigate the effects of COVID-19 on people with and without HIV was used as an exemplar of a research cohort that requires labeling of cognitive status). We evaluated interrater agreement in the first data set and compared it to interrater agreement in Epic—the EHR system used at MGB since 2015. The second data set was used to compare timings of manual to NAT adjudication, as the timing of manual adjudication was not available in the first data set.

By addressing the gaps in current chart review methods and leveraging existing NLP methods, we demonstrate that NAT increases both the efficiency and the interrater reliability of phenotyping cognitive status in EHR (relative to manual chart reviews) to build future research cohorts.

Methods

Clinical Settings and Data Sources

This diagnostic study was conducted at MGB—formerly Partners Healthcare—a private nonprofit organization comprising two major academic hospitals, community hospitals, and community health centers in the Boston metropolitan area. Data were sourced from the MGB Enterprise Data Warehouse that stores data from Epic. We evaluated NAT adjudication for phenotyping cognitive status on two distinct data sets. The first one included EHR data from January 1, 2017, to December 31, 2018, of 100 patients randomly selected from a larger data set that was expert-annotated via manual Epic chart reviews and reported elsewhere [15]. Specifically, this manually expert-annotated data set contained 1002 Medicare beneficiaries from the MGB ACO who were classified into (1) normal cognition, (2) borderline of normal cognition and mild cognitive impairment (MCI), (3) MCI, (4) borderline of MCI and dementia, or (5) dementia [15]. The experts graded their confidence in the adjudication as low, medium, moderate, or high. The 100 patients were randomly sampled from these 5 classes with 20 from each class, ensuring that each class had a similar distribution of confidence scores. The second data set included 527 MGB patients with a laboratory confirmed SARS-CoV-2 infection based on polymerase chain reaction testing between March 1 and December 31, 2020. The data set was created for a case-control study to investigate the effects of COVID-19 on people with and without HIV; EHR data up to 2 years prior to and any time after the index positive polymerase chain reaction test were used to investigate the performance of NAT adjudication.

Ethics Approval

This study was approved by the MGB Institutional Review Board (2015P001915).

Definition of Cognitive Impairment

In this study, to phenotype cognitive status, patients were annotated with three labels: (1) cognitively normal (CN), (2) cognitively impaired (CI), and (3) undetermined. Patients were labeled as CI if there was any documented suspicion or concern of memory or cognitive decline, whether based on symptoms, observations, or objective testing. This ranged from any dementia-related International Classification of Diseases (ICD) codes or medicines in the patients' charts to cognitive concerns—relayed by patients, family members or friends, or providers in the notes and phone logs—as these concerns often reflect an underlying change in cognition even if a cognitive evaluation is normal (in which case they prompt a diagnosis of subjective cognitive decline [27]). Conversely, to be annotated as CN, at least implicit evidence of no cognitive concerns was required (eg, the patient continued to work, clearly managed their own care or hobbies, and followed complicated instructions, or they had annual wellness or specialist notes with multisystem assessment and no mention of a cognitive concern). The strongest evidence for a CN annotation was a cognition test performed with an explicit note of intact cognition. If there was conflicting evidence of both cognitive impairment and evidence of no cognitive impairment in a patient's chart, the latest evidence or specialist notes (if any were available) informed the adjudication. Finally, patients were marked as "undetermined" if the EHR did not have sufficient information.

Data Preparation

Data query, preparation, and preprocessing steps are described in Multimedia Appendix 1. For each patient, the following EHR data from the relevant time period were extracted from the Enterprise Data Warehouse: (1) patient demographic information, including name, medical record number, birth date, sex, ethnic group, marital status, and educational level; (2) all clinical notes, including reason for visit, history, note text, encounter type, and MGB provider (including provider department, specialty, and qualifications); (3) current primary care provider; (4) patient care coordination note; (5) medication history and current medications; (6) magnetic resonance imaging and computerized tomography orders; (7) laboratory orders and results; (8) problem list, including ICD diagnoses and diagnosis codes; and (9) visit cancellations.

Several features were engineered from the EHR to facilitate assessment of cognitive status. Dementia-related medications and ICD codes (medications: galantamine, donepezil, rivastigmine, and memantine; ICD-9 codes: 290.X, 294.X, 331.X, and 780.93; ICD-10 codes: G30.X and G31.X) and laboratory tests (eg, vitamin B12, folate, and thyroid-stimulating hormone) related to assessment of cognitive status were identified and highlighted. The numbers of cancellations, no-shows, and refill requests, relative to the total number of encounters, were computed.

Finally, NLP was applied to the clinical notes. We curated two lists of regular expressions or keywords related to the presence or absence of both (1) cognitive impairment and (2) the functional impairment of activities of daily living (ADLs) or independent ADLs, respectively (Multimedia Appendices 2 and 3). We identified regular expression matches and highlighted these within the text of the notes with different colors for each category (eg, cognition vs ADLs) to facilitate their identification by the clinician. We applied a previously developed NLP model [26] to generate classification probabilities of the following classes for each note: CI, no CI, or neither. The notes were ranked based on these classification probabilities, and notes that the model predicted as indicative of CI were displayed at the top.

Development of an Annotation Tool

We designed and developed a web-based chart review and annotation tool, using the Python-based open-source Django web development framework with a SQLite database. We established data models for patient-level demographic and clinical data, encounter-level clinician notes, user account creation and authentication, and patient assignment to individual or multiple annotators (Multimedia Appendix 4). We created several user interfaces (ie, pages) to present the various data modalities in an integrated fashion for annotation.

Statistical Analysis

We evaluated NAT adjudication using three metrics: agreement with manual Epic chart reviews, assessment time, and interrater agreement. We evaluated agreement between manual Epic chart reviews and NAT adjudication as well as interrater agreement for NAT adjudication using Cohen κ, whereas assessment time in minutes was compared using a paired samples Wilcoxon test (also known as the Wilcoxon signed-rank test). There were no missing data for these variables. All analyses were conducted using the R statistical software (version 4.1.2; R Core Team).

Results

Patient Characteristics

The patient characteristics of the two data sets are shown in Table 1. The ACO data set comprised 100 patients (63/100, 63.0% were women; mean age 78.8, SD 7.4 years; 7/100, 7% racial or ethnic minorities, 1 missing; 51/100, 51.0% with a college degree or more, 3 missing; and 50/100, 50.0% were married). The COVID-19 data set comprised 527 patients (226/527, 42.9% women; mean age 52.6, SD 15.0 years; 318/527, 60.35% racial or ethnic minorities, 21 missing;

160/527, 30.4% college education or more, 62 missing; and 195/527, 37.0% married, 16 missing).

Table 1. Characteristics of Accountable Care Organization (ACO) and COVID-19 data sets used for NLP[a] annotation tool (NAT) evaluation.

Characteristics	Patients (N=627)	
	ACO data set (n=100)	COVID-19 data set (n=527)
Sex, n (%)		
Male	37 (37)	301 (57.1)
Female	63 (63)	226 (42.9)
Age (years), mean (SD)	78.8 (7.4)	52.6 (15)
Minorities, n (%)		
Black	4 (4)	163 (30.9)
Hispanic	2 (2)	138 (26.2)
Asian	1 (1)	16 (3)
Indigenous	0 (0)	1 (0.2)
College education, n (%)	51 (51)	160 (30.4)
Married, n (%)	50 (50.0)	195 (37)
Clinical characteristics		
Number of encounters, median (min-max)	164 (8-858)	106 (1-2474)
PCP[b] visit, n (%)	71 (71)	423 (80.3)
Dementia ICD[c] code and medication, n (%)	51 (51)	166 (5.3)

[a] NLP: natural language processing.

[b] PCP: primary care provider.

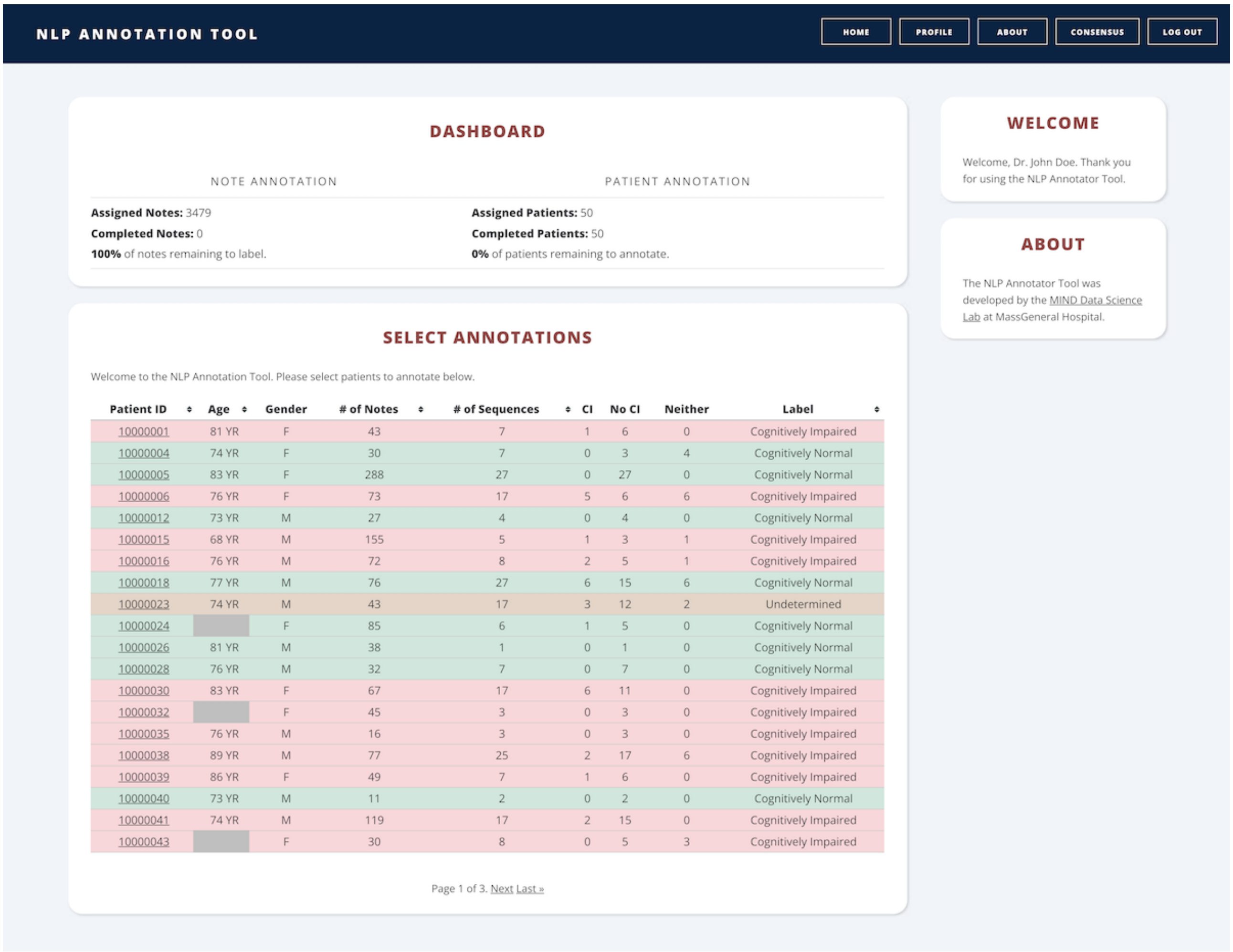
[c] ICD: International Classification of Diseases.

Features of NAT

Upon logging in to our annotation tool, an authenticated user is presented with a dashboard listing the patient IDs, ages, and sexes of their assigned patients (Figure 1). In addition, the total number of notes, the sequences within the notes that match a cognition or ADL keyword (Multimedia Appendices 1 and 2), and the number of notes for each predicted class (ie, cognition and ADL) are also presented. After annotation, the patient's label (CN, CI, or undetermined) is displayed with background colors reflecting the patient's annotated cognitive status.

Selecting a patient navigates the user to an annotation view summarizing the patient's demographic and clinical information (Figure 2A). Engineered features, including the total number of notes, encounters, no shows, cancellations, and refill requests, along with the patient care coordination note (if any), diagnosis ICD codes, and medications, are displayed (Figure 2B). Brain imaging and relevant laboratory tests, such as thyroid-stimulating hormone or vitamin B12, allow annotators to consider systemic causes of cognitive changes (Figure 2C). Finally, notes sorted by the predicted probability and with highlighted keywords are presented to expedite the review of the entire chart history during the relevant period for the clinical adjudication of cognitive status. Examples of the three predicted classes of notes (CN, CI, and undetermined) are shown in Figure 2D.

Figure 1. NAT dashboard: screenshot of the NAT dashboard displaying the current workload and assigned patients. A summary of patient information is displayed in each row, and the background reflects the cognitive status assigned to the patient. NAT: NLP annotation tool; NLP: natural language processing.



DASHBOARD

NOTE ANNOTATION

Assigned Notes: 3479
Completed Notes: 0
100% of notes remaining to label.

PATIENT ANNOTATION

Assigned Patients: 50
Completed Patients: 50
0% of patients remaining to annotate.

WELCOME

Welcome, Dr. John Doe. Thank you for using the NLP Annotator Tool.

ABOUT

The NLP Annotator Tool was developed by the MIND Data Science Lab at MassGeneral Hospital.

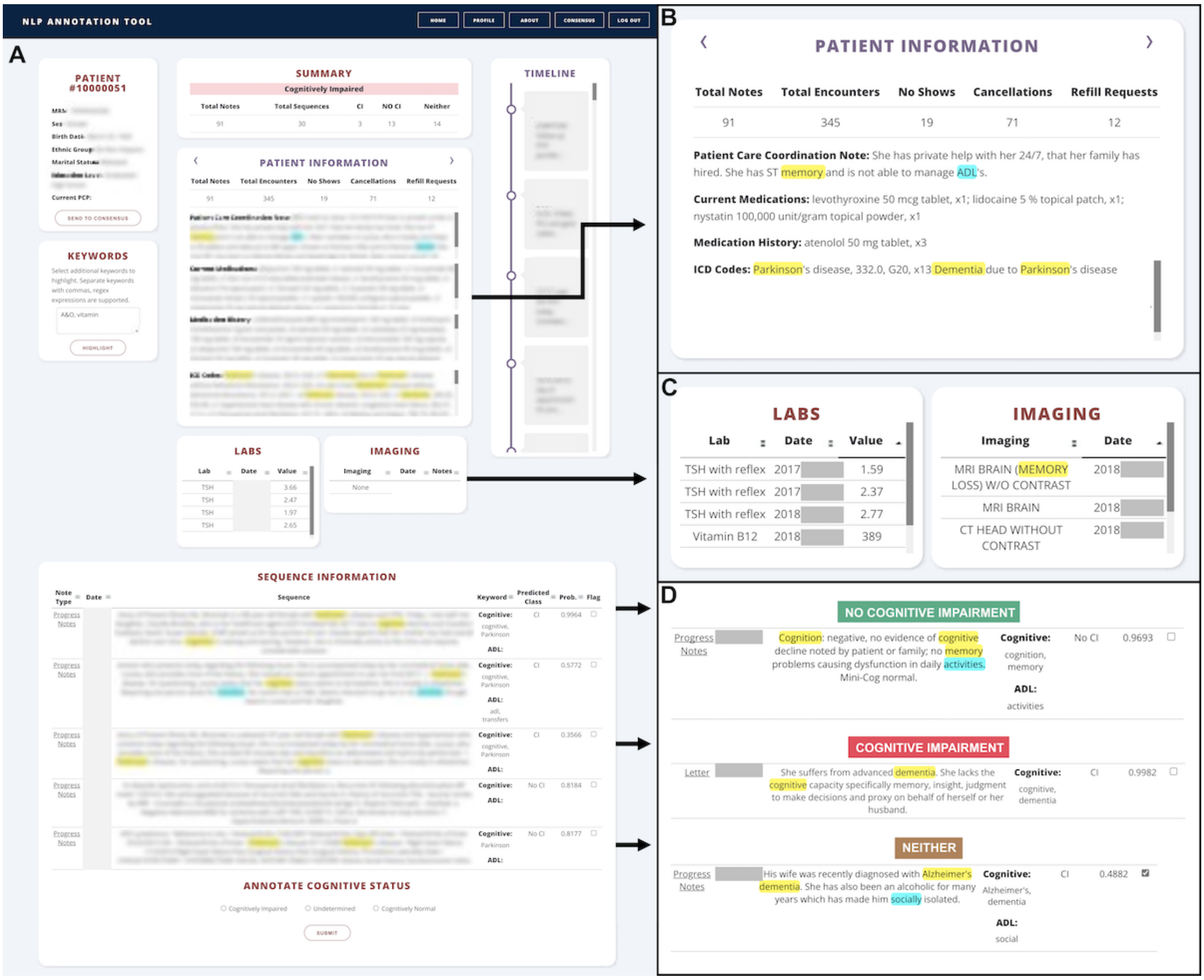
SELECT ANNOTATIONS

Welcome to the NLP Annotation Tool. Please select patients to annotate below.

Patient ID	Age	Gender	# of Notes	# of Sequences	CI	No CI	Neither	Label
10000001	81 YR	F	43	7	1	6	0	Cognitively Impaired
10000004	74 YR	F	30	7	0	3	4	Cognitively Normal
10000005	83 YR	F	288	27	0	27	0	Cognitively Normal
10000006	76 YR	F	73	17	5	6	6	Cognitively Impaired
10000012	73 YR	M	27	4	0	4	0	Cognitively Normal
10000015	68 YR	M	155	5	1	3	1	Cognitively Impaired
10000016	76 YR	M	72	8	2	5	1	Cognitively Impaired
10000018	77 YR	M	76	27	6	15	6	Cognitively Normal
10000023	74 YR	M	43	17	3	12	2	Undetermined
10000024	[redacted]	F	85	6	1	5	0	Cognitively Normal
10000026	81 YR	M	38	1	0	1	0	Cognitively Normal
10000028	76 YR	M	32	7	0	7	0	Cognitively Normal
10000030	83 YR	F	67	17	6	11	0	Cognitively Impaired
10000032	[redacted]	F	45	3	0	3	0	Cognitively Impaired
10000035	76 YR	M	16	3	0	3	0	Cognitively Impaired
10000038	89 YR	M	77	25	2	17	6	Cognitively Impaired
10000039	86 YR	F	49	7	1	6	0	Cognitively Impaired
10000040	73 YR	M	11	2	0	2	0	Cognitively Normal
10000041	74 YR	M	119	17	2	15	0	Cognitively Impaired
10000043	[redacted]	F	30	8	0	5	3	Cognitively Impaired

Page 1 of 3. Next Last »

Figure 2. Annotation view: (A) patient view displaying summary information at the top and sequences from clinical notes at the bottom; (B) the Patient Information box summarizes health care interaction, patient care coordination notes, current medications, and diagnosis codes; (C) laboratory tests and imaging conducted on the patient; (D) sample sequences from notes with dementia and activities of daily living (ADLs) keywords highlighted. Each sequence is classified as cognitive impairment (CI), no CI, or neither, with a probability, and allows annotators to flag incorrect classifications.



PATIENT INFORMATION

Total Notes	Total Encounters	No Shows	Cancellations	Refill Requests
91	345	19	71	12

Patient Care Coordination Note: She has private help with her 24/7, that her family has hired. She has ST memory and is not able to manage ADL's.

Current Medications: levothyroxine 50 mcg tablet, x1; lidocaine 5 % topical patch, x1; nystatin 100,000 unit/gram topical powder, x1

Medication History: atenolol 50 mg tablet, x3

ICD Codes: Parkinson's disease, 332.0, G20, x13 Dementia due to Parkinson's disease

LABS

Lab	Date	Value
TSH with reflex	2017 [redacted]	1.59
TSH with reflex	2017 [redacted]	2.37
TSH with reflex	2018 [redacted]	2.77
Vitamin B12	2018 [redacted]	389

IMAGING

Imaging	Date
MRI BRAIN (MEMORY LOSS) W/O CONTRAST	2018 [redacted]
MRI BRAIN	2018 [redacted]
CT HEAD WITHOUT CONTRAST	2018 [redacted]

NO COGNITIVE IMPAIRMENT

Progress Notes	[redacted]	Cognition: negative, no evidence of cognitive decline noted by patient or family; no memory problems causing dysfunction in daily activities. Mini-Cog normal.	**Cognitive:** cognition, memory **ADL:** activities	No CI	0.9693	☐

COGNITIVE IMPAIRMENT

Letter	[redacted]	She suffers from advanced dementia. She lacks the cognitive capacity specifically memory, insight, judgment to make decisions and proxy on behalf of herself or her husband.	**Cognitive:** cognitive, dementia	CI	0.9982	☐

NEITHER

Progress Notes	[redacted]	His wife was recently diagnosed with Alzheimer's dementia. She has also been an alcoholic for many years which has made him socially isolated.	**Cognitive:** Alzheimer's, dementia **ADL:** social	CI	0.4882	☑

Evaluation of NAT

Two teams of expert clinicians were randomly assigned patients and adjudicated the ACO data set, using NAT (team 1: LB, GKR, SSM; team 2: MBW and HA). We compared the phenotyping of cognitive status using NAT to manual chart reviews using Epic (labels were obtained from Moura et al [15]; patients who were not CN were grouped into the CI class). We removed patients annotated as “undetermined” in the set adjudicated using NAT, as they had little information in EHR to assess cognitive status and could not be directly compared to the labels obtained from Moura et al [15]. The agreement between NAT and manual Epic chart reviews was moderate for both team 1 (Cohen κ=0.68) and team 2 (Cohen κ=0.65) with a mean Cohen κ=0.67; the breakdown is shown in Figure 3A. Surprisingly, patients whose NAT label disagreed with the manual Epic chart reviews were annotated as CI using Epic and as CN using NAT. We manually reviewed the patients where the diagnostic labels disagreed; we found that NAT was able to highlight certain passages of text, such as “language, attention, and memory function are intact with good fund of knowledge”; the highlighted text facilitated the labeling of the patient as CN, whereas such phrases were easily missed in manual chart reviews. Moreover, if a patient had a transient cognitive deficit and was later evaluated as CN, for example, NAT presented all notes with highlighted evidence along with their dates in one view, making it easier to follow the sequence of events. The disagreements were mostly among patients annotated with a low confidence score in the Epic manually annotated data set [15] (Figure 3B). The interrater agreement of NAT adjudication between team 1 and team 2 was higher (Cohen κ=0.89) than the interrater agreement (Cohen κ=0.80) with manual Epic chart reviews reported in Moura et al [15].

Next, we compared the time required for phenotyping of cognitive status via NAT adjudication versus manual chart reviews in Epic. Four of the authors (DW, ER, HA, and SSM) adjudicated the full COVID-19 data set using NAT and recorded the annotation time for 129 patients. Two of the authors (HA and SSM) timed manual chart reviews in Epic for 32 randomly sampled patients. To ensure that a patient was not adjudicated using both methods by the same person, HA used Epic to perform chart reviews of patients adjudicated by SSM using NAT and vice versa. For most of the patients, the annotation time was substantially shorter with NAT as compared to manual chart reviews in Epic (Figure 3C). Adjudications using NAT provided substantial speed-up of annotations compared to manual chart reviews in Epic (time difference mean 1.4, SD 1.3 minutes; *P*<.001; ratio median 2.2, min-max 0.4-20). Additionally, we observed that clinicians spent more time using NAT on the first half of patients compared to the second half. This “learning effect” was not observed with manual Epic chart reviews. The breakdown of the cognitive status for the COVID-19 data set is shown in Figure 4. Notably, the cognitive status for 21.1% (n=111) of patients was undetermined, suggesting that there was little information in EHR to determine their cognitive status.

Figure 3. Comparison of adjudication with natural language processing (NLP)–powered annotation tool (NAT) and manual Epic chart reviews: (A) contingency table displaying adjudication with NAT versus Epic by team 1 (top row) and team 2 (bottom row); (B) distribution of confidence scores assigned in Epic manual chart reviews (Moura et al [15]) for agreements and disagreements between the two methods; (C) annotation time comparisons between NAT versus Epic.

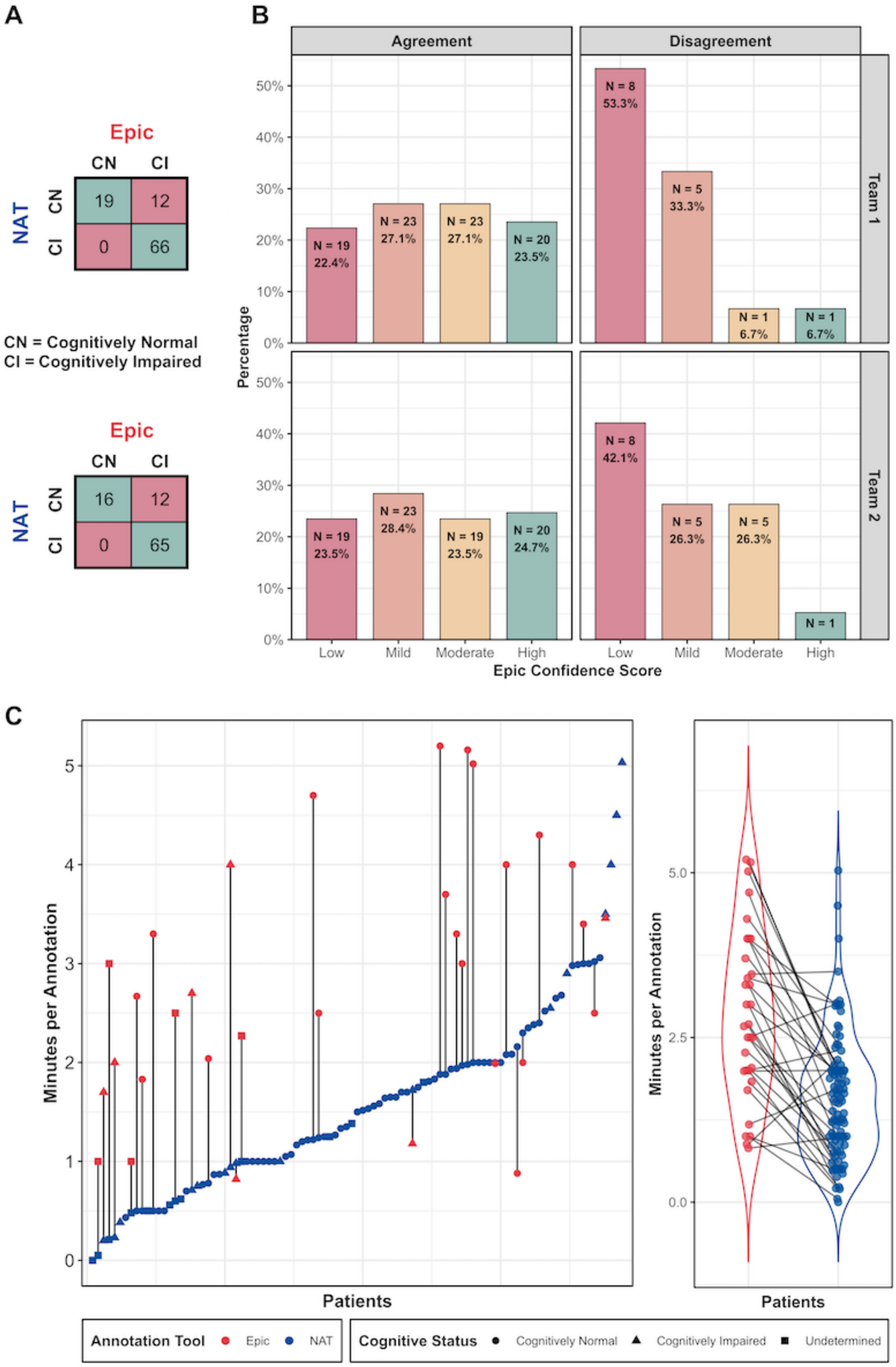
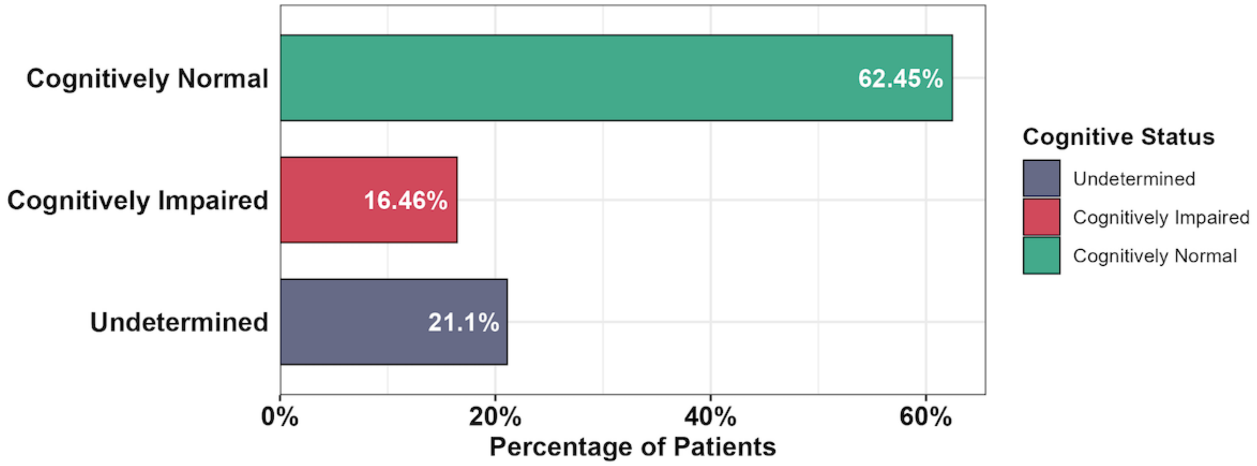


Figure 4. COVID-19 data set cognitive scores and distribution of cognitive scores in the COVID-19 data set.



Discussion

Principal Findings

In this study, we developed and evaluated a novel semiautomated NLP-powered annotation tool, NAT, to facilitate phenotyping of cognitive status. Clinical experts adjudicated the cognitive status of 627 patients at MGB health care using NAT or traditional chart reviews. NAT improves the efficiency and interrater reliability of chart review as compared to manual adjudication.

Strengths

Phenotyping methods have been applied to EHR to successfully identify patients with autism [28], diabetes [29], immunological diseases [30], and several chronic diseases [16]. EHR has been extensively used for dementia research, but the outcomes are typically defined by diagnosis codes or specialist diagnoses. Although phenotyping tools using NLP have been developed to detect cognitive impairment [18,19,24], they have been limited by their performance. In this study, we propose a novel semiautomated approach that combines NLP outputs with manual adjudication.

We selected this approach as it combines the automation of an NLP tool and the expert review required for phenotyping cognitive status. Phenotyping cognitive status requires the input from both structured (eg, diagnosis codes and medications) and unstructured (eg, clinical notes and images) data, and currently, there are no machine learning tools that integrate multiple data modalities. The approach has several advantages over manual chart reviews. Cognitive concerns are often subjective, and a significant amount of information is required to confidently ascertain the correct diagnosis. Since diagnoses are staged across months or years, individual notes across time must be evaluated together—NAT filters data for the period of interest and thus facilitates the adjudication process. Next, the absence of cognitive deficits is often difficult to adjudicate with confidence. In these cases, the annotator needs to review all notes to ensure there were no signs of cognitive impairment. NAT improves the efficiency of such tasks, as it automatically flags notes with signs of cognitive impairment as well as those with information on normal cognition and ranks them in order of importance. In addition, clinicians often use a wide variety of terms and phrases in clinical notes that can easily be missed in manual reviews. NAT, on the other hand, highlights all cognition-related patterns and phrases, decreasing the likelihood that the annotator might miss any information relevant to the decision-making task. Finally, NAT streamlines an established adjudication protocol and thus improves interrater agreement. NAT can, in principle, be extended to local hospitals and clinics that have digitized data but not an EHR system.

Limitations

This study has several limitations. First, NAT does not link to brain images, which may contain information relevant to brain function. Second, although NAT improves the efficiency of adjudicating cognitive status compared to manual chart reviews, it is not scalable to large data sets of thousands of patients. To scale to such sample sizes, fully automated machine learning algorithms that replicate the adjudication process are required. In the future, we plan to use NAT to create gold-standard data sets for training and validation of such machine learning algorithms for phenotyping cognitive status. Third, NAT adjudication was evaluated on data from a single health care system. Whether the cognition and ADL-related keywords apply to other health care settings is yet to be confirmed. The performance of the NLP tool [26] also needs to be evaluated with external data. Fourth, adjudicators were not blinded to identifiable information in EHR, which may have introduced biases in their labels. Tools, such as Philter, could be used in the future to remove protected health information in NAT [31]. Finally, research studies using EHR-based data sets are limited by the information available within the health care system and miss records of care outside the system. Such patients with missing information were labeled as "undetermined" in this study, but studies that use diagnosis codes for phenotyping of cognitive status may incorrectly label such patients as CN instead of distinguishing them as patients with insufficient information. Our study highlights the issue of missing information when phenotyping cognitive status in EHR, and consequently, the need for future work to minimize biases if such patients are excluded in a research study.

Conclusions

Although there is no substitute for a longitudinal cohort with formal cognitive evaluations to study Alzheimer disease and related dementias, leveraging EHR data with NLP holds promise. In this diagnostic study, we developed and evaluated a semiautomated NLP-powered annotation tool, NAT, to facilitate the phenotyping of cognitive status in EHRs. Expert clinicians adjudicated cognitive status of 627 patients from two distinct data sets; NAT had a high interrater agreement and improved the speed of annotations compared to manual chart reviews. Using NAT to adjudicate cognitive status would likely increase the feasibility and scalability of building gold-standard data sets for machine learning algorithms and research cohorts to study cognitive decline.

Acknowledgments

This study was supported by funding from National Institute on Aging awards (K08AG053380, R01AG073410, and P30AG062421), a National Institute of Mental Health award (K23MH115812), the James S McDonnell Foundation, and the Rappaport Fellowship. The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

ASP, BTH, DB, MBW, SSM, and SD conceptualized and designed the study. Acquisition, analysis, or interpretation of data was conducted by AN, CM, XL, TT, AK, HA, ER, DW, LB, GKR, LM, SZ, NMB, JH, and JD. The manuscript was drafted by AN, CM, and SD. All authors contributed to critical revision of the manuscript. Statistical analysis was performed by AN, CM, and XL. Funding was obtained by BTH, SSM, and SD. The study was supervised by SSM and SD.

Conflicts of Interest

SD, BTH, and ASP report research funding from Abbvie Inc. JD serves on a scientific review board for I-Mab Biopharma. NMB volunteers for the Epic Behavioral Health Subspecialty Steering Board.

Multimedia Appendix 1
Data query, preparation, and preprocessing steps.
[PDF File (Adobe PDF File), 253 KB - jmir_v24i8e40384_app1.pdf]

Multimedia Appendix 2
Regular expressions of dementia-related keywords.
[XLSX File (Microsoft Excel File), 7 KB - jmir_v24i8e40384_app2.xlsx]

Multimedia Appendix 3
Regular expressions of activities of daily living (ADLs) keywords.
[XLSX File (Microsoft Excel File), 6 KB - jmir_v24i8e40384_app3.xlsx]

Multimedia Appendix 4
Data model.
[PDF File (Adobe PDF File), 650 KB - jmir_v24i8e40384_app4.pdf]

References

1. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc 2020;2020:191-200 [FREE Full text] [Medline: 32477638]
2. Sendak M, Gao M, Nichols M, Lin A, Balu S. Machine learning in health care: a critical appraisal of challenges and opportunities. EGEMS (Wash DC) 2019 Jan 24;7(1):1 [FREE Full text] [doi: 10.5334/egems.287] [Medline: 30705919]
3. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 Jun;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
4. Wong J, Horwitz MM, Zhou L, Toh S. Using machine learning to identify health outcomes from electronic health record data. Curr Epidemiol Rep 2018 Dec;5(4):331-342 [FREE Full text] [doi: 10.1007/s40471-018-0165-9] [Medline: 30555773]
5. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
6. Zhang Y, Cai T, Yu S, Cho K, Hong C, Sun J, et al. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP). Nat Protoc 2019 Dec;14(12):3426-3444 [FREE Full text] [doi: 10.1038/s41596-019-0227-6] [Medline: 31748751]

7. Wilkinson J, Arnold KF, Murray EJ, van Smeden M, Carr K, Sippy R, et al. Time to reality check the promises of machine learning-powered precision medicine. Lancet Digit Health 2020 Dec;2(12):e677-e680 [FREE Full text] [doi: 10.1016/S2589-7500(20)30200-4] [Medline: 33328030]
8. Alzheimer’s Association. 2021 Alzheimer's disease facts and figures. Alzheimers Dement 2021 Mar;17(3):327-406. [doi: 10.1002/alz.12328] [Medline: 33756057]
9. Amjad H, Roth DL, Sheehan OC, Lyketsos CG, Wolff JL, Samus QM. Underdiagnosis of dementia: an observational study of patterns in diagnosis and awareness in US older adults. J Gen Intern Med 2018 Jul;33(7):1131-1138 [FREE Full text] [doi: 10.1007/s11606-018-4377-y] [Medline: 29508259]
10. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. Alzheimer Dis Assoc Disord 2009;23(4):306-314 [FREE Full text] [doi: 10.1097/WAD.0b013e3181a6bebc] [Medline: 19568149]
11. Taylor, Jr. DH, Østbye T, Langa KM, Weir D, Plassman BL. The accuracy of Medicare claims as an epidemiological tool: the case of dementia revisited. JAD 2009 Jul 23;17(4):807-815. [doi: 10.3233/jad-2009-1099]
12. Kotagal V, Langa KM, Plassman BL, Fisher GG, Giordani BJ, Wallace RB, et al. Factors associated with cognitive evaluations in the United States. Neurology 2014 Nov 26;84(1):64-71. [doi: 10.1212/wnl.0000000000001096]
13. Ostbye T, Taylor DH, Clipp EC, Scoyoc LV, Plassman BL. Identification of dementia: agreement among national survey data, medicare claims, and death certificates. Health Serv Res 2008 Mar;43(1 Pt 1):313-326 [FREE Full text] [doi: 10.1111/j.1475-6773.2007.00748.x] [Medline: 18211532]
14. Chen Y, Tysinger B, Crimmins E, Zissimopoulos JM. Analysis of dementia in the US population using Medicare claims: insights from linked survey and administrative claims data. Alzheimers Dement (N Y) 2019 Jun 06;5(1):197-207 [FREE Full text] [doi: 10.1016/j.trci.2019.04.003] [Medline: 31198838]
15. Moura LMVR, Festa N, Price M, Volya M, Benson NM, Zafar S, et al. Identifying Medicare beneficiaries with dementia. J Am Geriatr Soc 2021 Aug 26;69(8):2240-2251 [FREE Full text] [doi: 10.1111/jgs.17183] [Medline: 33901296]
16. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
17. Wei W, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J Am Med Inform Assoc 2016 Apr;23(e1):e20-e27 [FREE Full text] [doi: 10.1093/jamia/ocv130] [Medline: 26338219]
18. Gilmore-Bykovskyi AL, Block LM, Walljasper L, Hill N, Gleason C, Shah MN. Unstructured clinical documentation reflecting cognitive and behavioral dysfunction: toward an EHR-based phenotype for cognitive impairment. J Am Med Inform Assoc 2018 Sep 01;25(9):1206-1212 [FREE Full text] [doi: 10.1093/jamia/ocy070] [Medline: 29947805]
19. Reuben DB, Hackbarth AS, Wenger NS, Tan ZS, Jennings LA. An automated approach to identifying patients with dementia using electronic medical records. J Am Geriatr Soc 2017 Mar;65(3):658-659. [doi: 10.1111/jgs.14744] [Medline: 28152164]
20. Yarnall KSH, Pollak KI, Østbye T, Krause KM, Michener JL. Primary care: is there enough time for prevention? Am J Public Health 2003 Apr;93(4):635-641. [doi: 10.2105/ajph.93.4.635] [Medline: 12660210]
21. Boustani M, Perkins AJ, Fox C, Unverzagt F, Austrom MG, Fultz B, et al. Who refuses the diagnostic assessment for dementia in primary care? Int J Geriat Psychiatry 2006 Jun;21(6):556-563. [doi: 10.1002/gps.1524]
22. Fowler NR, Frame A, Perkins AJ, Gao S, Watson DP, Monahan P, et al. Traits of patients who screen positive for dementia and refuse diagnostic assessment. Alzheimers Dement (Amst) 2015 Jun;1(2):236-241 [FREE Full text] [doi: 10.1016/j.dadm.2015.01.002] [Medline: 26258162]
23. Barnes DE, Zhou J, Walker RL, Larson EB, Lee SJ, Boscardin WJ, et al. development and validation of eRADAR: a tool using EHR data to detect unrecognized dementia. J Am Geriatr Soc 2020 Jan;68(1):103-111. [doi: 10.1111/jgs.16182] [Medline: 31612463]
24. Amra S, O'Horo JC, Singh TD, Wilson GA, Kashyap R, Petersen R, et al. Derivation and validation of the automated search algorithms to identify cognitive impairment and dementia in electronic health records. J Crit Care 2017 Feb;37:202-205. [doi: 10.1016/j.jcrc.2016.09.026] [Medline: 27969571]
25. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Sep 21;7(9):e1000326 [FREE Full text] [doi: 10.1371/journal.pmed.1000326] [Medline: 20877712]
26. Tyagi T, Magdamo C, Noori A, Li Z, Liu X, Deodhar M, et al. Using deep learning to identify patients with cognitive impairment in electronic health records. ArXiv Preprint posted online on Nov 13, 2021 arXiv:2111.09115 [cs.CL]. [doi: 10.48550/arXiv.2111.09115]
27. van Harten AC, Mielke MM, Swenson-Dravis DM, Hagen CE, Edwards KK, Roberts RO, et al. Subjective cognitive decline and risk of MCI. Neurology 2018 Jun 29;91(4):e300-e312. [doi: 10.1212/wnl.0000000000005863]
28. Leroy G, Gu Y, Pettygrove S, Galindo MK, Arora A, Kurzius-Spencer M. Automated extraction of diagnostic criteria from electronic health records for autism spectrum disorders: development, evaluation, and application. J Med Internet Res 2018 Nov 07;20(11):e10497 [FREE Full text] [doi: 10.2196/10497] [Medline: 30404767]

29. Zheng T, Xie W, Xu L, He X, Zhang Y, You M, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform 2017 Dec;97:120-127 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.09.014] [Medline: 27919371]
30. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. J Allergy Clin Immunol 2020 Feb;145(2):463-469 [FREE Full text] [doi: 10.1016/j.jaci.2019.12.897] [Medline: 31883846]
31. Norgeot B, Muenzen K, Peterson TA, Fan X, Glicksberg BS, Schenk G, et al. Protected Health Information filter (Philter): accurately and securely de-identifying free-text clinical notes. NPJ Digit Med 2020 Apr 14;3(1):57 [FREE Full text] [doi: 10.1038/s41746-020-0258-y] [Medline: 32337372]

Abbreviations

ACO: Accountable Care Organization
ADL: activities of daily living
CI: cognitively impaired
CN: cognitively normal
EHR: electronic health record
ICD: International Classification of Diseases
MCI: mild cognitive impairment
MGB: Mass General Brigham
NAT: NLP annotation tool
NLP: natural language processing

Edited by G Eysenbach, C Basch; submitted 17.06.22; peer-reviewed by J Walsh, L Hopper; comments to author 13.07.22; revised version received 29.07.22; accepted 31.07.22; published 30.08.22.

Please cite as:
Noori A, Magdamo C, Liu X, Tyagi T, Li Z, Kondepudi A, Alabsi H, Rudmann E, Wilcox D, Brenner L, Robbins GK, Moura L, Zafar S, Benson NM, Hsu J, R Dickson J, Serrano-Pozo A, Hyman BT, Blacker D, Westover MB, Mukerji SS, Das S
Development and Evaluation of a Natural Language Processing Annotation Tool to Facilitate Phenotyping of Cognitive Status in Electronic Health Records: Diagnostic Study
J Med Internet Res 2022;24(8):e40384
URL: https://www.jmir.org/2022/8/e40384
doi:10.2196/40384
PMID:36040790

Original Paper

The Impact of a Place-Tailored Digital Health App Promoting Exercise Classes on African American Women's Physical Activity and Obesity: Simulation Study

Tiffany M Powell-Wiley[1,2], MD, MPH; Marie F Martinez[3,4], MSPH; Kosuke Tamura[5], PhD; Sam J Neally[1], BA; Kelly J O'Shea[3,4], BSFS; Kaveri Curlin[1], BS; Yardley Albarracin[3], MPH; Nithya P Vijayakumar[1], MD; Matthew Morgan[3], MBE; Erika Ortiz-Chaparro[1], MD; Sarah M Bartsch[3,4], MPH; Foster Osei Baah[1], PhD; Patrick T Wedlock[3,4], MSPH; Lola R Ortiz-Whittingham[1], BS; Sheryl Scannell[3,4], MS; Kameswari A Potharaju[1], BA; Samuel Randall[3], BS; Mario Solano Gonzales[3], BS; Molly Domino[3], BS; Kushi Ranganath[3], BS; Daniel Hertenstein[3], BS; Rafay Syed[3], MS; Colleen Weatherwax[3,4], MS; Bruce Y Lee[3,4], MD, MBA

[1]Social Determinants of Obesity and Cardiovascular Risk Laboratory, Cardiovascular Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, United States

[2]Intramural Research Program, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States

[3]Public Health Informatics, Computational, and Operations Research, CUNY Graduate School of Public Health and Health Policy, New York, NY, United States

[4]Center for Advanced Technology and Communication in Health, CUNY Graduate School of Public Health and Health Policy, New York, NY, United States

[5]Socio-Spatial Determinants of Health (SSDH) Laboratory, Population and Community Sciences Branch, Intramural Research Program, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States

Corresponding Author:
Tiffany M Powell-Wiley, MD, MPH
Social Determinants of Obesity and Cardiovascular Risk Laboratory
Cardiovascular Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute
National Institutes of Health
10 Center Drive
Bethesda, MD, 20892
United States
Phone: 1 (301) 547 5275
Email: tiffany.powell-wiley@nih.gov

Abstract

Background: The increasing prevalence of smartphone apps to help people find different services raises the question of whether apps to help people find physical activity (PA) locations would help better prevent and control having overweight or obesity.

Objective: The aim of this paper is to determine and quantify the potential impact of a digital health intervention for African American women prior to allocating financial resources toward implementation.

Methods: We developed our Virtual Population Obesity Prevention, agent-based model of Washington, DC, to simulate the impact of a place-tailored digital health app that provides information about free recreation center classes on PA, BMI, and overweight and obesity prevalence among African American women.

Results: When the app is introduced at the beginning of the simulation, with app engagement at 25% (eg, 25% [41,839/167,356] of women aware of the app; 25% [10,460/41,839] of those aware downloading the app; and 25% [2615/10,460] of those who download it receiving regular push notifications), and a 25% (25/100) baseline probability to exercise (eg, without the app), there are no statistically significant increases in PA levels or decreases in BMI or obesity prevalence over 5 years across the population. When 50% (83,678/167,356) of women are aware of the app; 58.23% (48,725/83,678) of those who are aware download it; and 55% (26,799/48,725) of those who download it receive regular push notifications, in line with existing studies on app usage, introducing the app on average increases PA and decreases weight or obesity prevalence, though the changes are not statistically significant. When app engagement increased to 75% (125,517/167,356) of women who were aware, 75% (94,138/125,517) of those who were aware downloading it, and 75% (70,603/94,138) of those who downloaded it opting into the app's push notifications, there were statistically significant changes in PA participation, minutes of PA and obesity prevalence.

Conclusions: Our study shows that a digital health app that helps identify recreation center classes does not result in substantive population-wide health effects at lower levels of app engagement. For the app to result in statistically significant increases in PA and reductions in obesity prevalence over 5 years, there needs to be at least 75% (125,517/167,356) of women aware of the app, 75% (94,138/125,517) of those aware of the app download it, and 75% (70,603/94,138) of those who download it opt into push notifications. Nevertheless, the app cannot fully overcome lack of access to recreation centers; therefore, public health administrators as well as parks and recreation agencies might consider incorporating this type of technology into multilevel interventions that also target the built environment and other social determinants of health.

(J Med Internet Res 2022;24(8):e30581) doi:10.2196/30581

KEYWORDS

computational modeling; digital health; physical activity; BMI; obesity; built environment; impact; app; exercise; simulation; intervention; women; African American; agent

Introduction

The increasing prevalence of smartphone apps to help people find different services (eg, Yelp and OpenTable to find restaurants, Fandango to find movie theaters, AllTrails to find hikes, GasBuddy to find gas stations, Expedia to find hotels, and Zillow to find homes and apartments) raises the question of whether apps to help people find physical activity (PA) locations (eg, ClassPass [1] and Fit Reserve [2]) would help to better prevent and control having overweight and obesity. Such place-tailored apps can help assemble, collate, and present information that may be available on different websites so that an individual can quickly find the closest location of interest. These place-tailored apps can be particularly helpful for PA locations and opportunities since they may exist in different and less obvious forms (eg, irregular timing of classes, walking and bike paths, outdoor tracks, and tennis or basketball courts). Such an app can also offer crowdsourced ratings of each location, details about specific services (eg, time, availability, costs, promotions, and deals), and even social connections with people who have the same interests or are in the same area. Previous studies have shown that people may not be aware of or have difficulty finding locations to engage in PA [3-5]. This may be the case in underresourced and otherwise disadvantaged communities where parks, affordable gyms, and other opportunities may be more difficult to find if they are in less-frequented or obscure locations, or if they are not regularly advertised or promoted [6]. African American women who live disproportionately in underresourced communities spend at least as much time as any other racial or ethnic group using apps and the internet (approximately 19 hours and 27 minutes each week versus 17 hours and 8 minutes each week), and approximately 80% of African American women own a smartphone [7], raising the possibility that this could be an effective means to help these women find PA opportunities. However, before such an app is rolled out in the "real world," it can be helpful to use simulation modeling to guide the design and test the potential impact of such an app. Such an approach is used in other fields (eg, aeronautical engineering and manufacturing) since running simulation models can take much less time and can be significantly less costly than conducting a real-world trial (which can take months to set up, recruit for, and implement). Moreover, once a trial is completed, one cannot go back and change the circumstances as they can in a simulation model. Therefore, we further developed our agent-based simulation model of Washington, DC to test the impact of such a place-tailored digital health app.

Methods

Ethics Approval

All authors' institutions were included in the institutional review board approval (IRB #00004203) at Johns Hopkins as the study began while certain members of the research team (MCF, KJO, YA, MM, SMB, PTW, SS, SR, MSG, MD, KR, DH, RS, and BYL) were based at Johns Hopkins.

Model of Washington, DC

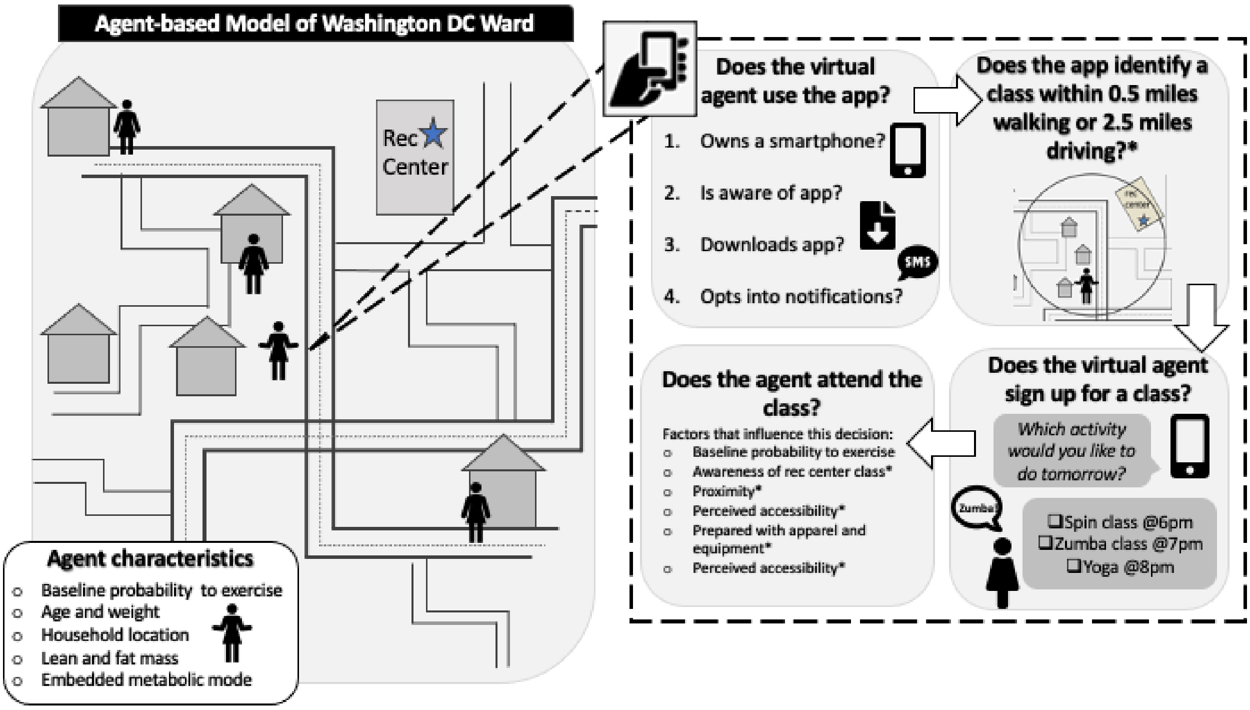
We used and further developed a Virtual Population Obesity Prevention, agent-based model of Washington, DC in 2020-2021 [8,9], which includes computer model–based representations of households, workplaces, and recreation centers throughout all 8 wards (similar to districts in other cities) in Washington, DC.

Agents Representing People

We represented each of the 167,356 African American women (aged 18-65 years) living in Washington, DC with a computer model–based agent. Each agent (ie, each African American woman in Washington, DC) has attributes for age, height, lean or fat mass, household location, work location, and income based on representative data for the region and population. Each agent also has an embedded metabolic model, which converts daily caloric intake and expenditure to corresponding lean or fat mass [10,11]. Caloric expenditure from exercise is determined by exercise intensity, duration, and the agent's current body weight [10,11]. Since individuals may vary in their inclination to exercise, each agent had a baseline probability of wanting to exercise each day. This accounts for an agent's past experiences and existing tendencies to exercise and includes factors such as household financial status, family responsibilities, chronic health conditions, and social influences. Different scenarios ranged this baseline probability from 10% (10/100) to 50% (50/100) to explore how this probability might affect the results.

In each simulated day, women may participate in a recreation center class, depending on a number of factors (Figure 1; Multimedia Appendix 1, Table S1 [8, 9, 12-23]), including the following: (1) her baseline probability to exercise (this accounts for an agent's past experience and existing tendencies to participate in recreation center classes), which we vary between simulation experiments; (2) objective accessibility to locations, based on the geographic locations of recreation centers [12], the distance individuals need to travel to reach these locations, and access to the types of transportation (eg, cars) that might be required to reach locations further away [24,25]; (3) perceived accessibility of locations [15], based on the individual's understanding and knowledge of nearby recreation centers; (4) awareness of classes at recreation centers; and (5) preparedness to exercise (whether or not she remembers her apparel and equipment).

Figure 1. A digital health app that helps locate and send reminders about recreation (rec) center classes. *Factors influenced by phone app.



Representations of Recreation Centers

Multimedia Appendix 1, Table S2 shows key characteristics (eg, number of recreation centers) for each ward. If an agent ultimately participates in a recreation center class, she is active for 50 minutes [12] at an intensity of approximately 6.5 metabolic equivalents [16].

Representations of Digital Health App

In the model, we represent a digital health app that helps locate and send reminders about in-person recreation center classes to increase the agents' likelihood of participation (Figure 1). This mobile app uses a database of public locations that have been previously identified as locations for PA, such as recreation centers in the case of this paper. Once this registry is verified, a geofence—a geographic boundary—can be created within the mobile app with a set distance surrounding the chosen location; in this case, our simulated app searched for recreation centers within 0.5 miles of the user. When the simulated mobile app detects that the user is within this defined boundary, it will generate a notification that will alert the user of the available resources in the area. Unlike existing fitness apps, this digital health app considers the geographic location of the user and the recreation centers to connect agents with recreation center classes that align with their neighborhood environment and schedules; prompts users to remind them about upcoming classes and what equipment they will need; and provides individually tailored information about class time, location, and necessary equipment to maximize user engagement (Figure 1). If an agent has a smartphone, downloads the app, and opts into notifications (Multimedia Appendix 1, Table S1), the app will send a question each evening asking the user which activity or class she would like to participate in the following day, thereby increasing an agent's knowledge of class schedules. After selecting the class, agents will receive a notification with a reminder of the class's time, location, and activity, as well as a reminder to bring clothes or equipment, thereby increasing an agent's probability of being prepared for and attending class. When representing the digital health app, we introduced it at the beginning of the 5-year simulation, but not all participants continued to use the app for the entire simulation duration (eg, we represented attrition, people discontinuing app use, during the 3 months following the introduction of the app; Multimedia Appendix 1, Table S1).

Representations of Engagement With the Digital Health App

Since only a certain percentage of the population may be aware that the app is available, we varied the proportion of women across the population who, in a given scenario, were aware of the app, subsequently downloaded the app, and then opted into push notifications (25%-75%). This means, 25% (41,839/167,356) of women are aware of the app, 25%

(10,460/41,841) of those who are aware download it, and 25% (2615/10,460) of those who download it receive regular push notifications from the app. We ranged this to 75% (125,517/167,356) of women aware of the app, 75% (94,138/125,517) of those who are aware download it, and 75% (70,603/94,138) of those who download it opt into the app's push notifications. Varying the level of user engagement across a range can help identify the thresholds of app engagement that result in observable and statistically significant impacts on PA and weight.

Simulation Experiments

We used the model of Washington, DC to simulate the impact of a digital health app on in-person recreation center class participation, recreation center class PA (minutes per week), subsequent changes in BMI, as well as the prevalence of obesity and the state of having overweight. Each simulation experiment consisted of running the model of Washington, DC and all 167,356 computer model–based agents, 10 times over 5 simulated years.

Validation

Validation consisted of comparing different model-generated metrics to observed values to determine if the model was representing what was occurring. For example, when we ran simulation runs, we saw that, on average, 2.1% (3514/167,356) of women were participating in recreation center classes daily compared to the observed 3.8% from the 2017 American Time Use Survey [17]. Since the people who exercised on one day will not necessarily be the same people who exercised on a different day, there will be a certain proportion of the population that exercised at least once over the course of the month. Thus, we also simulated the average percentage of women participating in recreation center classes at least once on a monthly basis (19.1% [31,965/176,356]) and compared this to the observed proportion of women participating in workout class activity on a monthly basis (16.1%), as reported by the Behavioral Risk Factors Surveillance System [18]. The model-generated data generally matched the observed data, and the differences are likely due to differences between populations and the classes available to that population. Further model validation details are available in Multimedia Appendix 1.

Results

No Mobile App

Table 1 shows PA from recreation center classes and weight-related outcomes after 5 years with no mobile app for different baseline probabilities to exercise. Figure 2 shows how the percent of women who exercised at least once in the simulation when there was no app varied by the Washington, DC ward. For example, Ward 6 had the highest percent of the population who exercised at least once (69.1% [4331/20,739], 95% CI 68.9%-69.2%), while Ward 7 had the lowest (48% [15,710/32,729], 95% CI 47.9%-48.0%) when there was no mobile app (25% [25/100] baseline probability to exercise). This trend in ward-level variation was consistent across all baseline exercise probabilities.

Table 1. Physical activity, overweight, obesity, BMI outcomes by baseline probability to exercise for different scenarios (eg, with and without digital health app).

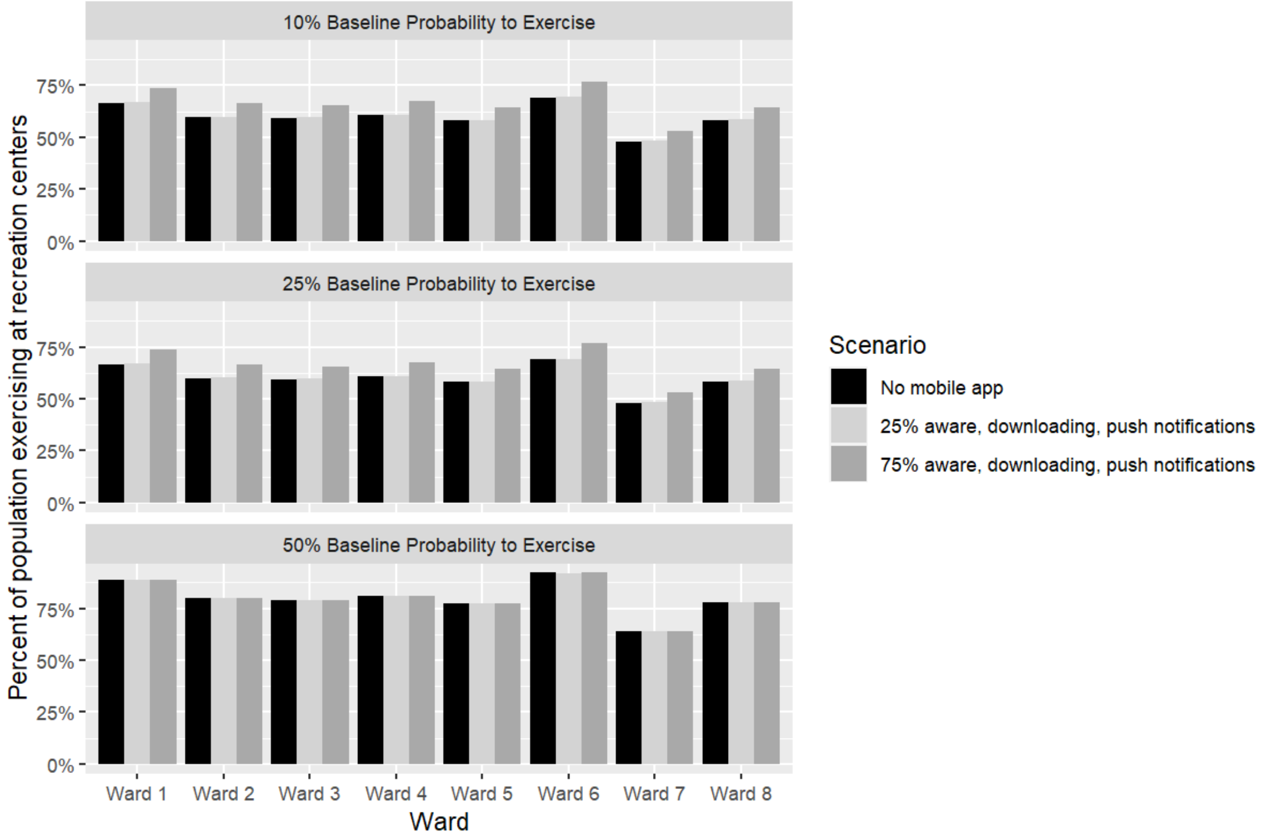
Simulation scenarios at each baseline probability to exercise	Percent of population exercising at recreation centers, mean (95% CI)	Average number of physical activity min/week, mean (95% CI)	Overweight prevalence, mean (95% CI)	Obesity prevalence, mean (95% CI)	Average BMI, mean (95% CI)	Average BMI among women with obesity, mean (95% CI)
10% (10/100) baseline probability to exercise						
No digital health app	58.66 (54.65-62.67)	36.97 (34.45-39.50)	24.44 (23.91-24.97)	56.10 (54.56-57.64)	30.16 (29.86-30.45)	34.20 (34.00-34.41)
Introducing place-tailored digital health app						
25%-25%-25%[a]	58.91 (54.87-62.94)	37.26 (34.71-39.81)	24.42 (23.88-24.96)	56.09 (54.53-57.65)	30.15 (29.86-30.45)	34.21 (34.00-34.43)
50%-50%-50%[b]	61.09 (56.92-65.26)	39.83 (37.12-42.54)	24.45 (23.91-24.98)	55.67 (54.15-57.19)	30.07 (29.78-30.36)	34.16 (33.94-34.37)
75%-75%-75%[c]	65.10 (60.64-69.56)	44.45 (41.41-47.50)	24.70 (24.21-25.20)	54.68 (53.12-56.25)	29.90 (29.60-30.19)	34.04 (33.83-34.26)
25% (25/100) baseline probability to exercise						
No digital health app	58.67 (54.66-62.68)	52.84 (49.23-56.45)	25.52 (25.04-26.01)	52.75 (51.06-54.43)	29.56 (29.27-29.86)	33.81 (33.61-34.01)
Introducing place-tailored digital health app						
25%-25%-25%[a]	58.92 (54.89-62.94)	53.25 (49.61-56.89)	25.54 (25.04-26.04)	52.62 (50.91-54.33)	29.56 (29.26-29.86)	33.83 (33.62-34.05)
50%-50%-50%[b]	61.17 (56.99-65.35)	56.98 (53.09-60.88)	26.24 (25.68-26.80)	51.25 (49.47-53.04)	29.44 (29.14-29.74)	33.83 (33.62-34.03)
75%-75%-75%[c]	65.10 (60.64-69.55)	63.52 (59.18-67.87)	27.72 (27.05-28.40)	48.66 (46.75-50.56)	29.23 (28.92-29.54)	33.82 (33.63-34.02)
50% (50/100) baseline probability to exercise						
No digital health app	78.30 (72.96-83.64)	86.33 (80.43-92.22)	27.88 (26.36-29.39)	44.42 (41.63-47.20)	28.38 (28.06-28.70)	33.00 (32.80-33.21)
Introducing place-tailored digital health app						
25%-25%-25%[a]	78.22 (72.86-83.58)	86.88 (80.93-92.83)	28.24 (26.81-29.67)	43.90 (41.27-46.52)	28.38 (28.05-28.71)	33.08 (32.84-33.31)
50%-50%-50%[b]	78.25 (72.90-83.60)	92.17 (85.89-98.45)	28.57 (27.17-29.96)	42.63 (40.03-45.23)	28.24 (27.91-28.57)	33.10 (32.86-33.34)
75%-75%-75%[c]	78.29 (72.95-83.63)	101.41 (94.48-108.33)	29.40 (28.15-30.66)	40.27 (37.75-42.78)	28.00 (27.65-28.34)	33.15 (32.91-33.40)

[a] 25% (41,839/167,356) aware of the app, 25% (10,460/41,839) of those who are aware download the app, and 25% (2615/10,460) of those who download it receive notifications.

[b] 50% (83,678/167,356) aware of the app, 50% (48,725/83,678) of those who are aware download the app, and 50% (26,799/48,725) of those who download it receive notifications.

[c] 75% (125,517/167,356) aware of the app, 75% (94,138/125,517) of those who are aware download app, and 75% (70,603/94,138) of those who download it receive app notifications.

Figure 2. Percent of women exercising with and without the mobile app within each ward in Washington, DC.



Impact of Introducing a Place-Tailored Mobile App That Connects Users to Recreation Center Classes

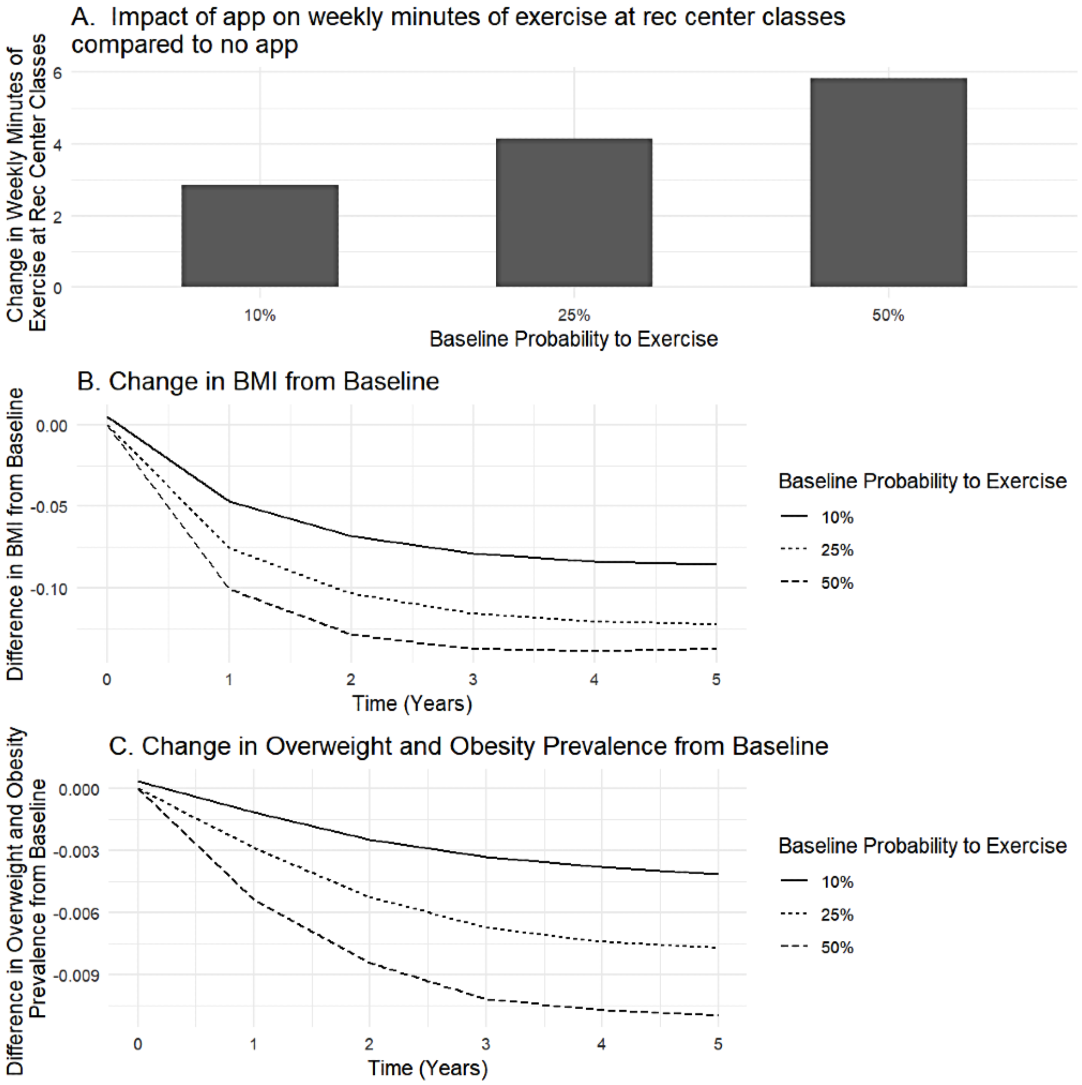
With lower levels of user engagement with the mobile phone app, that is 25% aware of app (41,839/167,356), 25% of those aware download app (10,460/41,839), and 25% (2615/10,460) of those who download it receive app notifications, the app had a negligible and nonsignificant impact on the additional minutes of PA (<1 minute), on the additional percent of women who ever exercise (0.2% [335/167,356]; Figure 2), and on reductions in obesity prevalence (0.1% [167/167,356]). Thus, even lower levels of app engagement (eg, below 25% [25/100]) would have no effect on physical activity and weight.

Increasing user engagement to approximately 50% (eg, 50% aware [83,678/167,356], 58.23% [48,725/83,678] of those who are aware download the app [19], and 55% [26,799/48,725] of those who download it receive regular push notifications [20]) resulted in moderate improvements to PA from recreation center classes and weight-related outcomes across the population. Figure 3 shows these observable changes to PA (panel A), BMI (panel B), and overweight and obesity prevalence (panel C). With a 10% (10/100) baseline probability of exercise, the PA minutes per week increase by 2.9 minutes (95% CI –1.4 to 17.9), BMI decreases by 0.09 kg/m^2 (95% CI –0.56 to 0.39), and obesity prevalence decreases by an absolute 0.43% (720/167,356; 95% CI –2.7% to 2.93%) at the end of the 5-year simulation. When baseline probability increases to 50% (50/100), there are larger increases in weekly PA minutes (5.4 minutes, 95% CI –4.1 to 15.8), and larger reductions in BMI (0.14 kg/m^2, 95% CI –0.67 to –0.4) and obesity prevalence (1.8% [3012/167,356]; 95% CI –2.6% to 6.2%).

The percent of women attending at least one recreation center class over the course of the simulation shows additional gains between when the baseline probability to exercise is between 10% (10/100; 2.43% [4067/167,356], 95% CI –4.24% to 9.1%) and 25% (25/100; 2.5% [4184/167,356]; 95% CI –4.2% to 9.2%). When the baseline probability to exercise is 50% (50/100), the percent of women exercising at least once hits a ceiling of 78% (130,538/167,356) (increase of 0.05% [84/167,356]; 95% CI –8.68% to 8.77%), due to the location and accessibility of recreation centers for some women. Thus, at lower probabilities to exercise (eg, 10%-25%), the app is more effective at increasing the number of women participating in at least one recreation center class (Figure 2). However, additional PA minutes per week from recreation center classes increase with baseline probability to exercise (eg, 4.14, 95% CI –1.9 to 10.2 vs 5.9, 95% CI –4.1 to 15.7 minutes per week at 25% [25/100] and 50% [50/100] baseline probabilities to exercise, respectively; Figure 3). Figure 3 also shows how reductions in BMI and overweight and obesity prevalence due to app use accrue over time during the 5-year simulation.

Further increasing app engagement to 75%, with 75% (125,517/167,356) of women aware of the app, 75% (94,138/125,517) of those who are aware downloading the app, and 75% (70,603/94,138) of those who download it opting into the app's push notifications resulted in statistically significant gains to PA and reductions in obesity prevalence. For example, weekly PA increased by 10.7 (95% CI 4.2-17.2) minutes per week, and obesity prevalence decreased by an absolute 4.09% (6,845/167,356; 95% CI 1.2%-7.0%) with 25% baseline exercise probability (Table 1).

Figure 3. Impact of mobile app on physical activity, BMI, as well as overweight and obesity prevalence at each baseline probability to exercise. Rec: recreation.



Ward-Level Impact of Place-Tailored Mobile App

The results varied substantially by ward. For example, at 25% (25/100) baseline probability to exercise (assuming 50% [83,678/167,356] aware, 50% [48,725/83,678] of those who are aware downloading the app, and 50% [26,799/48,725] of those who download it receiving app notifications), Ward 6 had the highest absolute increase in average PA minutes per week (4.85, 95% CI 4.58-5.11), and the greatest reduction in average BMI (–0.15 kg/m^2, 95% CI –0.19 to –0.11). However, Ward 7 had the lowest (3.39, 95% CI 3.24-3.53) increase in PA minutes per week and the smallest reduction in BMI (–0.09 kg/m^2; 95% CI –0.12 to –0.06). Changes in overweight and obesity prevalence also varied between wards and decreased by as much as 2.6% (539/20,739; 95% CI 2.3%-2.9%) in Ward 6, where participation in recreation center classes was highest and as little as 1.9% (622/32,729; 95% CI 1.7%-2.1%) in Ward 7 (25% baseline exercise probability).

Discussion

Principal Findings

Our simulation model of African American women in Washington, DC, and their use of a place-tailored digital health app to help identify recreation center classes shows that the app does not result in substantive population-wide health effects at lower levels of app engagement (eg, 25% of women are aware of the app, 25% of those aware of the app download it, and 25% of those who download it receive regular push notifications from the app). When 50% of women are aware of the app, 58.23% of those who are aware download the app, and 55% of those who download it receive regular push notifications from the app, there are observable changes in PA and weight across the population, but the impact is not statistically significant. For the app to result in statistically significant increases in PA and reductions to obesity prevalence over 5 years, there needs to be at least 75% of women who are aware of the app, 75% of those aware of the app downloading it, and 75% of those who download it opting into the app's push notifications. Thus, we demonstrated the minimum levels of engagement needed at the outset of a mobile phone app campaign (approximately 50% aware of the app, 50% of those who are aware download the app, and 50% of those who download it receive app notifications, assuming reductions in use over the first 3 months) to observe a change in PA and weight across the population. Studies have shown how perceived usefulness of an app, user-friendliness, backing from health care professionals, and continued engagement impact app usage [26,27] could be addressed through a structured marketing and communications strategy. Thus, future interventions should prioritize efforts to increase marketing for the place-tailored app to increase the percent of women who are aware of and use the app to reach the impactful threshold of engagement and obtain further benefits.

Further, our results show that a place-tailored app is more likely to be successful in increasing PA in those who already have a higher likelihood to exercise. While the results showed that the app was successful at encouraging individuals who have a low baseline probability (eg, 10% [10/100] and 25% [25/100]) to exercise to attend at least one new class over the course of the simulated period, this alone was not enough to drive a sustained change in regular exercise. The app did a better job at increasing the average duration of PA each week as baseline probability to exercise increased. This indicates that improving knowledge of recreation center classes, while important, should be coupled with interventions to help overcome personal and social barriers (eg, limited social support for PA or time constraints) that determine baseline exercise probability [28,29]. Place-tailored digital health apps could potentially address some of these barriers through the release of new features and functionality such as a social networking component [30,31].

Regardless of user engagement with the app, place-tailored digital health apps need to be combined with increasing physical access to recreation centers to see greater than additive effects in PA and subsequent health outcomes. There is a limit to a place-tailored app's impact because some individuals cannot access recreation centers due to the distance and lack of transportation (eg, access to car) from their home location. As shown in our results, there are clear disparities in the success of the app in improving health outcomes in neighborhoods with greater access to recreation centers (with nearly a 1.4-fold increase in the use of recreation center classes in these neighborhoods [eg, Ward 6]) compared to neighborhoods with less accessible recreation centers (eg, Ward 7), even with 75% of women who are aware of the app, 75% of those aware of the app download it, and 75% of those who download it opt into the app's push notifications. Past studies have shown that lower-income neighborhoods in many cities around the United States have less accessible PA locations and recreation centers [32]. Therefore, it is important for public health administrators and park and recreation agencies to consider pairing this type of digital health technology with improvements to recreation center access such as changes to the built environment, perceived safety, or transportation.

Our results also show that it takes time for the effect of the place-tailored mobile app to fully manifest (>2 years). In general, 1 year is not enough time to see an impact on BMI and overweight and obesity prevalence, as population-level effects on weight and subsequent health benefits accrue over years. This shows the need to continuously measure the value of intervention programs over a period of several years, since reductions in overweight and obesity prevalence may not be demonstrated immediately, and effective interventions may wrongly be deemed unsuccessful if evaluated too early. Accounting for this ramp-up period is important, as it can also take time for a new technology to be adopted and used. Our results show that the speed of the reduction in overweight and obesity prevalence in the population increases year after year as adoption rates increase, revealing a potential opportunity to increase momentum as more users adopt similar place-tailored digital health technology.

In addition to being able to simulate extended periods of time, another benefit of simulation modeling is that it can be adapted and refined over time. For example, simulation modeling can be used in conjunction with clinical trials [33,34] so that the model can continuously inform digital health phone app design and multipronged PA interventions. The simulation model can be run first, to help determine the impact of an app, which can then inform the implementation of a trial. Data and information from the trial can then further update the model. This iterative process can continue until the app or intervention is optimally designed.

Limitations

All models are simplifications of reality and cannot account for all possible factors that may affect PA decision-making. Our model included a few simplifying assumptions. For example, we did not account for objective accessibility to a recreation center near a woman's workplace and used the objective accessibility near the home as a proxy. In addition, since we wanted to demonstrate how to design an app that harnesses geographic location and the value of such an app, our study focused on the app locating and reminding individuals about in-person classes, rather than web-based classes. However, such an app may offer similar benefits for web-based classes such as reminding individuals about when classes are scheduled and what equipment is needed, while reducing potential geographic barriers to exercise. We also assumed that in-person classes are available (eg, not during a public health emergency such as the COVID-19 pandemic). When determining body weight changes for each woman, we assumed that compensatory eating did not occur. Our model simulated behavior of and used data specific to Washington, DC African American women, which may limit generalizability to other populations or geographic areas.

Conclusions

Our study shows that a digital health app that helps identify recreation center classes does not result in substantive population-wide health effects at lower levels of app engagement (eg, 25% of women who are aware of the app, 25% of those who are aware of the app download it, and 25% of those who download it receive regular push notifications from the app). For the app to result in statistically significant increases in PA and reductions to obesity prevalence over 5 years, there needs to be at least 75% of women aware of the app, 75% of those aware of the app download it, and 75% of those who download it opt into the app's push notifications. Even so, the app cannot fully overcome lack of access to recreation centers, and therefore, public health administrators as well as parks and recreation agencies might consider incorporating this type of technology into multilevel interventions that also target the built environment and other social determinants of health.

Acknowledgments

This project was supported by the National Institutes of Health (NIH) Intramural Research Program via grants ZIA HL006168, ZIA HL006225, ZIA HL006252, ZIA MD000010, and ZIA MD000020. The Social Determinants of Obesity and Cardiovascular Risk Laboratory is funded by the Division of Intramural Research at the National Heart, Lung, and Blood Institute (NHLBI) and the Intramural Research Program of the National Institute on Minority Health and Health Disparities (NIMHD). The Socio-Spatial Determinants of Health (SSDH) Laboratory is supported by the Intramural Research Program, NIMHD, NIH, and by the NIH Distinguished Scholars Programs.

This project was also supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) via grant U01HD086861 and 5R01HD086013-02, as well as the NICHD and Office of Behavioral and Social Sciences Research (OBSSR) under award number U54HD070725, the Agency for Healthcare Research and Quality (AHRQ) via grant 1R01HS028165-01, Models of Infectious Disease Agent Study (MIDAS) network under grant R01GM127512, as well as the NIH Common Fund, and the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number U54TR004279. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, US Department of Health and Human Services, NHLBI, NIMDH, NICHD, AHRQ, OBSSR, and MIDAS. None of the study sponsors had any role in the study design, collection, analysis, and interpretation of data, writing the report, nor the decision to submit the report for publication. All authors confirm that they had full access to all the data in the study and accept responsibility to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 21 KB - jmir_v24i8e30581_app1.docx]

References

1. ClassPass. URL: https://classpass.com/ [accessed 2022-07-13]
2. The premier fitness membership for studio and online classes. FitReserve. URL: https://www.fitreserve.com/ [accessed 2022-07-13]
3. James DCS, Harville C. Smartphone Usage, Social Media Engagement, and Willingness to Participate in mHealth Weight Management Research Among African American Women. Health Educ Behav 2018 Jun 13;45(3):315-322. [doi: 10.1177/1090198117714020] [Medline: 28606004]
4. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
5. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar 24;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
6. Estabrooks PA, Lee RE, Gyurcsik NC. Resources for physical activity participation: Does availability and accessibility differ by neighborhood socioeconomic status? ann. behav. med 2003 Apr;25(2):100-104. [doi: 10.1207/s15324796abm2502_05]
7. Reaching Black women across media platforms. Nielsen. URL: https://www.nielsen.com/us/en/insights/article/2017/reaching-black-women-across-media-platforms/ [accessed 2022-07-13]
8. Powell-Wiley TM, Wong MS, Adu-Brimpong J, Brown ST, Hertenstein DL, Zenkov E, et al. Simulating the Impact of Crime on African American Women's Physical Activity and Obesity. Obesity (Silver Spring) 2017 Dec;25(12):2149-2155 [FREE Full text] [doi: 10.1002/oby.22040] [Medline: 29086471]
9. Lee BY, Ferguson MC, Hertenstein DL, Adam A, Zenkov E, Wang PI, et al. Simulating the Impact of Sugar-Sweetened Beverage Warning Labels in Three Cities. Am J Prev Med 2018 Feb;54(2):197-204 [FREE Full text] [doi: 10.1016/j.amepre.2017.11.003] [Medline: 29249555]

10. Hall KD, Butte NF, Swinburn BA, Chow CC. Dynamics of childhood growth and obesity: development and validation of a quantitative mathematical model. The Lancet Diabetes & Endocrinology 2013 Oct;1(2):97-105. [doi: 10.1016/s2213-8587(13)70051-2]
11. Rahmandad H. Human growth and body weight dynamics: an integrative systems model. PLoS One 2014 Dec 5;9(12):e114609 [FREE Full text] [doi: 10.1371/journal.pone.0114609] [Medline: 25479101]
12. WebTrac Facility Search. DC Department of Parks and Recreation. URL: http://app.dpr.dc.gov/dprmap/details.asp?cid=2 [accessed 2019-06-14]
13. Wheaton WD, Cajka JC, Chasteen BM, Wagener DK, Cooley PC, Ganapathi L, et al. Synthesized Population Databases: A US Geospatial Database for Agent-Based Models. Methods Rep RTI Press 2009 May 01;2009(10):905 [FREE Full text] [doi: 10.3768/rtipress.2009.mr.0010.0905] [Medline: 20505787]
14. Adults Health and Wellness. DC Department of Parks and Recreation. URL: https://dpr.dc.gov/page/adults-health-wellness
15. Ries AV, Yan AF, Voorhees CC. The neighborhood recreational environment and physical activity among urban youth: an examination of public and private recreational facilities. J Community Health 2011 Aug 12;36(4):640-649. [doi: 10.1007/s10900-010-9355-1] [Medline: 21225329]
16. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Montoye HJ, Sallis JF, et al. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc 1993 Jan;25(1):71-80. [doi: 10.1249/00005768-199301000-00011] [Medline: 8292105]
17. American time use survey: arts activities, [United States], 2003-2017 (ICPSR 36268). National Archive of Data on Arts & Culture. 2017. URL: https://www.icpsr.umich.edu/web/NADAC/studies/36268/versions/V4 [accessed 2022-07-13]
18. Behavioral Risk Factor Surveillance System. DC Health. URL: https://dchealth.dc.gov/service/behavioral-risk-factor-surveillance-system [accessed 2018-01-01]
19. 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]
20. More than half of U.S. smartphone users ever get news alerts; about half of those who get them click for more information. Pew Research Center. 2016 Sep 07. URL: https://www.pewresearch.org/fact-tank/2016/09/08/more-than-half-of-smartphone-users-get-news-alerts-but-few-get-them-often/ft_16-09-08_newsalerts_main [accessed 2022-07-13]
21. Alpert D. Latest data shows plenty of car-free living in DC. Greater Greater Washington. 2012 Feb 21. URL: https://ggwash.org/view/12290/latest-data-shows-plenty-of-car-free-living-in-dc
22. Smith A. Record shares of Americans now own smartphones, have home broadband. PEW Research Center. 2017 Jan 12. URL: https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/
23. Perro J. Mobile Apps: What's A Good Retention Rate? Upland. 2016. URL: https://uplandsoftware.com/localytics/resources/blog/mobile-apps-whats-a-good-retention-rate/
24. Burke M, Brown AL. Distances people walk for transport. Road & Transport Research: A Journal of Australian and New Zealand Research and Practice 2007 Sep;16(3):16-29.
25. Yang Y, Diez-Roux AV. Walking distance by trip purpose and population subgroups. Am J Prev Med 2012 Jul;43(1):11-19 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.015] [Medline: 22704740]
26. Kim SC, Yoon D, Han EK. Antecedents of mobile app usage among smartphone users. Journal of Marketing Communications 2014 Aug 22;22(6):653-670. [doi: 10.1080/13527266.2014.951065]
27. McLean G. Examining the determinants and outcomes of mobile app engagement - A longitudinal perspective. Computers in Human Behavior 2018 Jul;84:392-403. [doi: 10.1016/j.chb.2018.03.015]
28. Michie S, West R. Behaviour change theory and evidence: a presentation to Government. Health Psychology Review 2013 Mar;7(1):1-22. [doi: 10.1080/17437199.2011.649445]
29. Suls J, Mogavero JN, Falzon L, Pescatello LS, Hennessy EA, Davidson KW. Health behaviour change in cardiovascular disease prevention and management: meta-review of behaviour change techniques to affect self-regulation. Health Psychol Rev 2020 Mar 29;14(1):43-65 [FREE Full text] [doi: 10.1080/17437199.2019.1691622] [Medline: 31707938]
30. Cho H, Porras T, Baik D, Beauchemin M, Schnall R. Understanding the predisposing, enabling, and reinforcing factors influencing the use of a mobile-based HIV management app: A real-world usability evaluation. Int J Med Inform 2018 Sep;117:88-95 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.007] [Medline: 30032969]
31. Tong HL, Coiera E, Laranjo L. Using a Mobile Social Networking App to Promote Physical Activity: A Qualitative Study of Users' Perspectives. J Med Internet Res 2018 Dec 21;20(12):e11439 [FREE Full text] [doi: 10.2196/11439] [Medline: 30578201]
32. Moore LV, Diez Roux AV, Evenson KR, McGinn AP, Brines SJ. Availability of recreational resources in minority and low socioeconomic status areas. Am J Prev Med 2008 Jan;34(1):16-22 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.021] [Medline: 18083446]
33. Lee BY, Bartsch SM, Mui Y, Haidari LA, Spiker ML, Gittelsohn J. A systems approach to obesity. Nutr Rev 2017 Jan 03;75(suppl 1):94-106 [FREE Full text] [doi: 10.1093/nutrit/nuw049] [Medline: 28049754]

34. Lee BY, Ferguson MC, Cox SN, Phan PH. Big Data and Systems Methods: The Next Frontier to Tackling the Global Obesity Epidemic. Obesity (Silver Spring) 2021 Feb 09;29(2):263-264 [FREE Full text] [doi: 10.1002/oby.23062] [Medline: 33421308]

Abbreviations

PA: physical activity

Edited by A Mavragani; submitted 21.05.21; peer-reviewed by L Rovniak, A Chwałczyńska; comments to author 30.07.21; revised version received 17.12.21; accepted 30.06.22; published 22.08.22.

Please cite as:
Powell-Wiley TM, Martinez MF, Tamura K, Neally SJ, O'Shea KJ, Curlin K, Albarracin Y, Vijayakumar NP, Morgan M, Ortiz-Chaparro E, Bartsch SM, Osei Baah F, Wedlock PT, Ortiz-Whittingham LR, Scannell S, Potharaju KA, Randall S, Solano Gonzales M, Domino M, Ranganath K, Hertenstein D, Syed R, Weatherwax C, Lee BY
The Impact of a Place-Tailored Digital Health App Promoting Exercise Classes on African American Women's Physical Activity and Obesity: Simulation Study
J Med Internet Res 2022;24(8):e30581
URL: https://www.jmir.org/2022/8/e30581
doi:10.2196/30581
PMID:35994313

Original Paper

A Questionnaire-Based Ensemble Learning Model to Predict the Diagnosis of Vertigo: Model Development and Validation Study

Fangzhou Yu[1*], MD; Peixia Wu[2*], PhD; Haowen Deng[3*], BMS; Jingfang Wu[1,4,5], PhD; Shan Sun[5,6], PhD; Huiqian Yu[1,4,5], PhD; Jianming Yang[7], MMed; Xianyang Luo[8], MMed; Jing He[8], MMed; Xiulan Ma[9], MD; Junxiong Wen[9], MMed; Danhong Qiu[10], MMed; Guohui Nie[11], PhD; Rizhao Liu[11], MMed; Guohua Hu[12], PhD; Tao Chen[12], PhD; Cheng Zhang[3*], PhD; Huawei Li[1,4,5,6,13*], MD, PhD

[1]Department of Otorhinolaryngology, Eye & ENT Hospital, Fudan University, Shanghai, China
[2]Nursing Department, Eye & ENT Hospital, Fudan University, Shanghai, China
[3]Department of Information Management and Information Systems, Fudan University, Shanghai, China
[4]State Key Laboratory of Medical Neurobiology and Ministry of Education Frontiers Center for Brain Science, Fudan University, Shanghai, China
[5]National Health Commission Key Laboratory of Hearing Medicine, Fudan University, Shanghai, China
[6]Institutes of Brain Science and the Collaborative Innovation Center for Brain Science, Fudan University, Shanghai, China
[7]Department of Otorhinolaryngology-Head and Neck Surgery, The Second Affiliated Hospital of Anhui Medical University, Hefei, China
[8]Department of Otolaryngology-Head and Neck Surgery, The First Affiliated Hospital, Medical College, Xiamen University, Xiamen, China
[9]Department of Otolaryngology-Head and Neck Surgery, Shengjing Hospital of China Medical University, Shenyang, China
[10]Department of Otolaryngology, Shanghai Pudong Hospital, Shanghai, China
[11]Department of Otolaryngology, Shenzhen Second People's Hospital, Shenzhen, China
[12]Department of Otolaryngology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[13]Institutes of Biomedical Sciences, Fudan University, Shanghai, China
[*]these authors contributed equally

Corresponding Author:
Huawei Li, MD, PhD
Department of Otorhinolaryngology
Eye & ENT Hospital
Fudan University
Room 611, Building 9
No. 83, Fenyang Road, Xuhui District
Shanghai, 200031
China
Phone: 86 021 64377134 ext 2669
Email: hwli@shmu.edu.cn

Abstract

Background: Questionnaires have been used in the past 2 decades to predict the diagnosis of vertigo and assist clinical decision-making. A questionnaire-based machine learning model is expected to improve the efficiency of diagnosis of vestibular disorders.

Objective: This study aims to develop and validate a questionnaire-based machine learning model that predicts the diagnosis of vertigo.

Methods: In this multicenter prospective study, patients presenting with vertigo entered a consecutive cohort at their first visit to the ENT and vertigo clinics of 7 tertiary referral centers from August 2019 to March 2021, with a follow-up period of 2 months. All participants completed a diagnostic questionnaire after eligibility screening. Patients who received only 1 final diagnosis by their treating specialists for their primary complaint were included in model development and validation. The data of patients enrolled before February 1, 2021 were used for modeling and cross-validation, while patients enrolled afterward entered external validation.

Results: A total of 1693 patients were enrolled, with a response rate of 96.2% (1693/1760). The median age was 51 (IQR 38-61) years, with 991 (58.5%) females; 1041 (61.5%) patients received the final diagnosis during the study period. Among them, 928

(54.8%) patients were included in model development and validation, and 113 (6.7%) patients who enrolled later were used as a test set for external validation. They were classified into 5 diagnostic categories. We compared 9 candidate machine learning methods, and the recalibrated model of light gradient boosting machine achieved the best performance, with an area under the curve of 0.937 (95% CI 0.917-0.962) in cross-validation and 0.954 (95% CI 0.944-0.967) in external validation.

Conclusions: The questionnaire-based light gradient boosting machine was able to predict common vestibular disorders and assist decision-making in ENT and vertigo clinics. Further studies with a larger sample size and the participation of neurologists will help assess the generalization and robustness of this machine learning method.

(J Med Internet Res 2022;24(8):e34126) doi:10.2196/34126

KEYWORDS

vestibular disorders; machine learning; diagnostic model; vertigo; ENT; questionnaire

Introduction

Dizziness and vertigo are the major complaints of patients with vestibular disorders, with an estimated lifetime prevalence of dizziness (including vertigo) of 15%-35% [1]. Dizziness and vertigo are incapacitating and considerably impact patients' quality of life. These conditions often lead to activity restriction and are closely associated with psychiatric disorders such as anxiety, phobic, and somatoform disorders [1-3]. Patients with dizziness and vertigo are also at a higher risk of falls and fall-related injuries, especially older people [4]. However, the diagnosis of vestibular disorders is challenging and time-consuming. It involves a variety of vestibular and neurological causes and complex pathological processes, leading to misdiagnosis and potentially widespread overuse of imaging among vertiginous patients [5-8]. Consequent delays in diagnosis can worsen the functional and psychological consequences of the disease.

The application of artificial intelligence in diagnosing dizziness and vertigo dates back more than 30 years. Expert systems such as *Vertigo* [9], *Carrusel* [10], and *One* [11] consist of knowledge bases with fixed diagnostic rules. They infer through nonadaptive algorithms that were unable to learn from patients' data. Different machine learning algorithms, including genetic algorithms, neural networks, Bayesian methods, k-nearest neighbors, and support vector machines, have also been employed to analyze patient data from *One* [12-16]. The predictive accuracy was 90%-97% for 6 common otoneurologic diagnoses and 76.8%-82.4% for 9 diagnostic categories. *EMBalance* is a comprehensive platform that was launched in 2015 to assist the diagnosis, treatment, and evolution of balance disorders by using ensemble learning methods based on decision trees (Adaptive Boosting) [17,18]. There has been a shift from pure knowledge-driven to data-driven methodology in computer-aided diagnosis of vestibular disorders.

Except *Vertigo*, all of the models mentioned above are based on patients' medical history and examinations combined with necessary tests, while in practice, patient history alone provides important clues to possible diagnosis and further evaluation [19]. Numerous questionnaires for dizziness and vertigo have emerged during the past 2 decades to assist the clinical diagnosis of vestibular disorders [20-27]. Most of these studies used simple statistical models, typically logistic regression, validated with the same data as modeling [26-28]. Few studies have tried to apply machine learning algorithms. However, the accuracy of these models was not as good as that of simple statistical models owing to small data sets or inappropriate choice of modeling data [29,30].

This study is part of the Otogenic Vertigo Artificial Intelligence Research (OVerAIR) study, in which the overarching purpose is to build a comprehensive platform that integrates diagnosis, treatment, rehabilitation, and follow-up in a cohort of patients with otogenic vertigo by using artificial intelligence. The specific aims of this study include developing and verifying a diagnostic platform for vertigo and assisting clinical decision-making by using machine learning techniques and further exploring the effectiveness and clinical utility of the proposed platform.

Methods

Study Design

Patients presenting with a new complaint of vertigo or dizziness according to the classification of vestibular symptoms by the Barany Society [31] were enrolled consecutively from the ENT and vertigo clinics of Eye & ENT Hospital of Fudan University, The Second Hospital of Anhui Medical University, The First Affiliated Hospital of Xiamen University, Shengjing Hospital of China Medical University, Shanghai Pudong Hospital, Shenzhen Second People's Hospital, and The First Affiliated Hospital of Chongqing Medical University from August 2019 through March 2021. At their first interview with an ENT specialist, patients completed the electronic version of the questionnaire via a tablet or smartphone after giving informed consent. Those who were unable to read and complete the questionnaire by themselves answered the questions read by the researchers. We did not interfere with the normal medical procedures of the patients. Patients were scheduled for a next visit as the specialist considered necessary; therefore, they did not stick to a fixed follow-up time.

Ethics Approval

This study was approved by the Institutional Review Boards of all participating centers (approval 2019091). This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guidelines [32].

Outcomes

Each patient went through routine history collection followed by complete otoneurological examinations, and further workup (ie, pure tone audiometry, vestibular testing, computed tomography, and magnetic resonance imaging) was prescribed when necessary. The clinical diagnosis given by ENT specialists with more than 5 years of clinical experience who were blinded to questionnaire responses was used as the reference diagnosis. The reference diagnostic standards include practice guidelines for benign paroxysmal positional vertigo (BPPV) by the American Academy of Otolaryngology-Head and Neck Surgery [33] and diagnostic criteria for vestibular disorders (including vestibular migraine [34], Meniere disease [35], persistent postural-perceptual dizziness [36], vestibular paroxysmia [37], and bilateral vestibulopathy [38]) by the Barany society. Patients with typical clinical features who did not meet the criteria of definite diagnosis were given probable diagnosis. Patients without a specific diagnosis within 2 months or who stopped coming for visits before reaching a final diagnosis were labeled undetermined.

Questionnaire Development

The diagnostic questionnaire was developed through an iterative process that mainly consisted of the following 3 stages.

1. Focus group and panel meeting: First, a focus group discussion and 3 follow-up panel meetings were convened to identify the commonly seen peripheral vestibular disorders in ENT clinics. In this process, 16 disorders were identified and the featured manifestations of each disorder were listed. The literature of diagnostic or practice guidelines for each disorder was searched and the pertinent ones were carefully reviewed. After that, the initial questionnaire composed of 43 items was drafted.
2. Patient interview: Fifteen patients who presented with vertigo in our ENT clinic were interviewed for the understandability and easiness of filling out the questionnaire. Two patients reported that it was too long and time-consuming. Another 3 complained of being asked too many questions such as heart disease and medication taken, which seemed unrelated to their vertigo condition. At this stage, the wording of the questionnaire was thoroughly simplified and 6 questions were deleted.
3. Expert group meeting: At a national conference, 12 experts (from ENT, neurology, vestibular examination, and rehabilitation) were invited to evaluate the suitability and clarity of the questionnaire, and they put forward suggestions for further revision. During this process, the items were reordered and some were combined or omitted.

Statistical Analysis

We compared 9 candidate machine learning methods to screen for the one with the best performance. Five non–ensemble learning algorithms were considered, namely, decision tree [39], ridge regression [40], logistic regression (with L2-regularization) [41], support vector classification [42], and support vector classification with stochastic gradient descent [43]. Ensemble learning refers to a general meta approach that strategically improves predictive performance by combining the predictions from multiple models. Four of the ensemble learning methods were implemented, namely, random forest [44], Adaptive Boosting [45], gradient boosting decision tree [46], and light gradient boosting machine (LGBM) [47]. We took bootstrapped cross-validation that randomly sampled data into train and validation sets by 7:3, which were repeated 100 times with replacement [48]. Models were trained on the training set and evaluated based on the prediction performance on the validation set. The best model was selected and tuned based on the average prediction performance over the 100 validation set. The area under the curve (AUC) was used to evaluate the performance of the models. In multiclass prediction, sensitivity, specificity, likelihood ratio, and AUC were calculated through a one-vs-rest scheme (microaverage). Then, recalibration was performed using calibration curves [49] and Brier scores [50] to adjust the difference between the predicted probability and observed proportion of each diagnostic category. External validation was performed using the data of the newest patients in the cohort (enrolled during the last 2 months), which constituted the test set. The 95% CIs of all the metrics were calculated through bootstrapping.

The missing values of Boolean variables were imputed with False in the main results, and sensitivity analysis was conducted by comparing different imputation strategies (ie, without imputation or imputation with True). All machine learning algorithms were implemented in Python, and the code is available in online resources. Hyperparameters are set to default according to the state-of-art machine learning package: sklearn.

Robustness and Sample Size Analysis

As a data-driven prediction approach for boosting clinical diagnosis, it is necessary to verify that the number of samples is enough for model development and validation. Following Riley [51] and Riley et al [52], we quantified the sufficiency of sample size in terms of the global shrinkage factor and the minimal number of samples. The criterion of enough sample size is to ensure a shrinkage factor >0.9. Further, given the acceptable shrinkage factor (eg, 0.9), the necessary size of the samples to develop a prediction model can be estimated based on the Cox-Snell ratio of explained variance.

Further, the increased flexibility of modern techniques implies that larger sample sizes may be required for reliable estimation compared with classical methods such as logistic regression. Thus, we followed the approach of van der Ploeg et al [53] to evaluate our best model LGBM’s sensitivity on sample size. The training set is of different sizes and subsampled from the development set. Each training set size is repeated 30 times to eliminate randomness, while the average AUC measures the performance on the test set.

Important Variables

To measure the importance of variables, we first evaluated multivariate feature importance according to information gains in cross-validation and selected the top 20 important variables. Then, to figure out feature importance in individual diagnostic categories, each selected variable was used to predict the 5 diagnostic categories independently, and univariate variable importance was measured in terms of AUC.

Results

Overview of the Diagnostic Questionnaire

The final questionnaire consists of 23 items that incorporated branching logic. The full version of the questionnaire is available in Multimedia Appendix 1. The contents of the items are shown in Textbox 1.

Textbox 1. Items in the diagnostic questionnaire.

- One question on the characteristic of the symptom: was the head spinning or not? If not, then the kind of dizziness needs to be specified (heavy/muddled head, staggering, or other)
- Three questions on the frequency, duration, and duration it has been since the first vertigo attack
- One question on the condition of hearing loss, that is, which side and how it changes
- Three questions on the condition of tinnitus, aural fullness, and earache, that is, which side and whether it changes before and after the attack should be specified (aggravate before/during the attack, relieve after the attack)
- One question on the presence of headache, specifically the time of headache attack and relevant family history
- One question on accompanied photophobia or phonophobia
- One question on unsteadiness during, after, or without vertigo attacks
- One question on whether symptoms worsen when standing or walking
- Two questions on the condition of fall, consciousness state, and whether there was incontinence during the attack
- Five questions on the triggering factors of vertigo, that is, lying down, turning over, getting up quickly, holding breath, loud stimulation, in some special scenes, special foods or smells, fatigue, insomnia, and getting angry
- One question on whether it is cervical vertigo, that is, upper limb numbness and pain or neck pain
- One question on prodrome, that is, cold, fever, and diarrhea before onset
- One question on the medical history of otological disorders, that is, otorrhea, otitis media, ear surgery
- One question on head and neck trauma and surgery history

Demographic Characteristics of the Participants

A prospective cohort of 1693 patients was enrolled from the ENT and vertigo clinics of 7 participating centers (Table 1). The response rate was 96.2% (1693/1760, 67 declined participation). Of the 1693 enrolled patients, 1041 (61.5%) received 1 final diagnosis by the treating specialists, 14 (0.8%) had more than one diagnosis, 145 (8.6%) had a probable diagnosis, while the other 493 (29.1%) did not receive the final diagnosis within 2 months. The final diagnoses were found to be unevenly distributed. The most common diagnoses were BPPV, vestibular migraine, sudden sensorineural hearing loss with vestibular dysfunction (SSNHL-V), and Meniere disease. Since only patients with 1 final diagnosis were included in the model development and validation, 1041 patients (median age 50 [IQR 38-61] years, 608 [58.4%] females) in the 5 diagnostic categories were included in the model development and validation. Less frequent diagnoses with no more than 20 cases were labeled as “others” for the moment because there were not sufficient cases for them to form separate categories.

Of the 1041 patients, 928 were classified into the training set (for modeling and cross-validation) and 113 were included in the test set (Table 2). Figure 1 shows the study flowchart. The details of the training set and test set are described in Table 2.

Table 1. Demographic characteristics of the participants (N=1693).

Characteristic	Value
Age (years), median (IQR)	51 (38-61)
Sex, n (%)	
Female	991 (58.5)
Male	702 (41.6)
Diagnoses, n (%)	
Benign paroxysmal positional vertigo	398 (23.5)
Vestibular migraine	203 (12)
Meniere disease	194 (11.5)
Sudden sensorineural hearing loss with vestibular dysfunction	173 (10.2)
Others[a]	73 (4.3)
Multiple diagnosis	14 (0.8)
Probable diagnosis	145 (8.6)
Undetermined	493 (29.1)

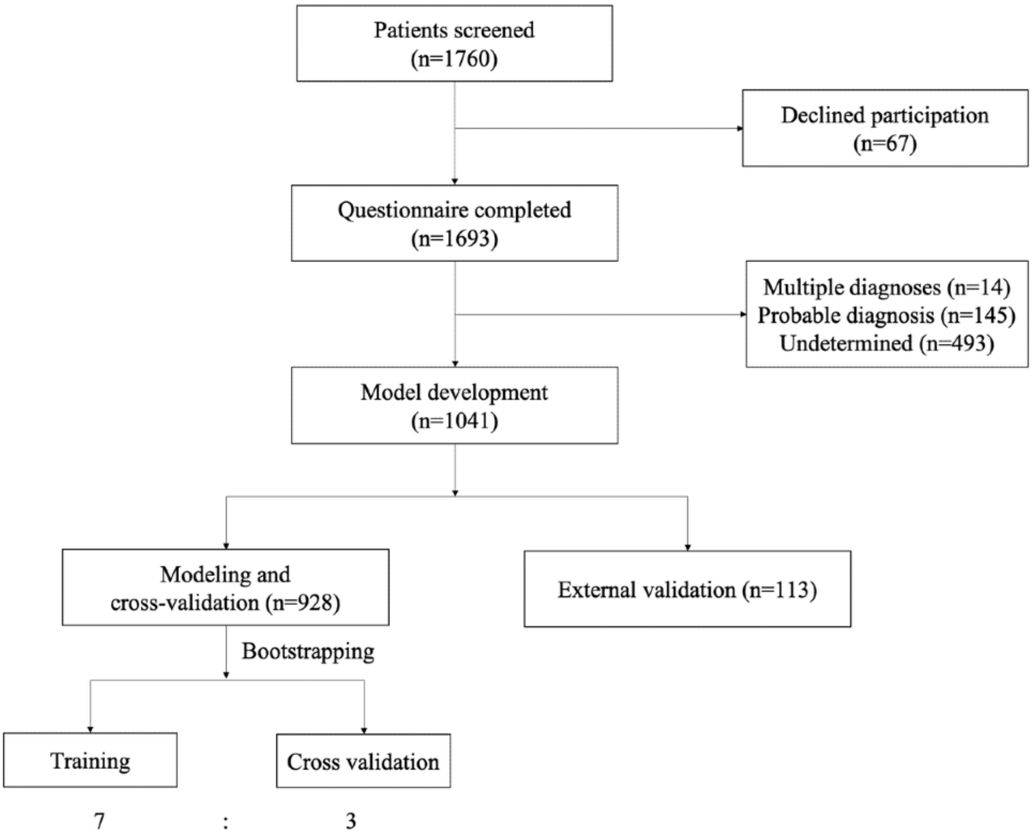
[a] This category included vestibular neuritis, persistent postural-perceptual dizziness, psychogenic dizziness, delayed endolymphatic hydrops, vestibular paroxysmia, cervicogenic vertigo, acoustic neuroma, presbyvestibulopathy, light cupula, Ramsay-Hunt syndrome, labyrinthine fistula, and superior semicircular canal dehiscence syndrome.

Table 2. Characteristics of the training data set and test set.

Characteristic	Training set (n=928)	Test set (n=113)
Age (years), median (IQR)	50 (37-60)	53 (41-63)
Sex, n (%)		
Female	536 (57.8)	72 (63.7)
Male	392 (42.2)	41 (36.3)
Diagnoses, n (%)		
Benign paroxysmal positional vertigo	348 (37.5)	50 (44.2)
Vestibular migraine	182 (19.6)	21 (18.6)
Meniere disease	168 (18.1)	26 (23)
Sudden sensorineural hearing loss with vestibular dysfunction	164 (17.6)	9 (8)
Others[a]	66 (7.1)	7 (6.2)

[a] This category included vestibular neuritis, persistent postural-perceptual dizziness, psychogenic dizziness, delayed endolymphatic hydrops, vestibular paroxysmia, cervicogenic vertigo, acoustic neuroma, presbyvestibulopathy, light cupula, Ramsay-Hunt syndrome, labyrinthine fistula, and superior semicircular canal dehiscence syndrome.

Figure 1. Patients with a new vertigo or dizziness complaint were screened between August 2019 and March 2021. Diagnoses were recorded within 2 months of follow-up.



Development and Validation of the Model

The LGBM model had the highest AUC of 0.937 (95% CI 0.917-0.962) and the lowest Brier score of 0.057 (95% CI 0.049-0.068) among the 9 models in cross-validation (Table 3). Therefore, it was recalibrated and used as the final predictive model.

For sensitivity analysis, when imputing the missing value with mode (the most frequent label), the AUC and Brier score of all 9 methods dropped (Table 4). Note that LGBM does not rely on imputation methods; therefore, it can directly utilize the information from *missing* to achieve a better prediction performance. LGBM without imputation performs as well as the recalibrated LGBM (imputed with 0), which verifies the robustness of our method. Ensemble learning methods performed better than non–ensemble learning methods except logistic regression with LASSO in cross-validation, indicating that the introduction of ensemble learning in vertigo diagnosis is effective across specific ensemble approaches. Further, LGBM performs better than other methods in AUC and Brier scores.

The receiver operating characteristic curves of the recalibrated LGBM model in cross-validation are shown in Figure 2. Table 5 presents the AUC, sensitivity, specificity, likelihood ratios, and accuracy in different diagnostic categories in both cross and external validation. The model made highly accurate prediction for SSNHL-V (AUC>0.98, positive likelihood ratio [+LR]>20, negative likelihood ratio [–LR]<0.05), accurate prediction for BPPV and Meniere disease (AUC>0.95, sensitivity>0.8, specificity>0.9, accuracy>0.9, +LR>10, –LR<0.2), and showed fair discriminative ability for vestibular migraine (AUC 0.9, 95% CI 0.87-0.92). The prediction of other diagnoses was unstable owing to the limited sample size and great heterogeneity in this category, with an AUC ranging from 0.771 to 0.929 in cross-validation and 0.879 to 0.957 in external validation.

Calibration curves in cross-validation (Figure 3) properly estimated the probability of Meniere disease and vestibular migraine and slightly underestimated the probability of SSNHL-V and BPPV. The predictions for other diagnoses were relatively conservative, as it was less likely to give probabilities close to 0 or 1. The Brier score was 0.058 (95% CI 0.049-0.068) in cross-validation, which suggested that the predicted probabilities fitted well with the actual proportions of the diagnoses. We also applied our methods to the external data set. The results indicated that the selected best model, LGBM, was of generalization ability in predicting vertigo diagnosis, achieving an AUC of 0.958 (95% CI 0.951-0.969). Meanwhile, LGBM also performed better than the second-best method, logistic regression, which achieved an AUC of 0.939 (95% CI 0.925-0.956) in external validation. The multivariable feature importance in terms of information gain is shown in Table 6.

The analysis of the global shrinkage factor of each diagnostic category and sensitivity analysis results indicated that the sample size of this study is sufficient for model development. See Multimedia Appendix 2 for more details of sample size analysis. Then, to figure out feature importance in individual diagnostic categories, each of the top 20 contributing variables in Table 6 was used to predict the 5 diagnostic categories independently, and univariate variable importance was measured in terms of AUC (Figure 4).

Table 3. The prediction performance of candidate algorithms.

Method	Area under the curve (95% CI)	Brier score (95% CI)
Non–ensemble learning		
Decision tree	0.765 (0.726-0.798)	0.125 (0.104-0.146)
Ridge regression	0.803 (0.780-0.831)	0.087 (0.071-0.104)
Logistic regression	0.928 (0.907-0.956)	0.060 (0.051-0.069)
Support vector classification	0.501 (0.499-0.505)	0.239 (0.220-0.258)
Stochastic gradient descent	0.733 (0.611-0.824)	0.141 (0.083-0.254)
Ensemble learning		
Random forest	0.924 (0.900-0.949)	0.063 (0.056-0.070)
Adaptive Boosting	0.851 (0.793-0.901)	0.148 (0.144-0.151)
Gradient boosting decision tree	0.925 (0.902-0.951)	0.064 (0.053-0.076)
Light gradient boosting machine	0.935 (0.913-0.960)	0.057 (0.047-0.067)
Recalibrated light gradient boosting machine	0.937 (0.917-0.962)	0.058 (0.049-0.068)

Table 4. Performance of different algorithms while imputing missing data with mode.

Method	Area under the curve (95% CI)	Brier score (95% CI)
Non–ensemble learning		
Decision tree	0.746 (0.690-0.791)	0.137 (0.114-0.169)
Ridge regression	0.788 (0.733-0.817)	0.096 (0.076-0.121)
Logistic regression	0.921 (0.900-0.943)	0.067 (0.057-0.082)
Support vector classification	0.500 (0.500-0.500)	0.240 (0.222-0.258)
Stochastic gradient descent	0.727 (0.578-0.819)	0.148 (0.090-0.251)
Ensemble learning		
Random forest	0.919 (0.896-0.939)	0.068 (0.061-0.078)
Adaptive Boosting	0.833 (0.741-0.887)	0.148 (0.143-0.156)
Gradient boosting decision tree	0.915 (0.888-0.935)	0.073 (0.059-0.093)
Light gradient boosting machine	0.929 (0.906-0.950)	0.062 (0.055-0.072)
Light gradient boosting machine (without imputation)	0.935 (0.916-0.956)	0.057 (0.049-0.065)

Figure 2. The receiver operating characteristic curves (solid lines) with 95% CI (between 2 dashed lines) for each diagnostic category. The performance of each diagnostic category was evaluated through one-vs-rest scheme. BPPV: benign paroxysmal positional vertigo; SSNHL-V: sudden sensorineural hearing loss with vertigo.

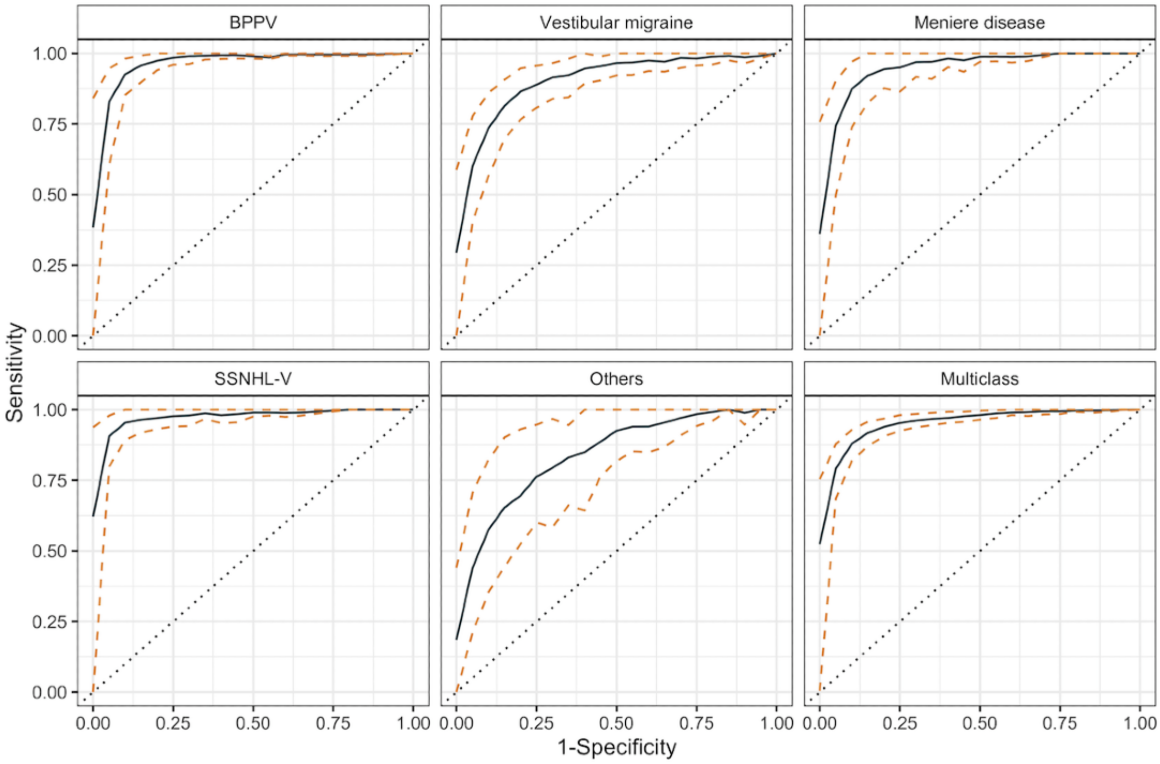


Table 5. Predictive ability in different diagnostic categories.

	AUC[a] (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	+LR[b] (95% CI)	–LR[c] (95% CI)	Accuracy (95% CI)
Benign paroxysmal positional vertigo						
CV[d]	0.97 (0.96-0.99)	0.94 (0.87-0.99)	0.92 (0.85-0.97)	13.23 (6.55-29.3)	0.07 (0.01-0.14)	0.92 (0.89-0.95)
EV[e]	0.98 (0.97-0.99)	0.97 (0.92-1)	0.90 (0.83-0.94)	10.23 (5.88-17.92)	0.04 (0-0.09)	0.93 (0.90-0.96)
Vestibular migraine						
CV	0.91 (0.87-0.94)	0.86 (0.76-0.95)	0.85 (0.74-0.95)	6.58 (3.56-13.93)	0.17 (0.07-0.27)	0.85 (0.78-0.92)
EV	0.9 (0.87-0.92)	0.66 (0.52-0.76)	0.90 (0.85-0.96)	7.38 (4.71-12.05)	0.38 (0.26-0.51)	0.86 (0.82-0.88)
Sudden sensorineural hearing loss with vertigo						
CV	0.99 (0.97-1)	0.95 (0.88-1)	0.95 (0.90-0.99)	25.07 (9.39-67.93)	0.05 (0-0.12)	0.95 (0.91-0.98)
EV	1.00 (1.00-1.00)	1.00 (1.00-1.00)	0.98 (0.97-1.00)	Inf[f] (34.67-Inf)	0.00 (0.00-0.00)	0.98 (0.97-1)
Meniere disease						
CV	0.96 (0.93-0.98)	0.92 (0.81-1)	0.90 (0.82-0.96)	10.79 (5.28-22)	0.09 (0-0.21)	0.90 (0.84-0.95)
EV	0.97 (0.97-0.98)	0.82 (0.69-0.88)	0.98 (0.95-0.99)	Inf (18.4-Inf)	0.19 (0.12-0.31)	0.94 (0.91-0.96)
Others						
CV	0.86 (0.77-0.93)	0.83 (0.66-1)	0.78 (0.55-0.93)	4.44 (2.10-9.77)	0.21 (0-0.44)	0.78 (0.57-0.91)
EV	0.92 (0.88-0.96)	0.74 (0.50-0.86)	0.90 (0.85-0.94)	7.59 (5.05-12.02)	0.38 (0.26-0.51)	0.89 (0.85-0.93)

[a] AUC: area under the curve.

[b] +LR: positive likelihood ratio.

[c] –LR: negative likelihood ratio.

[d] CV: cross-validation.

[e] EV: external validation.

[f] Inf: Positive likelihood ratio was infinity because specificity was 1.

Figure 3. Calibration curves (blue solid lines) with pointwise 95% confidence limits (grey ribbon) on the validation data based on recalibrated light gradient boosting machine model. BPPV: benign paroxysmal positional vertigo; SSNHL-V: sudden sensorineural hearing loss with vertigo.

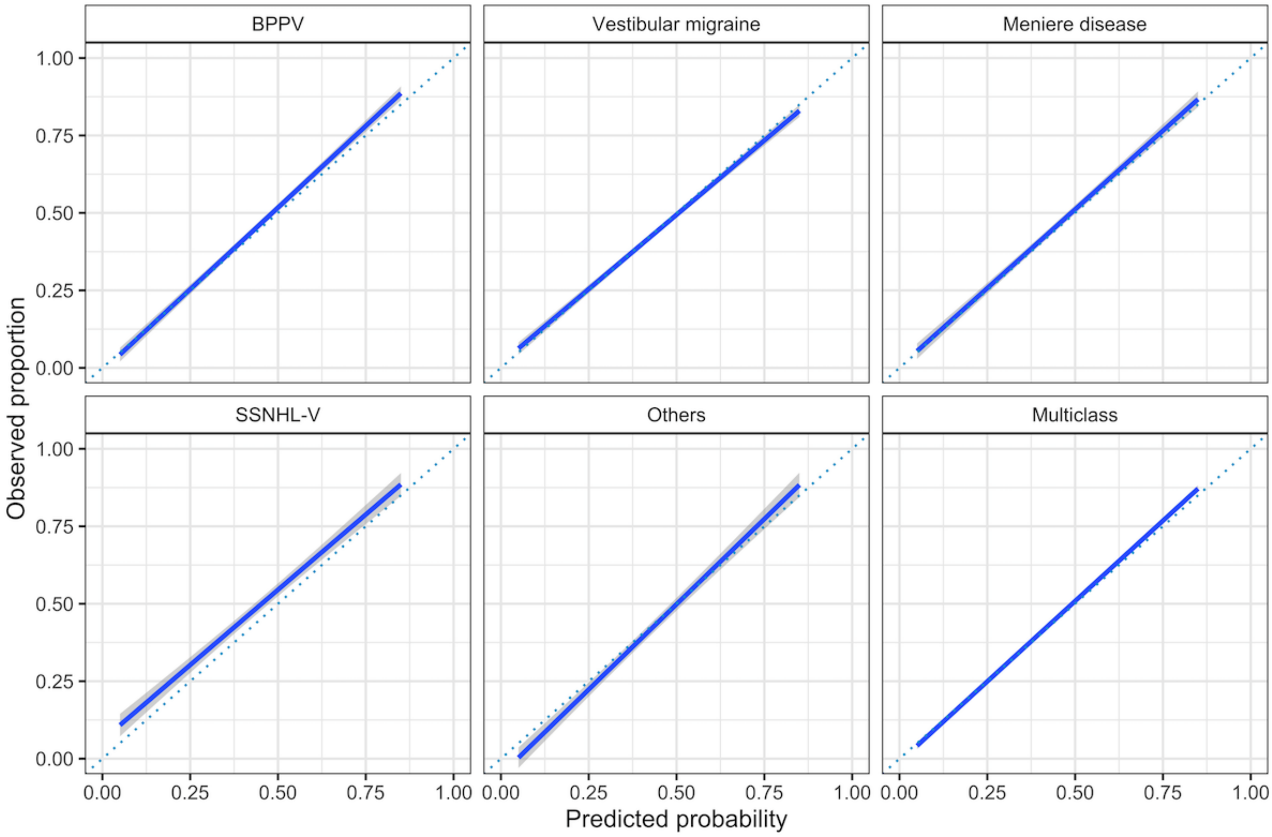
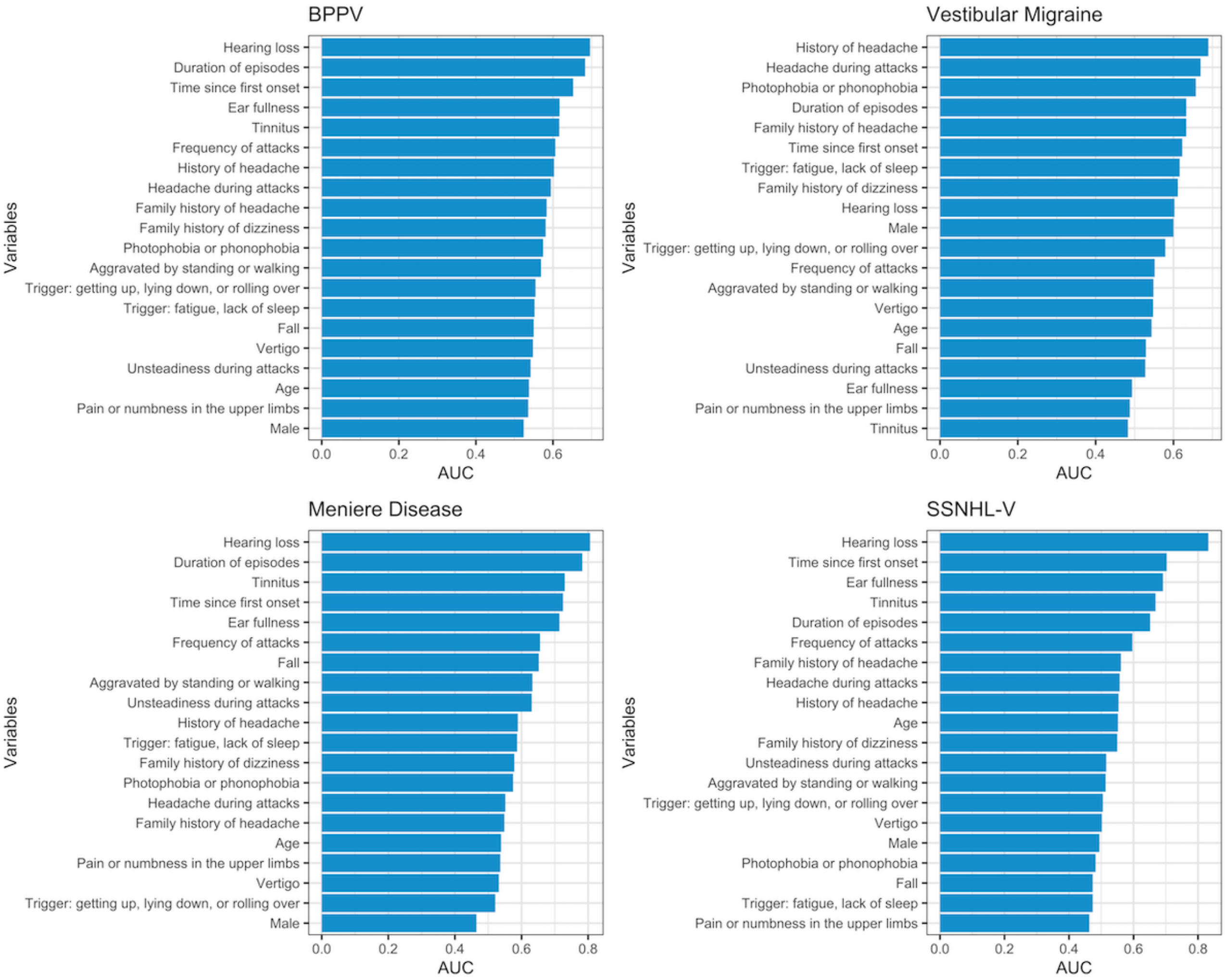


Table 6. Multivariable feature importance in light gradient boosting machine model.

Variable	Feature importance
Sudden hearing loss	1039.8
Duration of episodes	912.3
Hearing loss	694.8
Time since first onset	468.1
Trigger: getting up, lying down, or rolling over	358.0
Age	255.6
History of headache	250.6
Frequency of attacks	221.4
Fluctuating hearing loss	186.3
Photophobia or phonophobia	185.7
Time since first hearing loss	183.7
Recurring symptoms	155.9
Tinnitus	135.5
Ear fullness	135.4
Headache during attacks	117.7
Aggravated by standing or walking	80.4
Trigger: fatigue, lack of sleep	69.7
Vertigo	65.0
Pain or numbness in the upper limbs	62.4
Unsteadiness during attacks	59.5
Family history of headache	54.2
Male	54.1
Fall	47.3
Loss of consciousness, incontinence	44.6
Tinnitus: aggravated before an attack, alleviated after an attack	36.7
Trigger: visual stimuli	31.0
Trigger: sound and pressure	23.0
Unsteadiness: after first onset	22.4
Prodrome: cold, fever, vomiting, or diarrhea	22.0
Family history of dizziness	17.4
Trigger: certain foods	15.9
Otalgia	11.6
Conscious when falling	9.8
History of otitis media or ear surgery	7.2
Tinnitus: worsen during vertigo	4.5
Fluctuating: gradually worsen	0.0
Unsteadiness between attackss	0.0
Recent history of head and neck trauma or surgery	0.0

Figure 4. Area under the curve in univariate prediction was used as the estimate of variable importance. AUC: area under the curve; BPPV: benign paroxysmal positional vertigo; SSNHL-V: sudden sensorineural hearing loss with vertigo.



Discussion

Principal Findings

In this multicenter prospective cohort study, a questionnaire was developed to diagnose vertigo, and an LGBM model was developed using patients' historical data collected through the questionnaire. This is, to our knowledge, the first questionnaire-based machine learning model to predict multiple diagnoses of vertigo. Because all the patients in this study were from ENT and vertigo clinics, the distribution of diagnoses differs from that in previous studies conducted in neurology and balance clinics [19-21,26]. There was a much higher prevalence of SSNHL-V (173/1693, 10.2%) and a lower prevalence of vestibular neuritis (22/1693, 1.3%) in our study.

Our model outperformed previously reported questionnaire-based statistical models in predicting common vestibular diagnoses [20,21,26]. A possible explanation is that machine learning methods are better at dealing with potentially nonlinear relationships and overfitting. Additionally, given the subjectivity of patient-reported historical information, data-driven models are better fits in questionnaire-based prediction than knowledge-driven models [9,11,54,55]. Compared with previous machine learning diagnostic systems that used comprehensive patient history data, physical examination, and laboratory tests, our questionnaire-based diagnostic model has its merits [13-17]. First, medical history provides important clues to the cause of vertigo, based on which the doctor will try to confirm or exclude a presumptive diagnosis. Therefore, a questionnaire-based diagnostic tool can provide early decision support according to patient history and help reduce unnecessary workup. Further, since questionnaire data come directly from patients, the model's performance does not rely on the accurate interpretation of patient history by professionals. Besides, considering the limited accessibility of specific tests (eg, pure tone audiometry, caloric test, video head impulse test), a questionnaire requiring no special equipment is suitable across different clinical settings. However, a questionnaire-based diagnostic model also has intrinsic limitations. Patient-reported medical history can be imprecise because it can be easily affected by recall bias, misinterpretation, emotional state of the patients, and other subjective factors. Meanwhile, for patients with only nonspecific symptoms, physical examination and laboratory testing are more important diagnostic tools. Patient history should always be combined with objective evidence to make a more reliable diagnosis. Therefore, it is necessary to introduce physical examination and laboratory test results into the system in the future to make a comprehensive stepwise diagnostic prediction.

Limitations

This study had the following limitations. The uneven distribution of diagnoses made it difficult for the model to give accurate predictions of rare diagnoses. In order to reduce potential noise, we included only patients with 1 final diagnosis in modeling. The exclusion of patients with undetermined diagnosis was a potential source of bias. There were several reasons that these patients did not receive a specific diagnosis. In some cases, patients with BPPV might experience spontaneous remission while waiting for the scheduled positional test and treatment (1-2 weeks later), which also explains the relatively low prevalence of BPPV in our cohort than that in other ENT clinics [56]. The exclusion of these patients could reduce noise and improve model performance. Besides, some patients only experienced transient symptoms without observable structural, functional, or psychological changes; therefore, no specific diagnosis was given. Moreover, while a majority of patients completed all the necessary examinations within the follow-up, it was also possible that some rare causes were not determined within 2 months, possibly adding to the imbalance of data. Nevertheless, as the cohort expands, more patients with rare diagnoses will be included, which will enable the model to predict rare diagnoses with higher accuracy. We can also manage the influence of imbalanced data during modeling. Meanwhile, the observed AUC in external validation was higher than that in cross-validation, which could be accounted for by the relatively small sample size of the test set. More participants with definite diagnosis are needed for providing further validation. Finally, since this study was conducted in the ENT and vertigo clinic of tertiary centers, the predictive power of the model is yet to be verified in different clinical settings.

Conclusion

This study presents the first questionnaire-based machine learning model for the prediction of common vestibular disorders. The model achieved strong predictive ability for BPPV, vestibular migraine, Meniere disease, and SSNHL-V by using an ensemble learning method LGBM. As part of the OVerAIR platform, it can be used to assist clinical decision-making in ENT clinics and help with the remote diagnosis of BPPV. We have also been working on a smartphone app that integrates the questionnaire with referral, follow-up, treatment, and rehabilitation to improve the health outcomes of patients with vertigo. The next phase of the OVerAIR study will involve the participation of neurologists, which is expected to improve the model's predictive ability for central vertigo and help assess its generalization and robustness.

Acknowledgments

This study was supported by the Capacity Building Project for interdisciplinary diagnosis and treatment of major diseases (otogenic vertigo), Shanghai Municipal Health Commission, Shanghai Municipal Key Clinical Specialty (shslczdzk00801), General Project of Scientific Research Fund of Shanghai Health Committee (grant 202040286), and the National Natural Science Foundation of China (91846302 and 72033003). The funding sources played no part in the design and conduct of this study; collection, management, analysis, and interpretation of data; writing of the report; or the decision to submit the manuscript for publication.

Authors' Contributions

FY and HD had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. FY, PW, HD, CZ, and HL were responsible for the concept and design of this study. All authors contributed to the acquisition, analysis, or interpretation of data and critical revision of the manuscript for important intellectual content. FY, PW, and HD drafted the manuscript. FY and HD performed the statistical analysis. HL and CZ obtained the funding for this study. FY, PW, HD, JW, HY, CZ, and HL provided administrative, technical, or material support. HL and CZ supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The final version of the diagnostic questionnaire for vertigo applied in this study.

[DOCX File , 21 KB - jmir_v24i8e34126_app1.docx]

Multimedia Appendix 2

Sample size analysis.

[DOCX File , 119 KB - jmir_v24i8e34126_app2.docx]

References

1. Eckhardt-Henn A, Best C, Bense S, Breuer P, Diener G, Tschan R, et al. Psychiatric comorbidity in different organic vertigo syndromes. J Neurol 2008 Mar;255(3):420-428. [doi: 10.1007/s00415-008-0697-x] [Medline: 18338198]

2. Best C, Eckhardt-Henn A, Tschan R, Dieterich M. Psychiatric morbidity and comorbidity in different vestibular vertigo syndromes. Results of a prospective longitudinal study over one year. J Neurol 2009 Jan;256(1):58-65. [doi: 10.1007/s00415-009-0038-8] [Medline: 19221849]
3. Lahmann C, Henningsen P, Brandt T, Strupp M, Jahn K, Dieterich M, et al. Psychiatric comorbidity and psychosocial impairment among patients with vertigo and dizziness. J Neurol Neurosurg Psychiatry 2015 Mar;86(3):302-308. [doi: 10.1136/jnnp-2014-307601] [Medline: 24963122]
4. de Moraes SA, Soares WJDS, Ferriolli E, Perracini MR. Prevalence and correlates of dizziness in community-dwelling older people: a cross sectional population based study. BMC Geriatr 2013 Jan 04;13:4 [FREE Full text] [doi: 10.1186/1471-2318-13-4] [Medline: 23290128]
5. Grill E, Strupp M, Müller M, Jahn K. Health services utilization of patients with vertigo in primary care: a retrospective cohort study. J Neurol 2014 Aug;261(8):1492-1498. [doi: 10.1007/s00415-014-7367-y] [Medline: 24817192]
6. Ahsan SF, Syamal MN, Yaremchuk K, Peterson E, Seidman M. The costs and utility of imaging in evaluating dizzy patients in the emergency room. Laryngoscope 2013 Sep;123(9):2250-2253. [doi: 10.1002/lary.23798] [Medline: 23821602]
7. Bakhit M, Heidarian A, Ehsani S, Delphi M, Latifi SM. Clinical assessment of dizzy patients: the necessity and role of diagnostic tests. Glob J Health Sci 2014 Mar 24;6(3):194-199 [FREE Full text] [doi: 10.5539/gjhs.v6n3p194] [Medline: 24762362]
8. Royl G, Ploner CJ, Leithner C. Dizziness in the emergency room: diagnoses and misdiagnoses. Eur Neurol 2011;66(5):256-263. [doi: 10.1159/000331046] [Medline: 21986277]
9. Mira E, Buizza A, Magenes G, Manfrin M, Schmid R. Expert systems as a diagnostic aid in otoneurology. ORL J Otorhinolaryngol Relat Spec 1990;52(2):96-103. [doi: 10.1159/000276113] [Medline: 2183129]
10. Gavilán C, Gallego J, Gavilán J. 'Carrusel': an expert system for vestibular diagnosis. Acta Otolaryngol 1990;110(3-4):161-167. [doi: 10.3109/00016489009122532] [Medline: 2239202]
11. Auramo Y, Juhola M, Pyykkö I. An expert system for the computer-aided diagnosis of dizziness and vertigo. Med Inform (Lond) 1993;18(4):293-305. [doi: 10.3109/14639239309025318] [Medline: 8072338]
12. Laurikkala JPS, Kentala EL, Juhola M, Pyvkkö IV. A novel machine learning program applied to discover otological diagnoses. Scand Audiol Suppl 2001(52):100-102. [doi: 10.1080/010503901300007218] [Medline: 11318434]
13. Siermala M, Juhola M, Kentala E. Neural network classification of otoneurological data and its visualization. Comput Biol Med 2008 Aug;38(8):858-866. [doi: 10.1016/j.compbiomed.2008.05.002] [Medline: 18572156]
14. Miettinen K, Juhola M. Classification of otoneurological cases according to Bayesian probabilistic models. J Med Syst 2010 Apr;34(2):119-130. [doi: 10.1007/s10916-008-9223-z] [Medline: 20433050]
15. Varpa K, Joutsijoki H, Iltanen K, Juhola M. Applying one-vs-one and one-vs-all classifiers in k-nearest neighbour method and support vector machines to an otoneurological multi-class problem. Stud Health Technol Inform 2011;169:579-583. [Medline: 21893815]
16. Joutsijoki H, Varpa K, Iltanen K, Juhola M. Machine learning approach to an otoneurological classification problem. 2013 Presented at: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 3-7, 2013; Osaka, Japan p. 1294-1297.
17. Exarchos T, Rigas G, Bibas A, Kikidis D, Nikitas C, Wuyts F, et al. Mining balance disorders' data for the development of diagnostic decision support systems. Comput Biol Med 2016 Oct 01;77:240-248. [doi: 10.1016/j.compbiomed.2016.08.016] [Medline: 27619194]
18. Exarchos T, Bellos C, Bakola I, Kikidis D, Bibas A, Koutsouris D, et al. Management and modeling of balance disorders using decision support systems: the EMBALANCE project. In: GeNeDis 2014: Adv Exp Med Biol. Cham: Springer; 2015:61-67.
19. Zhao JG, Piccirillo JF, Spitznagel EL, Kallogjeri D, Goebel JA. Predictive capability of historical data for diagnosis of dizziness. Otol Neurotol 2011 Feb;32(2):284-290 [FREE Full text] [doi: 10.1097/MAO.0b013e318204aad3] [Medline: 21178804]
20. Bayer O, Warninghoff J, Straube A. Diagnostic indices for vertiginous diseases. BMC Neurol 2010 Oct 25;10:98 [FREE Full text] [doi: 10.1186/1471-2377-10-98] [Medline: 20973968]
21. Friedland DR, Tarima S, Erbe C, Miles A. Development of a Statistical Model for the Prediction of Common Vestibular Diagnoses. JAMA Otolaryngol Head Neck Surg 2016 Apr;142(4):351-356 [FREE Full text] [doi: 10.1001/jamaoto.2015.3663] [Medline: 26913615]
22. Imai T, Higashi-Shingai K, Takimoto Y, Masumura C, Hattori K, Inohara H. New scoring system of an interview for the diagnosis of benign paroxysmal positional vertigo. Acta Otolaryngol 2016;136(3):283-288. [doi: 10.3109/00016489.2015.1121547] [Medline: 26838579]
23. Lapenna R, Faralli M, Del Zompo MR, Cipriani L, Mobaraki PD, Ricci G. Reliability of an anamnestic questionnaire for the diagnosis of benign paroxysmal positional vertigo in the elderly. Aging Clin Exp Res 2016 Oct;28(5):881-888. [doi: 10.1007/s40520-015-0487-0] [Medline: 26537236]
24. Chen W, Shu L, Wang Q, Pan H, Wu J, Fang J, et al. Validation of 5-item and 2-item questionnaires in Chinese version of Dizziness Handicap Inventory for screening objective benign paroxysmal positional vertigo. Neurol Sci 2016 Aug;37(8):1241-1246. [doi: 10.1007/s10072-016-2573-2] [Medline: 27071688]

25. Li L, Qi X, Liu J, Wang Z. Formulation and evaluation of diagnostic questionnaire for benign paroxysmal positional vertigo. Journal of the Neurological Sciences 2017 Oct;381:148-149. [doi: 10.1016/j.jns.2017.08.438]
26. Roland L, Kallogjeri D, Sinks B, Rauch S, Shepard N, White J, et al. Utility of an Abbreviated Dizziness Questionnaire to Differentiate Between Causes of Vertigo and Guide Appropriate Referral: A Multicenter Prospective Blinded Study. Otol Neurotol 2015 Dec;36(10):1687-1694 [FREE Full text] [doi: 10.1097/MAO.0000000000000884] [Medline: 26485598]
27. Kim H, Song J, Zhong L, Yang X, Kim J. Questionnaire-based diagnosis of benign paroxysmal positional vertigo. Neurology 2019 Dec 30;94(9):e942-e949. [doi: 10.1212/wnl.0000000000008876]
28. Britt CJ, Ward BK, Owusu Y, Friedland D, Russell JO, Weinreich HM. Assessment of a Statistical Algorithm for the Prediction of Benign Paroxysmal Positional Vertigo. JAMA Otolaryngol Head Neck Surg 2018 Oct 01;144(10):883-886 [FREE Full text] [doi: 10.1001/jamaoto.2018.1657] [Medline: 30178063]
29. Richburg H, Povinelli R, Friedland D. Direct-to-patient survey for diagnosis of benign paroxysmal positional vertigo. 2018 Dec Presented at: 17th IEEE International Conference on Machine Learning and Applications; December; Orlando, FL, USA p. 332-337.
30. Masankaran L, Viyanon W, Mahasittiwat V. Classification of benign paroxysmal positioning vertigo types from dizziness handicap inventory using machine learning techniques. 2018 Presented at: International Conference on Intelligent Informatics and Biomedical Sciences; 2018; Bangkok, Thailand p. 209-214.
31. Bisdorff A, Von Brevern M, Lempert T, Newman-Toker DE. Classification of vestibular symptoms: Towards an international classification of vestibular disorders. VES 2009 Oct 01;19(1-2):1-13. [doi: 10.3233/ves-2009-0343]
32. Collins GS, Reitsma JB, Altman DG, Moons KGM, TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. Circulation 2015 Jan 13;131(2):211-219 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.014508] [Medline: 25561516]
33. Bhattacharyya N, Gubbels SP, Schwartz SR, Edlow JA, El-Kashlan H, Fife T, et al. Clinical Practice Guideline: Benign Paroxysmal Positional Vertigo (Update). Otolaryngol Head Neck Surg 2017 Mar;156(3_suppl):S1-S47. [doi: 10.1177/0194599816689667] [Medline: 28248609]
34. Lempert T, Olesen J, Furman J, Waterston J, Seemungal B, Carey J, et al. Vestibular migraine: Diagnostic criteria. VES 2012 Nov 01;22(4):167-172. [doi: 10.3233/ves-2012-0453]
35. Lopez-Escamez JA, Carey J, Chung W, Goebel JA, Magnusson M, Mandalà M, et al. Diagnostic criteria for Menière's disease. VES 2015 Mar 01;25(1):1-7. [doi: 10.3233/ves-150549]
36. Staab JP, Eckhardt-Henn A, Horii A, Jacob R, Strupp M, Brandt T, et al. Diagnostic criteria for persistent postural-perceptual dizziness (PPPD): Consensus document of the committee for the Classification of Vestibular Disorders of the Bárány Society. VES 2017 Oct 21;27(4):191-208. [doi: 10.3233/ves-170622]
37. Strupp M, Lopez-Escamez JA, Kim J, Straumann D, Jen JC, Carey J, et al. Vestibular paroxysmia: Diagnostic criteria. VES 2017 Jan 27;26(5-6):409-415. [doi: 10.3233/ves-160589]
38. Strupp M, Kim J, Murofushi T, Straumann D, Jen JC, Rosengren SM, et al. Bilateral vestibulopathy: Diagnostic criteria Consensus document of the Classification Committee of the Bárány Society. VES 2017 Oct 21;27(4):177-189. [doi: 10.3233/ves-170619]
39. Loh W. Classification and regression trees. WIREs Data Mining Knowl Discov 2011 Jan 06;1(1):14-23. [doi: 10.1002/widm.8]
40. Rifkin RM, Lippert RA. Notes on regularized least squares. CSAIL Technical Reports. 2007 May 01. URL: http://hdl.handle.net/1721.1/37318 [accessed 2022-06-25]
41. Hosmer JDW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Hoboken, New Jersey: John Wiley & Sons; 2013.
42. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. Data Science ASSN. 2003. URL: http://www.datascienceassn.org/sites/default/files/Practical%20Guide%20to%20Support%20Vector%20Classification.pdf [accessed 2021-02-19]
43. Bottou L. Stochastic gradient descent tricks. In: Montavon G, Orr G, Müller KR, editors. Neural Networks: Tricks of the Trade. Berlin, Heidelberg: Springer; 2012:3-540.
44. Breiman L. Random forests. Machine learning 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
45. Hastie T, Rosset S, Zhu J, Zou H. Multi-class adaboost. Statistics and its Interface 2009;2(3):349-360. [doi: 10.4310/SII.2009.v2.n3.a8]
46. Friedman JH. Stochastic gradient boosting. Computational Statistics & Data Analysis 2002 Feb;38(4):367-378. [doi: 10.1016/S0167-9473(01)00065-2]
47. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. 2017 Dec Presented at: Advances in Neural Information Processing Systems 30 (NIPS 2017); December 4-9, 2017; Long Beach, CA, USA.
48. Rao C, Wu Y. Linear model selection by cross-validation. Journal of Statistical Planning and Inference 2005 Jan;128(1):231-240. [doi: 10.1016/j.jspi.2003.10.004]
49. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016 Jun;74:167-176. [doi: 10.1016/j.jclinepi.2015.12.005] [Medline: 26772608]

50. Harrell JFE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, USA: Springer; 2015.
51. Riley RD. Correction to: Minimum sample size for developing a multivariable prediction model: Part II-binary and time-to-event outcomes by Riley RD, Snell KI, Ensor J, et al. Stat Med 2019 Dec 30;38(30):5672 [FREE Full text] [doi: 10.1002/sim.8409] [Medline: 31793031]
52. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. Stat Med 2019 Mar 30;38(7):1262-1275. [doi: 10.1002/sim.7993] [Medline: 30347470]
53. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014 Dec 22;14:137 [FREE Full text] [doi: 10.1186/1471-2288-14-137] [Medline: 25532820]
54. Dong C, Wang Y, Zhang Q, Wang N. The methodology of Dynamic Uncertain Causality Graph for intelligent diagnosis of vertigo. Comput Methods Programs Biomed 2014;113(1):162-174. [doi: 10.1016/j.cmpb.2013.10.002] [Medline: 24176413]
55. Feil K, Feuerecker R, Goldschagg N, Strobl R, Brandt T, von Müller A, et al. Predictive Capability of an iPad-Based Medical Device (mex) for the Diagnosis of Vertigo and Dizziness. Front Neurol 2018;9:29 [FREE Full text] [doi: 10.3389/fneur.2018.00029] [Medline: 29535671]
56. Parker I, Hartel G, Paratz J, Choy N, Rahmann A. A Systematic Review of the Reported Proportions of Diagnoses for Dizziness and Vertigo. Otol Neurotol 2019 Jan;40(1):6-15. [doi: 10.1097/MAO.0000000000002044] [Medline: 30439765]

Abbreviations

AUC: area under the curve
BPPV: benign paroxysmal positional vertigo
LGBM: light gradient boosting machine
OVerAIR: Otogenic Vertigo Artificial Intelligence Research
SSNHL-V: sudden sensorineural hearing loss with vestibular dysfunction

Edited by R Kukafka; submitted 12.10.21; peer-reviewed by R Bajpai, S Kim; comments to author 30.01.22; revised version received 14.02.22; accepted 13.06.22; published 03.08.22.

Please cite as:
Yu F, Wu P, Deng H, Wu J, Sun S, Yu H, Yang J, Luo X, He J, Ma X, Wen J, Qiu D, Nie G, Liu R, Hu G, Chen T, Zhang C, Li H
A Questionnaire-Based Ensemble Learning Model to Predict the Diagnosis of Vertigo: Model Development and Validation Study
J Med Internet Res 2022;24(8):e34126
URL: https://www.jmir.org/2022/8/e34126
doi:10.2196/34126
PMID:35921135

Original Paper

Predicting Mortality in Intensive Care Unit Patients With Heart Failure Using an Interpretable Machine Learning Model: Retrospective Cohort Study

Jili Li[1*], BS; Siru Liu[2*], PhD; Yundi Hu[3], BDS; Lingfeng Zhu[4], BCS; Yujia Mao[1], BS; Jialin Liu[5], MD

[1]West China School of Medicine, Sichuan University, Chengdu, China
[2]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
[3]School of Data Science, Fudan University, Shanghai, China
[4]Department of Computer Science, Sichuan University, Chengdu, China
[5]Department of Medical Informatics, West China Hospital, Sichuan University, Chengdu, China
[*]these authors contributed equally

Corresponding Author:
Jialin Liu, MD
Department of Medical Informatics
West China Hospital
Sichuan University
No 37 Guoxue Road
Chengdu, 610041
China
Phone: 86 28 85422306
Fax: 86 28 85582944
Email: dljl8@163.com

Abstract

Background: Heart failure (HF) is a common disease and a major public health problem. HF mortality prediction is critical for developing individualized prevention and treatment plans. However, due to their lack of interpretability, most HF mortality prediction models have not yet reached clinical practice.

Objective: We aimed to develop an interpretable model to predict the mortality risk for patients with HF in intensive care units (ICUs) and used the SHapley Additive exPlanation (SHAP) method to explain the extreme gradient boosting (XGBoost) model and explore prognostic factors for HF.

Methods: In this retrospective cohort study, we achieved model development and performance comparison on the eICU Collaborative Research Database (eICU-CRD). We extracted data during the first 24 hours of each ICU admission, and the data set was randomly divided, with 70% used for model training and 30% used for model validation. The prediction performance of the XGBoost model was compared with three other machine learning models by the area under the curve. We used the SHAP method to explain the XGBoost model.

Results: A total of 2798 eligible patients with HF were included in the final cohort for this study. The observed in-hospital mortality of patients with HF was 9.97%. Comparatively, the XGBoost model had the highest predictive performance among four models with an area under the curve (AUC) of 0.824 (95% CI 0.7766-0.8708), whereas support vector machine had the poorest generalization ability (AUC=0.701, 95% CI 0.6433-0.7582). The decision curve showed that the net benefit of the XGBoost model surpassed those of other machine learning models at 10%~28% threshold probabilities. The SHAP method reveals the top 20 predictors of HF according to the importance ranking, and the average of the blood urea nitrogen was recognized as the most important predictor variable.

Conclusions: The interpretable predictive model helps physicians more accurately predict the mortality risk in ICU patients with HF, and therefore, provides better treatment plans and optimal resource allocation for their patients. In addition, the interpretable framework can increase the transparency of the model and facilitate understanding the reliability of the predictive model for the physicians.

(J Med Internet Res 2022;24(8):e38082) doi:10.2196/38082

KEYWORDS

heart failure; mortality; intensive care unit; prediction; XGBoost; SHAP; SHapley Additive exPlanation

Introduction

Heart failure (HF), the terminal phase of many cardiovascular disorders, is a major health care issue with an approximate prevalence of 26 million worldwide and more than 1 million hospital admissions annually in both the United States and Europe [1]. Projections show that by 2030 over 8 million Americans will have HF, leading to an increase of 46% from 2012 [2]. Each year, HF costs an estimated US $108 billion, constituting 2% of the health care budget globally, and it is predicted to continue to rise, yet half of it is possibly avoidable [3]. As COVID-19 continues to spread across the world, HF, a severe complication, is associated with poor outcome and death from COVID-19 [4,5].

The critically ill patients in intensive care units (ICUs) demand intensive care services and highly qualified multidisciplinary assistance [6]. Although ICU plays an integral role in maintaining patients' life, this also implies the workforce shortage, limited medical resources, and heavy economic burden [7]. Therefore, early hospital mortality detection for patients is necessary and may assist in delivering proper care and providing clinical decision support [8].

In recent years, artificial intelligence has been widely used to explore the early warning predictors of many diseases. Given the inherent powerful feature of capturing the nonlinear relationships with machine learning algorithms, more researchers advocate the use of new prediction models based on machine learning to support appropriate treatment for patients rather than traditional illness severity classification systems such as SOFA, APACHE II, or SAPS II [9-11]. Although a large number of predictive models have shown promising performance in research, the evidence for their application in clinical setting and interpretable risk prediction models to aid disease prognosis are still limited [12-15].

The purpose of our study is to develop an interpretable model to predict the risk mortality for patients with HF in the ICU, using the free and open critical care database—the eICU Collaborative Research Database (eICU-CRD). In addition, the SHapley Additive exPlanations (SHAP) method is used to explain the extreme gradient boosting (ie, XGBoost) model and explore prognostic factors for HF.

Methods

Data Source

The eICU-CRD (version 2.0) is a publicly available multicenter database [16], containing deidentified data associated with over 200,000 admissions to ICUs at 208 hospitals of the United States between 2014-2015. It records all patients, demographics, vital sign measurements, diagnosis information, and treatment information in detail [17].

Ethical Considerations

Ethical approval and individual patient consent was not necessary because all the protected health information was anonymized.

Study Population

All patients in the eICU-CRD database were considered. The inclusion criteria were as follows: (1) patients were diagnosed with HF according to the International Classification of Diseases, ninth and tenth Revision codes (Multimedia Appendix 1); (2) the diagnosis priority label was "primary" when admitted to the ICU in 24 hours; (3) the ICU stay was more than 1 day; and (4) patients were aged 18 years or older. Patients who had more than 30% missing values were excluded [18].

Predictor Variables

The prediction outcome of the study is the probability of in-hospital mortality, defined as patient's condition upon leaving the hospital. Based on previous studies [19-22] and experts' opinion (a total of 6 independent medical professionals and cardiologists in West China Hospital of Sichuan University), demographics, comorbidities, vital signs, and laboratory findings (Multimedia Appendix 2) were extracted from the eICU-CRD, using Structured Query Language (MySQL) queries (version 5.7.33; Oracle Corporation). The following tables from eICU-CRD were used: "diagnosis," "intakeoutput," "lab," "patient," and "nursecharting." Except for the demographic characteristics, other variables were collected during the first 24 hours of each ICU admission. Furthermore, to avoid overfitting, Least Absolute Shrinkage and Selection Operator (LASSO) is used to select and filter the variables [23,24].

Missing Data Handling

Variables with missing data are a common occurrence in eICU-CRD. However, analyses that ignore missing data have the potential to produce biased results. Therefore, we used multiple imputation for missing data [25]. All selected variables contained <30% missing values. Data were assumed missing at random and were imputed using fully conditional specification with the "mice" package (version 3.13.0) for R (version 4.1.0; R Core Team).

Machine Learning Explainable Tool

The interpretation of the prediction model is performed by SHAP, which is a unified approach to calculate the contribution and influence of each feature toward the final predictions precisely [26]. The SHAP values can show how much each predictor contributes, either positively or negatively, to the target variable. Besides, each observation in the data set could be interpreted by the particular set of SHAP values.

Statistical Analysis

All statistical analysis and calculations were performed using R software and Python (version 3.8.0; Python Software Foundation). The categorical variables are expressed as total numbers and percentages, and the χ^2 test or Fisher exact test (expected frequency <10) is used to compare the differences between groups. The continuous variables are expressed as median and IQR, and the Wilcoxon rank sum test is used when comparing the two groups.

Four machine learning models—XGBoost, logistic regression (LR), random forest (RF), and support vector machine (SVM)—were used to develop the predictive models. The prediction performance of each model was evaluated by the area under the receiver operating characteristic curve. Moreover, we calculated the accuracy, sensitivity, positive predictive values, negative predictive values, and F_1 score when the prediction specificity was fixed at 85%. Additionally, to assess the utility of models for decision-making by quantifying the net benefit at different threshold probabilities, decision curve analysis (DCA) was conducted [27].

Results

Patient Characteristics

Among 17,029 patients with HF in eICU-CRD, a total of 2798 adult patients diagnosed with primary HF were included in the final cohort for this study. The patient screening process is shown in Figure 1. The data set was randomly divided into 2 parts: 70% (n=1958) of the data were used for model training, and 30% (n=840) of the data were used for model validation. The LASSO regularization process resulted in 24 potential predictors on the basis of 1958 patients in the training data set, which were used for model developing. Patients in the nonsurvivor group were older than the ones in the survivor group (*P*<.001). The hospital mortality rate was 9.96% (195/1958) in the training data set and 10% (84/840) in the testing data set (Multimedia Appendix 3). Table 1 shows the comparisons of predictor variables between survivors and nonsurvivors during hospitalization.

Figure 1. Flowchart of patient selection. ICD: International Classification of Diseases; ICU: intensive care unit.

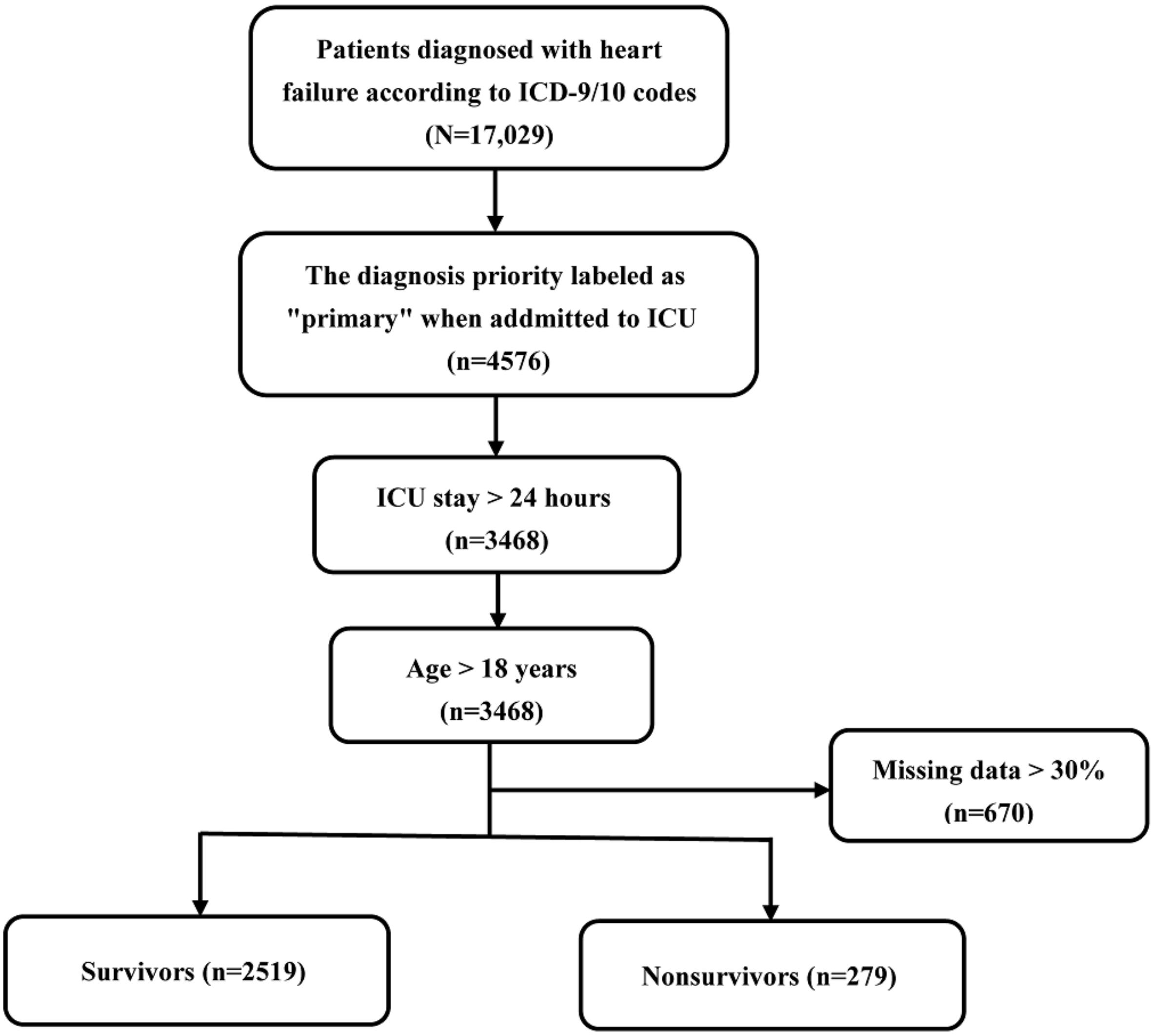


Table 1. All predictor variables for patients with heart failure (N=2798).

	Survivors (n=2519)	Nonsurvivors (n=279)	*P* value
Age (years), median (IQR)	71 (60-80)	76 (66-82)	<.001
Gender (male), n (%)	1338 (53.1)	170 (60.9)	.02
Comorbidities, n (%)			
Hypertension	654 (26)	46 (16.5)	<.001
Acute renal failure	441 (17.5)	78 (28.0)	<.001
Vital signs, median (IQR)			
Heartrate_min[a]	70 (61-80)	74 (62-86)	<.001
Respiratory rate_avg[b]	20.1 (17.8-23.0)	21.8 (19.0-26.0)	<.001
Respiratory rate_max[c]	27 (24-32)	32 (26-38)	<.001
Nibp[d]_systolic_avg	120.0 (107.1-134.8)	109.0 (100.1-121.4)	<.001
Nibp_systolic_min	95 (83-110)	84 (72-97)	<.001
Nibp_diastolic_min	49 (41-57)	45 (35-52.5)	<.001
Temperature_max	37 (37-37)	37 (37-38)	.03
Temperature_min	36 (36-37)	36 (36-37)	.007
Laboratory variables, median (IQR)			
Urineoutput	1550 (599-2750)	875 (140-1900)	<.001
SpO_2[e]_min	92 (88-95)	90 (84.5-94)	<.001
SpO_2_avg	96.6 (95.1-98.0)	96.5 (94.5-97.9)	.04
Anion_gap_max	11.0 (9.0-14.0)	12.0 (10.0-15.0)	<.001
Creatinine_min	1.45 (1.01-2.30)	1.70 (1.19-2.50)	.001
Blood_urea_nitrogen_avg	30.0 (21.0-47.6)	42.0 (28.0-58.5)	<.001
Calcium_min	8.6 (8.1-9.0)	8.5 (7.9-8.9)	.005
Chloride_min	101 (97-104)	99 (95-104)	.01
Platelets×1000_min	193 (149-249)	180 (140-235.5)	.008
White_blood_cell×1000_min	9.1 (6.8-12.1)	10.9 (7.6-15.7)	<.001
RDW[f]_min	15.7 (14.4-17.3)	16.4 (15.0-18.2)	<.001
Hemoglobin_max	10.6 (9.2-12.3)	10.4 (8.95-12.0)	.059

[a] Min: minimum.
[b] Avg: average.
[c] Max: maximum.
[d] Nibp: noninvasive blood pressure.
[e] SpO_2: O_2 saturation.
[f] RDW: red blood cell distribution width.

Model Building and Evaluation

Within the training data set, the XGBoost, LR, RF, and SVM models were established, and the testing data set obtained AUCs of 0.824, 0.800, 0.779, and 0.701, respectively (Table 2 and Figure 2). Comparatively, XGBoost had the highest predictive performance among the four models (AUC=0.824, 95% CI 0.7766-0.8708), whereas SVM had the poorest generalization ability (AUC=0.701, 95% CI 0.6433-0.7582). DCA was performed for four machine learning models in the testing data set to compare the net benefit of the best model and alternative approaches for clinical decision-making. Clinical net benefit is defined as the minimum probability of disease, when further intervention would be warranted [28]. The plot measures the net benefit at different threshold probabilities. The orange line in Figure 3 represents the assumption that all patients received intervention, whereas the yellow line represents that none of the patients received intervention. Due to the heterogeneous profile of the study population, treatment strategies informed by any of the four machine learning–based models are superior to the default strategies of treating all or no patient. The net benefit of the XGBoost model surpassed those of the other machine learning models at 10%~28% threshold probabilities (Figure 3).

Table 2. Performance of each model for prediction.

Model	AUC[a] (%)	Sensitivity (%)	F_1 score	Accuracy (%)	PPV[b]	NPV[c]
XGBoost	0.824	0.595	0.407	0.826	0.307	0.950
LR[d]	0.800	0.607	0.413	0.827	0.311	0.951
RF[e]	0.779	0.571	0.392	0.823	0.298	0.947
SVM[f]	0.701	0.345	0.258	0.801	0.204	0.921

[a] AUC: area under the curve.

[b] PPV: positive predictive value.

[c] NPV: negative predictive value.

[d] LR: logistic regression.

[e] RF: random forest.

[f] SVM: support vector machine.

Figure 2. The receiver operating characteristic curve among the four models for patients with heart failure. SVM: support vector machine.

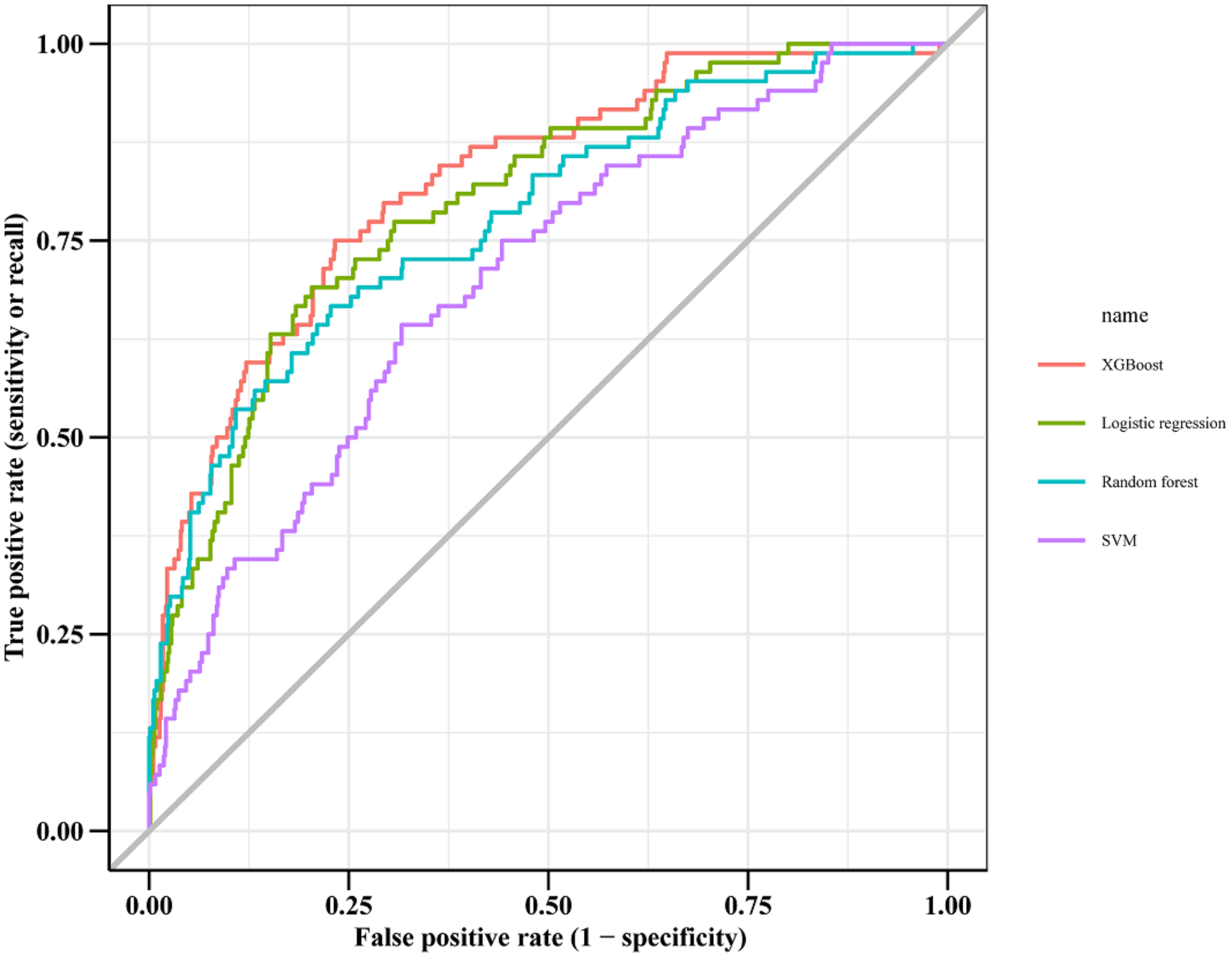
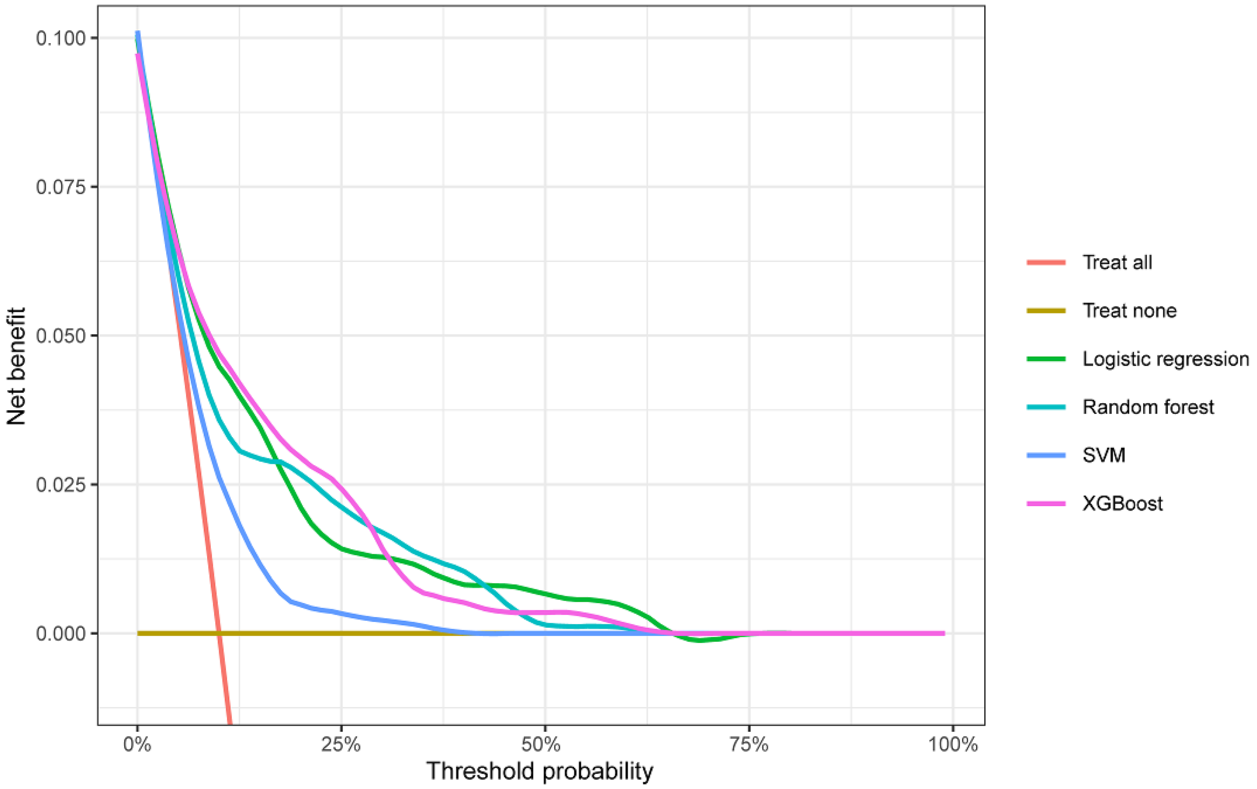


Figure 3. Decision curve analysis of four models plotting the net benefit at different threshold probabilities. SVM: support vector machine.



Explanation of XGBoost Model With the SHAP Method

The SHAP algorithm was used to obtain the importance of each predictor variable to the outcome predicted by the XGBoost model. The variable importance plot lists the most significant variables in a descending order (Figure 4). The average of blood urea nitrogen (BUN) had the strongest predictive value for all prediction horizons, followed quite closely by the age factor, the average of noninvasive systolic blood pressure, urine output, and the maximum of respiratory rate. Furthermore, to detect the positive and negative relationships of the predictors with the target result, SHAP values were applied to uncover the mortality risk factors. As presented in Figure 5, the horizontal location shows whether the effect of that value is associated with a higher or lower prediction and the color shows whether that variable is high (in red) or low (in blue) for that observation; we can see that increases in the average BUN has a positive impact and push the prediction toward mortality, whereas increases in urine output has a negative impact and push the prediction toward survival.

Figure 4. The weights of variables importance. Avg: average; BUN: blood urea nitrogen; max: maximum; min: minimum; NIBP: noninvasive blood pressure; RDW: red blood cell distribution width; SHAP: SHapley Additive exPlanation; SpO_2: O_2 saturation; WBC: white blood cell.

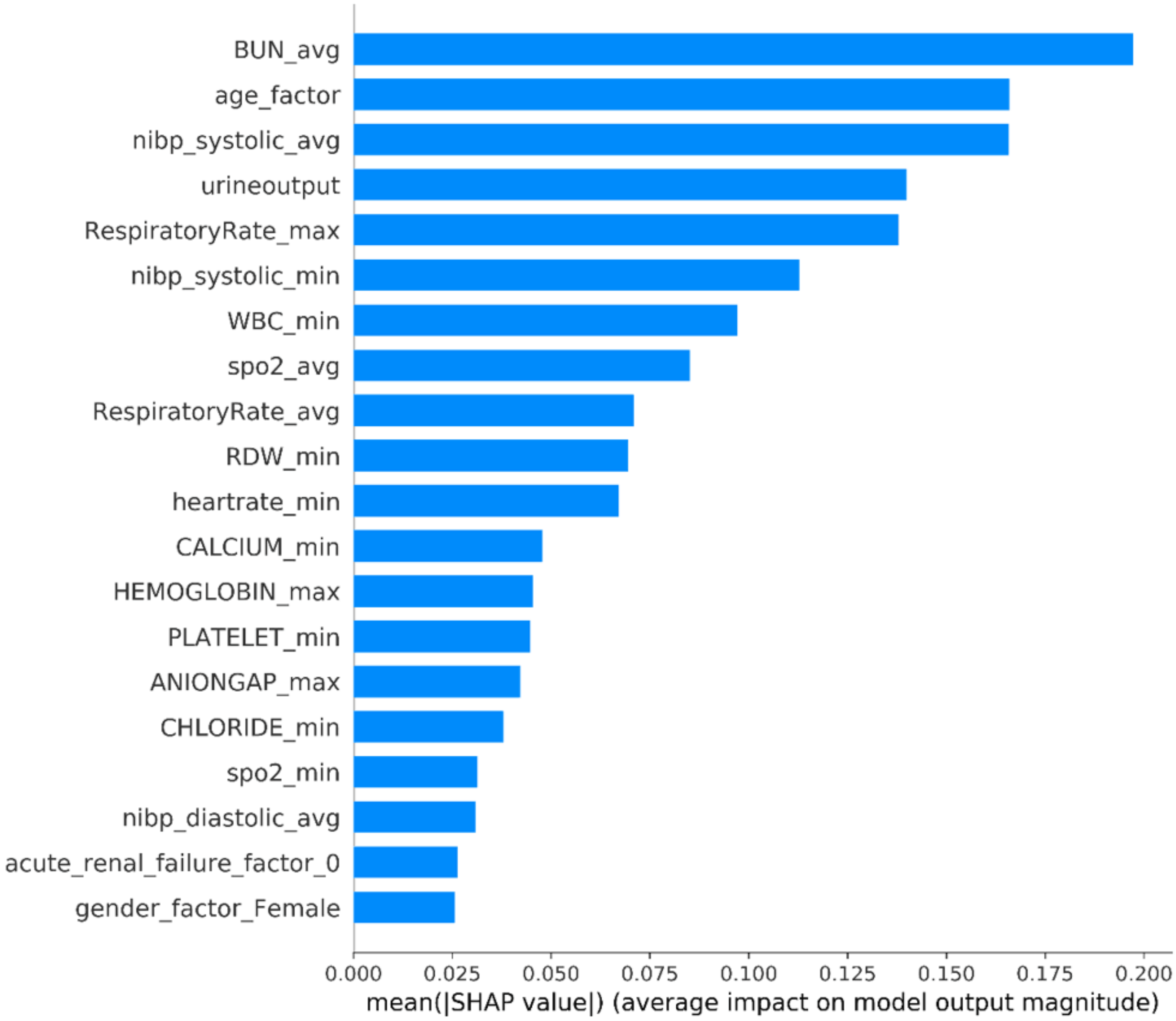
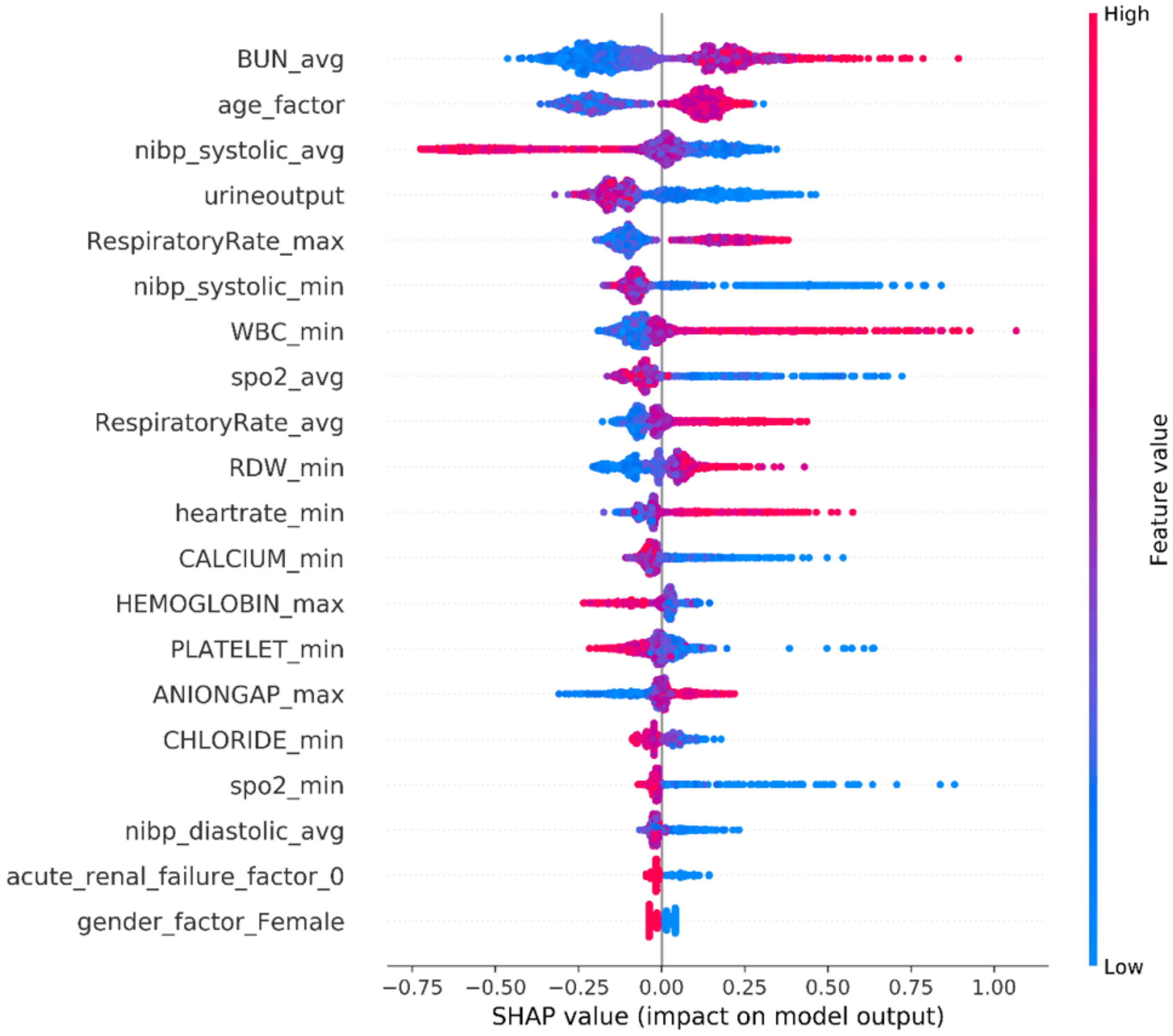


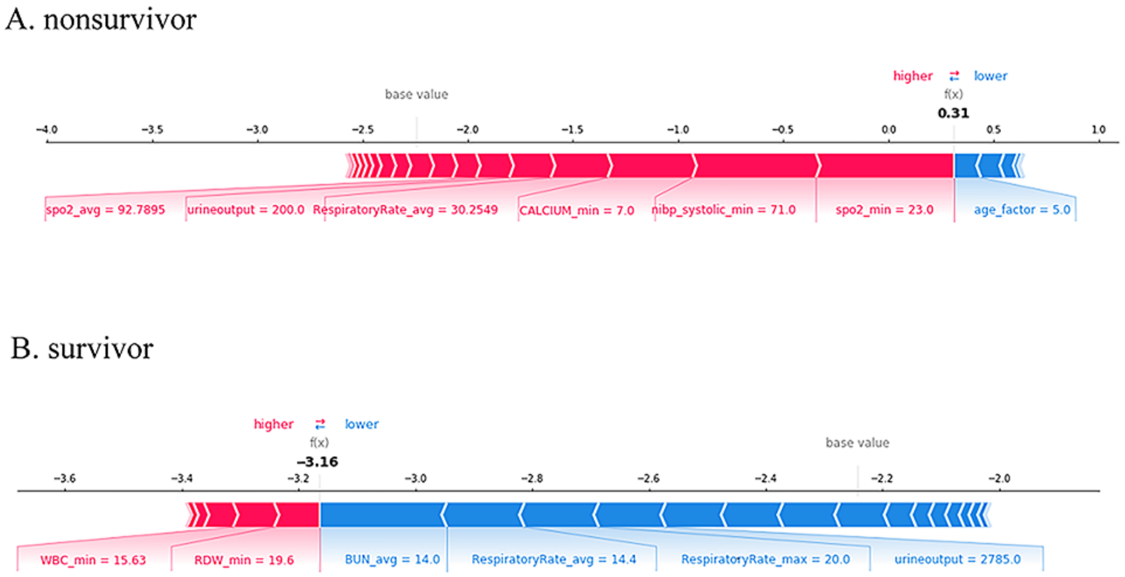
Figure 5. The SHapley Additive exPlanation (SHAP) values. Avg: average; BUN: blood urea nitrogen; max: maximum; min: minimum; NIBP: noninvasive blood pressure; RDW: red blood cell distribution width; SpO_2: O_2 saturation; WBC: white blood cell.



SHAP Individual Force Plots

Figure 6 shows the individual force plots for patients who (A) did not survive and (B) survived. The SHAP values indicate the prediction-related feature of individual patients and the contribution of each feature to the mortality prediction. The bold-faced numbers are the probabilistic predicted values (f(x)), whereas the base values are the values predicted without giving input to the model. The f(x) is the log odds ratio of each observation. The red features (on the left) indicate features that increase the mortality risk, and the blue features indicate features that decrease the mortality risk. The length of the arrows helps visualize the magnitude of the effect on the prediction. The longer the arrow, the larger the effect.

Figure 6. SHapley Additive exPlanation (SHAP) force plot for two selected patients.



A. nonsurvivor

B. survivor

Discussion

Principal Findings

In this retrospective cohort study of a large-scale public ICU database, we developed and validated four machine learning algorithms to predict the mortality of patients with HF. The XGBoost model outperforms the performance of LR, RF, and SVM. The SHAP method is used to explain the XGBoost model, which ensures the model performance and clinical interpretability. This will help physicians better understand the decision-making process of the model and facilitates the use of prediction results. Besides, to avoid ineffective clinical interventions, the relevant threshold probability range of DCA that we focused on was between 15% and 25%, during which XGBoost performed better. In critical care research, XGBoost has been widely used to predict the in-hospital mortality of patients and may assist clinicians in decision-making [29-31]. However, the mortality of patients with HF included in the final cohort is just 9.97%. Although DCA shows that the XGBoost model is better than the two default strategies, the positive predictive value is just 0.307 when the prediction specificity is fixed at 85%. Therefore, the XGBoost model may not be fully acceptable to provide decision-making support for clinicians. Evaluation of the benefits of earlier prediction of mortality and its additional cost is necessary in clinical practice.

Using SHAP to explain the XGBoost model, we identified some important variables associated with in-hospital mortality of patients with HF. In this study, the average BUN was recognized as the most important predictor variable. As a renal function marker to measure the amount of nitrogen in blood that comes from protein metabolism, previous studies also showed that BUN was the key predictor of HF mortality prediction with machine learning algorithms [32,33]. Kazory [34] concludes that BUN may be a biomarker of neurohormonal activation in patients with HF. From the perspective of pathophysiology, the activity of sympathetic nervous systems and the renin-angiotensin-aldosterone system increases with the aggravation of HF, which causes the vasoconstriction of the afferent arterioles. A reduction in renal perfusion further leads to water and sodium retention and promotes urea reabsorption, ultimately resulting in an increased BUN. However, further research is needed to explore the applicability of this SHAP method, due to the lack of an external validation cohort.

Limitations

This study had some limitations. First, our data were extracted from a publicly available database, and some variables were missing. For example, we intended to include more predictor variables that may affect in-hospital mortality such as prothrombin time as well as brain natriuretic peptide and lactate; however, the missing values were over 70%. Second, all data were derived from the ICU patients from the United States, so the applicability of our model remained unclear in other populations. Third, our mortality prediction models were based on data available within 24 hours of each ICU admission, which may neglect the subsequent events that change the prognosis and cause confounders to some extent. Fourth, due to lack of an external validation cohort, the applicability of the developed XGBoost model may not be very efficient in clinical practice. Currently, we are trying to collect data of patients with HF in ICUs from West China Hospital of Sichuan University. Although we have obtained some preliminary data, it is not feasible for the external prospective validation because of the limited sample size.

Conclusions

We developed the interpretable XGBoost prediction model that has the best performance in estimating the mortality risk in patients with HF. In addition, the interpretable machine learning approach can be applied to accurately explore the risk factors of patients with HF and enhance physicians' trust in prediction models. This will help physicians identify patients with HF who have a high mortality risk so as to timely apply appropriate treatments for them.

Authors' Contributions

J Liu, J Li, and SL conceived the study. SL, J Liu, J Li, YH, LZ, and YM performed the analysis, interpreted the results, and drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material 1: the ICD 9/10 codes.
[XLS File (Microsoft Excel File), 25 KB - jmir_v24i8e38082_app1.xls]

Multimedia Appendix 2
Supplementary material 2: the selected predictor variables.
[DOCX File , 17 KB - jmir_v24i8e38082_app2.docx]

Multimedia Appendix 3
Supplementary material 3: all predictor variables for patients in the training and testing data set.
[DOCX File , 21 KB - jmir_v24i8e38082_app3.docx]

References

1. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol 2014 Apr 1;63(12):1123-1133 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.053] [Medline: 24491689]
2. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail 2013 May;6(3):606-619 [FREE Full text] [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]
3. Liao L, Allen LA, Whellan DJ. Economic burden of heart failure in the elderly. Pharmacoeconomics 2008;26(6):447-462. [Medline: 18489197]
4. Kim HJ, Park M, Shin JI, Park J, Kim D, Jeon J, et al. Associations of heart failure with susceptibility and severe complications of COVID-19: a nationwide cohort study. J Med Virol 2021 Nov 05:1138-1145 [FREE Full text] [doi: 10.1002/jmv.27435] [Medline: 34738248]
5. Yonas E, Alwi I, Pranata R, Huang I, Lim MA, Gutierrez EJ, et al. Effect of heart failure on the outcome of COVID-19 - a meta analysis and systematic review. Am J Emerg Med 2021 Aug;46:204-211 [FREE Full text] [doi: 10.1016/j.ajem.2020.07.009] [Medline: 33071085]
6. Monteiro F, Meloni F, Baranauskas JA, Macedo AA. Prediction of mortality in intensive care units: a multivariate feature selection. J Biomed Inform 2020 Jul;107:103456 [FREE Full text] [doi: 10.1016/j.jbi.2020.103456] [Medline: 32454242]
7. Kramer AA, Dasta JF, Kane-Gill SL. The impact of mortality on total costs within the ICU. Crit Care Med 2017 Sep;45(9):1457-1463. [doi: 10.1097/CCM.0000000000002563] [Medline: 28658024]
8. Khurrum M, Asmar S, Joseph B. Telemedicine in the ICU: innovation in the critical care process. J Intensive Care Med 2021 Dec;36(12):1377-1384. [doi: 10.1177/0885066620968518] [Medline: 33111599]
9. Kang MW, Kim J, Kim DK, Oh K, Joo KW, Kim YS, et al. Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. Crit Care 2020 Feb 06;24(1):42 [FREE Full text] [doi: 10.1186/s13054-020-2752-7] [Medline: 32028984]
10. Liu C, Liu X, Mao Z, Hu P, Li X, Hu J, et al. Interpretable machine learning model for early prediction of mortality in ICU patients with rhabdomyolysis. Med Sci Sports Exerc 2021 Sep 01;53(9):1826-1834. [doi: 10.1249/MSS.0000000000002674] [Medline: 33787533]
11. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van DLMJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. Lancet Respir Med 2015 Jan;3(1):42-52 [FREE Full text] [doi: 10.1016/S2213-2600(14)70239-5] [Medline: 25466337]
12. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, et al. Prognostic Assessment of COVID-19 in the intensive care unit by machine learning methods: model development and validation. J Med Internet Res 2020 Nov 11;22(11):e23128 [FREE Full text] [doi: 10.2196/23128] [Medline: 33035175]

13. Stenwig E, Salvi G, Rossi PS, Skjærvold NK. Comparative analysis of explainable machine learning prediction models for hospital mortality. BMC Med Res Methodol 2022 Feb 27;22(1):53 [FREE Full text] [doi: 10.1186/s12874-022-01540-w] [Medline: 35220950]
14. Wang K, Tian J, Zheng C, Yang H, Ren J, Liu Y, et al. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. Comput Biol Med 2021 Oct;137:104813 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104813] [Medline: 34481185]
15. Zihni E, Madai VI, Livne M, Galinovic I, Khalil AA, Fiebach JB, et al. Opening the black box of artificial intelligence for clinical decision support: a study predicting stroke outcome. PLoS One 2020;15(4):e0231166 [FREE Full text] [doi: 10.1371/journal.pone.0231166] [Medline: 32251471]
16. eICU Collaborative Research Database. URL: https://eicu-crd.mit.edu/ [accessed 2022-07-20]
17. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 2018 Sep 11;5:180178 [FREE Full text] [doi: 10.1038/sdata.2018.178] [Medline: 30204154]
18. Li K, Shi Q, Liu S, Xie Y, Liu J. Predicting in-hospital mortality in ICU patients with sepsis using gradient boosting decision tree. Medicine (Baltimore) 2021 May 14;100(19):e25813 [FREE Full text] [doi: 10.1097/MD.0000000000025813] [Medline: 34106618]
19. Abraham WT, Fonarow GC, Albert NM, Stough WG, Gheorghiade M, Greenberg BH, OPTIMIZE-HF InvestigatorsCoordinators. Predictors of in-hospital mortality in patients hospitalized for heart failure: insights from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). J Am Coll Cardiol 2008 Jul 29;52(5):347-356 [FREE Full text] [doi: 10.1016/j.jacc.2008.04.028] [Medline: 18652942]
20. Fonarow GC, Adams KF, Abraham WT, Yancy CW, Boscardin WJ, ADHERE Scientific Advisory Committee, Study Group,Investigators. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA 2005 Feb 02;293(5):572-580. [doi: 10.1001/jama.293.5.572] [Medline: 15687312]
21. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985 Oct;13(10):818-829. [Medline: 3928249]
22. Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, Peterson ED, American Heart Association Get With the Guidelines-Heart Failure Program. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program. Circ Cardiovasc Qual Outcomes 2010 Jan;3(1):25-32. [doi: 10.1161/CIRCOUTCOMES.109.854877] [Medline: 20123668]
23. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Stat Methodol 1996;58(1):267-288 [FREE Full text] [doi: 10.1111/j.2517-6161.1996.tb02080.x]
24. Muthukrishnan R, Rohini R. LASSO: A feature selection technique in predictive modeling for machine learning. 2016 Presented at: IEEE international conference on advances in computer applications (ICACA); Oct 24; Coimbatore, India. [doi: 10.1109/icaca.2016.7887916]
25. Blazek K, van Zwieten A, Saglimbene V, Teixeira-Pinto A. A practical guide to multiple imputation of missing data in nephrology. Kidney Int 2021 Jan;99(1):68-74. [doi: 10.1016/j.kint.2020.07.035] [Medline: 32822702]
26. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17; Dec 4-9; Long Beach, CA p. 4768-4777.
27. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol 2018 Dec;74(6):796-804 [FREE Full text] [doi: 10.1016/j.eururo.2018.08.038] [Medline: 30241973]
28. Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanapragasam VJ. Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the Surveillance, Epidemiology, and End Results (SEER) database. Lancet Digit Health 2021 Mar;3(3):e158-e165 [FREE Full text] [doi: 10.1016/S2589-7500(20)30314-9] [Medline: 33549512]
29. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. J Transl Med 2020 Dec 07;18(1):462 [FREE Full text] [doi: 10.1186/s12967-020-02620-5] [Medline: 33287854]
30. Liu J, Wu J, Liu S, Li M, Hu K, Li K. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. PLoS One 2021;16(2):e0246306 [FREE Full text] [doi: 10.1371/journal.pone.0246306] [Medline: 33539390]
31. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Crit Care 2019 Apr 08;23(1):112 [FREE Full text] [doi: 10.1186/s13054-019-2411-z] [Medline: 30961662]
32. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, Urey MA, et al. Improving risk prediction in heart failure using machine learning. Eur J Heart Fail 2020 Jan;22(1):139-147 [FREE Full text] [doi: 10.1002/ejhf.1628] [Medline: 31721391]
33. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC Heart Fail 2020 Jan;8(1):12-21 [FREE Full text] [doi: 10.1016/j.jchf.2019.06.013] [Medline: 31606361]

34. Kazory A. Emergence of blood urea nitrogen as a biomarker of neurohormonal activation in heart failure. Am J Cardiol 2010 Sep 01;106(5):694-700. [doi: 10.1016/j.amjcard.2010.04.024] [Medline: 20723648]

Abbreviations

AUC: area under the curve
BUN: blood urea nitrogen
DCA: decision curve analysis
eICU-CRD: eICU Collaborative Research Database
HF: heart failure
ICU: intensive care unit
LASSO: Least Absolute Shrinkage and Selection Operator
LR: logistic regression
RF: random forest
SHAP: SHapley Additive exPlanation
SVM: support vector machine

Edited by R Kukafka; submitted 18.03.22; peer-reviewed by L Xu, Z Zhang; comments to author 31.05.22; revised version received 07.07.22; accepted 15.07.22; published 09.08.22.

Please cite as:
Li J, Liu S, Hu Y, Zhu L, Mao Y, Liu J
Predicting Mortality in Intensive Care Unit Patients With Heart Failure Using an Interpretable Machine Learning Model: Retrospective Cohort Study
J Med Internet Res 2022;24(8):e38082
URL: https://www.jmir.org/2022/8/e38082
doi:10.2196/38082
PMID:35943767

Original Paper

The Adoption of Artificial Intelligence in Health Care and Social Services in Australia: Findings From a Methodologically Innovative National Survey of Values and Attitudes (the AVA-AI Study)

Sebastian Isbanner[1], BBE, MBA, PhD; Pauline O'Shaughnessy[2], BActSt, MActSt, PhD; David Steel[2], AStat, BSc, MSc, PhD; Scarlet Wilcock[3], BA, LLB, GDLP, PhD; Stacy Carter[4], BAppSci, MPH, PhD

[1]Social Marketing @ Griffith, Griffith Business School, Griffith University, Brisbane, Australia

[2]School of Mathematics and Applied Statistics, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia

[3]Australian Research Council Centre of Excellence for Automated Decision-Making and Society, The University of Sydney Law School, The University of Sydney, Sydney, Australia

[4]Australian Centre for Health Engagement Evidence and Values, Faculty of the Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, Australia

Corresponding Author:
Stacy Carter, BAppSci, MPH, PhD
Australian Centre for Health Engagement Evidence and Values
Faculty of the Arts, Social Sciences and Humanities
University of Wollongong
Northfields Ave
Wollongong, 2522
Australia
Phone: 61 2 4221 3243
Email: stacyc@uow.edu.au

Abstract

Background: Artificial intelligence (AI) for use in health care and social services is rapidly developing, but this has significant ethical, legal, and social implications. Theoretical and conceptual research in AI ethics needs to be complemented with empirical research to understand the values and judgments of members of the public, who will be the ultimate recipients of AI-enabled services.

Objective: The aim of the Australian Values and Attitudes on AI (AVA-AI) study was to assess and compare Australians' general and particular judgments regarding the use of AI, compare Australians' judgments regarding different health care and social service applications of AI, and determine the attributes of health care and social service AI systems that Australians consider most important.

Methods: We conducted a survey of the Australian population using an innovative sampling and weighting methodology involving 2 sample components: one from an omnibus survey using a sample selected using scientific probability sampling methods and one from a nonprobability-sampled web-based panel. The web-based panel sample was calibrated to the omnibus survey sample using behavioral, lifestyle, and sociodemographic variables. Univariate and bivariate analyses were performed.

Results: We included weighted responses from 1950 Australians in the web-based panel along with a further 2498 responses from the omnibus survey for a subset of questions. Both weighted samples were sociodemographically well spread. An estimated 60% of Australians support the development of AI in general but, in specific health care scenarios, this diminishes to between 27% and 43% and, for social service scenarios, between 31% and 39%. Although all ethical and social dimensions of AI presented were rated as important, accuracy was consistently the most important and reducing costs the least important. Speed was also consistently lower in importance. In total, 4 in 5 Australians valued continued human contact and discretion in service provision more than any speed, accuracy, or convenience that AI systems might provide.

Conclusions: The ethical and social dimensions of AI systems matter to Australians. Most think AI systems should augment rather than replace humans in the provision of both health care and social services. Although expressing broad support for AI, people made finely tuned judgments about the acceptability of particular AI applications with different potential benefits and downsides. Further qualitative research is needed to understand the reasons underpinning these judgments. The participation of ethicists, social scientists, and the public can help guide AI development and implementation, particularly in sensitive and value-laden domains such as health care and social services.

(J Med Internet Res 2022;24(8):e37611) doi:10.2196/37611

KEYWORDS

artificial intelligence; surveys and questionnaires; consumer health informatics; social welfare; bioethics; social values

Introduction

Background

Artificial intelligence (AI) and automation are accelerating in many fields driven by an increase in the availability of massive linked data sets, cloud computing, more powerful processors, and the development of new types of algorithms, particularly in the field of machine learning. In this paper, AI will be broadly conceptualized, consistent with the Australian Council of Learned Academies definition, as "a collection of interrelated technologies used to solve problems and perform tasks that, when humans do them, requires thinking" [1]. These technologies are being applied in social services, including to automate eligibility verification, target and personalize welfare services, and aid in the detection of fraud and debt liability [2,3]. Health care, initially slow to adopt AI, is also seeing rapid development for applications including health service planning and resource allocation, triage, screening and diagnosis, prognostication, robotics in applications such as aged care, and health advice chatbots [4-6]. These areas of practice—social services and health care—have traditionally been provided via extensive human-to-human contact by staff with professional autonomy and the capacity to exercise discretion in handling the problems of service users or patients.

Ethical, Legal, and Social Implications of AI

A growing body of literature acknowledges the complex ethical, legal, and social implications (ELSI) of AI deployment [1,7,8]. In the 2010s, many intergovernmental, academic, and industry groups examined the ELSI of AI in a general sense, producing lists of high-level principles for AI ethics [1,7,9,10] often reminiscent of existing frameworks in bioethics [11]. In parallel, a set of approaches that foreground the significance of power, oppression, discrimination, and injustice has been developed, contextualized in the sociotechnical systems in which AI is embedded [12]. Other work critiques the corporate AI sector for establishing AI ethics boards and documents while persisting with unethical practices and points to the difficulties faced by AI ethics researchers when working inside corporations producing AI systems [13]. The abstract principles and frameworks that have proliferated in AI ethics offer accessible ways in to ethical debates, but they cannot be sufficient to address ethical issues in practice [14]. There are now calls to complement ethical frameworks with other forms of knowledge, including analysis of detailed use cases and investigation of what members of the public think and value regarding the use of AI [15]. Two linked cases are the focus of this study: the use of AI in health services and in social services, which are an important social determinant of health especially for marginalized and disadvantaged populations.

For the first case, health care AI, research on ELSI has been rapidly expanding since 2019. In a 2020 review, Morley et al [16] highlighted 3 groups of ELSI issues for health care AI: epistemic concerns (that the evidence on which health care AI is based is inconclusive, inscrutable, or misguided), normative concerns (highlighting unfairness and the potential for transformative unintended consequences), and concerns about the ability to either identify algorithmic harm or ascribe responsibility for it. Another 2020 review focused on health care emphasized the potential to worsen outcomes or cost-effectiveness, the problem of transportability (that algorithms may not work equally well in different populations), automation bias (that humans tend to be too willing to accept that algorithmic systems are correct), the potential to intensify inequities, the risk of clinical deskilling, increased threats to data protection and privacy, lack of contestability of algorithmic decisions, the need to preserve clinician and patient autonomy, and the potential to undermine trust in health care systems [17]. A 2021 scoping review on health care AI ELSI highlighted data privacy and security, trust in AI, accountability and responsibility, and bias as key ethical issues for health care AI [18]. Also in 2021, Goirand et al [19] identified 84 AI-specific ethics frameworks relevant to health and >11 principles recurring across these while noting that few frameworks had been implemented in practice. In parallel, empirical evidence demonstrates a continuing need to address the ELSI of health care AI. A well-known example is an AI system used to allocate health care in many US health services that allocated more care to White patients than to Black patients, even when the Black patients had greater need, because the AI learned from historical underservicing that Black patients had lower care requirements [20].

Regarding our second case, AI in the social services, ELSI research is also gaining momentum, particularly as part of broader inquiries into the digital welfare state or in relation to high-profile examples of technology failure [2,21,22]. This research highlights the potential of AI to improve the consistency and accuracy of welfare decision-making and increase cost-efficiency. However, it also raises grave concerns regarding the social costs associated with implementing AI in the social services, particularly for vulnerable populations. For example, the pioneering ethnographic study by Eubanks [21] of AI and automation technologies in the United States in 2018 illustrates how new technologies can disempower poor citizens, intensify existing patterns of discrimination, and *automate inequality*. Similar concerns have been raised in Australia in relation to the Online Compliance Intervention known as *robodebt*. The scheme automated the calculation of welfare debts based on an income-averaging algorithm. The legality of the algorithm was successfully challenged before a domestic court in 2019, culminating in an Aus $1.8 billion (US $1.25

billion) class action lawsuit against the Australian government and prompting significant public and scholarly criticism of the scheme [23].

AI applications in the welfare sector pose novel challenges to legal and regulatory compliance. Many AI systems, including robodebt, have been designed and implemented in the absence of proper legal frameworks or in contravention of prevailing laws and administrative principles [2,24]. Other high-profile examples include the System Risk Indication system of the Dutch government, which was used to predict an individual's risk of welfare fraud. System Risk Indication was successfully challenged based on the fact that the system breached the right to privacy contained in the European Convention on Human Rights [2]. Such cases have prompted a growing body of literature concentrated on the legal and human rights implications of AI in the social services. The recent report by the United Nations Special Rapporteur on Extreme Poverty [2] calls for a human rights–based approach to digital regulation in social protection systems, which has prompted further research on AI and human rights principles [25].

Existing Research on Perceptions of the ELSI of Using AI, Including in Health Care and Social Services

An approach to thinking about the ELSI of AI is to examine public attitudes and judgments toward these technologies. In areas such as health care and social services, this includes the attitudes and judgments of patients and service users. A small body of literature exists on general attitudes toward AI. In 2018, Zhang and Dafoe [26] surveyed 2000 American adults and found mixed support for developing AI and strong support for the idea that AI should be carefully managed. In April 2020, the Monash Data Futures Institute surveyed 2019 Australians on their attitudes toward AI, adapting some questions from Zhang and Dafoe [27]. They found that Australians did not consider themselves knowledgeable about AI, but 62.4% expressed support for the development of AI. When asked whether they supported the use of AI in particular fields, respondents were most supportive of AI use in health (44.1% strong support) and medicine (43% strong support) and less supportive of AI use in *equality and inclusion* (21.5% strong support) and public and social sector management (20.2% strong support). Respondents tended to agree that AI would do more social good than harm overall [27].

Research on the attitudes of patients and service users is developing; most research to date—such as this study—has been speculative, asking informants about their views or intentions rather than their direct experience of AI. Studies asking patients to imagine the use of AI in their care generally report broad acceptance [28-30] in areas including skin cancer screening and neurosurgery. Promises of greater diagnostic accuracy are well received [30], and sharing deidentified health data for the development of medical AI may be acceptable to most [28]. A study reported experiences with a diabetic retinopathy screening AI—96% of patients were satisfied or very satisfied [31]. However, respondents in most studies also express concerns. Regarding skin cancer screening, concerns included inaccurate or limited training sets; lack of context; lack of physical examination; operator dependence; data protection; and potential errors, including false negatives and false positives [28,30]. In the context of neurosurgery, respondents wanted a human neurosurgeon to remain in control [29]. Finally, a study of patients with cancer in China suggested that despite reporting that they *believed in* both diagnoses and therapeutic advice given by an AI (90% and 85%, respectively), when this differed from the advice given by a human clinician, most patients would prefer to take the human clinician's recommendation (88% and 91%, respectively) [32].

Research examining public and professional attitudes toward AI in the welfare sector is very limited. To the authors' knowledge, research is yet to explore citizens' general attitudes toward AI in the domain of welfare provision. However, there is a small body of research documenting service users' experiences of specific AI applications in the social services, particularly users' negative experiences of exclusion and discrimination [21,33], providing context-specific insights into system users' experiences of AI and illustrating the high-stakes nature of implementing AI in this domain. This work, together with some small-scale, mostly qualitative studies involving frontline social service staff [34-38], illustrates the complex and dynamic relationship between AI and the routines of social welfare professionals and indicates mixed reactions to these systems among staff. For example, the study by Zejnilović et al [36] of counselors in a Portuguese employment service in 2020 found high levels of distrust and generally negative perceptions of an AI system used to score clients' risk of long-term unemployment. However, the survey data also indicated that workers would continue to rely on the system even if it became optional, suggesting that respondents harbor mixed feelings about the system.

The Australian Values and Attitudes on Artificial Intelligence (AVA-AI) study set out to understand Australians' values and attitudes regarding the use of AI in health care and social services. Australia has been relatively slow to approve and adopt medical AI compared, for example, with the United Kingdom and the United States. The adoption of AI and automation technologies in the social services is comparatively advanced in Australia, although its development has been uneven and marked by controversy, including the case of robodebt. Multiple stakeholders are now confronting the opportunities and risks of these technologies. Policy makers need high-quality evidence of what Australians consider acceptable or unacceptable to ensure that their decision-making is legitimate. This study used an innovative methodology to survey Australians regarding these questions. Our aims were to understand Australians' front-of-mind normative judgments about the use of AI, especially in the underresearched fields of social services and health care, and what attributes of AIs they would consider to be most important if those AIs were to be deployed in health care and social services. Although parallel literature seeks to model the characteristics of AI that predict acceptance [39], this work has the complementary aim of seeking to understand the prevalence and patterning of different normative judgments about AI.

The research questions answered in this study are as follows: (1) How do Australians' general judgments regarding the use of AI compare with their judgments regarding the particular uses of AI in health care and social services? (2) Do Australians make different judgments about different health care and social service applications of AI? (3) What attributes of health care and social service AI systems do Australians consider most important?

Methods

Aims

The *AVA-AI study* was conducted to (1) provide information on Australians' attitudes and values regarding AI, especially in health care and social services, and (2) allow for analysis of how these vary across different subpopulations and are associated with people's sociodemographic characteristics and familiarity with technology. This study focuses on attitudes and values, how they differ for different scenarios, and the relative importance of different attributes of health care and social service AI. A selection of concepts from AI ethics relevant to understanding this study is outlined in Textbox 1. Analyses across different subpopulations will be reported in future papers.

Textbox 1. Concepts from artificial intelligence (AI) ethics used in the Australian Values and Attitudes on Artificial Intelligence (AVA-AI) study.

Concept and meaning in the context of AI ethics

- Accuracy: the degree to which an AI can perform tasks without errors. In the context of screening or targeting, for example, this would include the ability of the AI to detect a condition or identify a person without false positives (where a case is identified as having a condition or being a target when they do not fit the criteria). It also includes the ability of the AI to avoid false negatives (where a case is identified as not having a condition or not being a target when they do fit the criteria).
- Algorithmic targeting: the use of AI to find people with a certain profile, often predictively (eg, to identify people likely to be unable to find work or people likely to commit a crime).
- Autonomous machine decision-making: situations in which an AI makes a decision that would previously have been made only by a person, for example, whether a person has a condition or whether a person is eligible for a social security payment.
- Contestability: whether machine decision-making can be effectively challenged. Contestability is to some extent dependent on explainability but is also dependent on policy settings.
- Explainability: whether it is possible to explain how an AI makes a decision. For some forms of AI, especially deep learning algorithms, humans do not explicitly instruct the AI on what basis it should make decisions. This makes explainability potentially more challenging, leading such algorithms to be labeled as *black box* algorithms.
- Deskilling: when tasks previously undertaken by humans are delegated to AI, humans lose their ability to complete those tasks; that is, they deskill in relation to those tasks.
- Fair treatment: AI systems tend to reflect human bias; this relates to the concept of justice, which is complex and multidimensional. Doing justice is unlikely to entail treating everyone identically as different people have different needs and opportunities. In the AVA-AI study, we asked respondents how important it was to "know that the system treats everyone fairly" to capture an intuitive judgment of a system's capacity to deal justly or unjustly with different individuals and populations.
- Personal tailoring: the ability of an AI, by comparing the data of an individual with large, linked data sets, to recommend services or interventions that respond to the particularity of an individual's situation.
- Privacy: freedom from intrusion into personal matters, including the ability to control personal information about oneself.
- Responsibility: a complex and multidimensional concept, which attributes moral or legal duties and moral or legal blame, including for errors or harms.

Instrument Development

When designing the study, there were no existing instruments we could adopt. We used a question from the 2018 survey by Zhang and Dafoe [26] and developed other questions based on a review of the AI ethics literature. Before the study commenced, the instrument underwent multiple rounds of input from investigators and expert colleagues, as well as cognitive testing.

Final Instrument Design

In addition to sociodemographic variables, the survey asked about the use of AI in health care and welfare. Questions were of 2 types. The first type, in the form of *How muchdo you support or oppose*, presented a 5-point scale. Questions of this type asked about the development of AI in general (B01, taken from Zhang and Dafoe [26], running from *strongly support* to *strongly oppose*) and the use of AI in 6 particular health care and welfare AI scenarios for which potential advantages and disadvantages were presented in a balanced way (C03-C05 and D03-D05, for which the 5-point scale ran from *I support this use of AI* to *I oppose this use of AI*; Multimedia Appendix 1). A final question of this type (E01) asked respondents to indicate what they valued more on a 5-point scale: *Quicker, more convenient, more accurate health and social services* or *More human contact and discretion in health and social services.* This trade-off asked respondents to evaluate a bundle of benefits commonly attributed to AI-enabled services against a bundle of benefits commonly attributed to services provided by human professionals.

The second type of question presented a scenario involving AI use and then asked respondents to consider 7 ELSI dimensions or values (eg, *getting an answer quickly* and *getting an accurate answer*) and rate how important each dimension was to them personally on a scale from *extremely important* to *not at all* *important*. There were 4 questions of this type: 2 with health care scenarios (C01-C02) and 2 with welfare scenarios (D01-D02). Module C presented health care questions and module D presented welfare questions; respondents were randomly allocated to receive module C or D first, and the order of presentation of the values was also randomized. Table 1 summarizes the variables presented as well as the concepts each question was designed to assess. Note that the dimensions or values were identical for module C and D questions except that the health care questions had an item about responsibility, including mistakes (reflecting the status quo of medical professional autonomy), whereas the social service questions had an item about personal tailoring (reflecting a promised potential benefit of AI in social services).

The final survey instrument is provided in Multimedia Appendix 1.

Table 1. Summary of the variables collected in the Australian Values and Attitudes on Artificial Intelligence (AI) study.

Type of variable	Question number and variable	Concepts tested
General support or opposition	• B01—how much do you support or oppose the development of AI in general (with multiple examples given)?[a]	• Broad support for or opposition to AI

Importance of different attributes of AI in health care scenarios	• C01—machine reads medical test, diagnoses, and recommends treatment
	• C02—machine triages when you are unwell
		• In relation to:
		  • C01—delegation of clinical decisions to an autonomous machine
		  • C02—automating decisions about need for health care services (time-sensitive)
		• Importance of:
		  • Explanation
		  • Speed
		  • Accuracy
		  • Human contact
		  • Reducing system costs
		  • Fair treatment
		  • Responsibility

Importance of different attributes of AI in welfare scenarios	• D01—machine processes application for unemployment benefits (data sharing required)
	• D02—chatbot advises about carer payments
		• In relation to:
		  • D01—foregoing privacy as a barrier to access services
		  • D02—automation of information services
		• Importance of:
		  • Explanation
		  • Speed
		  • Accuracy
		  • Human contact
		  • Reducing system costs
		  • Fair treatment
		  • Personal tailoring

Support for or opposition to AI in specific health care scenarios	• C03—nonexplainable hospital algorithms • C04—data sharing for quality care • C05—deskilling physicians	• C03—importance of explainable machine recommendations • C04—importance of privacy (balanced against quality of care) • C05—importance of retaining human clinical skills
Support for or opposition to AI in specific welfare scenarios	• D03—targeted compliance checking • D04—nonexplainable job services • D05—automated assignment of parent support with limited contestability	• D03—algorithmic targeting of punitive policy • D04—importance of explainable machine recommendations • D05—importance of contestability (balanced against accuracy)
Speed—human contact	• E01—trade-off between quicker, more convenient, more accurate health care and social services and more human contact and discretion in health care and social services	• E01—speed and convenience and accuracy vs human contact and discretion
Sociodemographic	• Age, gender, concession card type, and employment status; household income, education, household type, language other than English spoken at home, and general health • Centrelink payment, employment field, relevant experience, relevant degree, life satisfaction, and disability	• Descriptive variables collected using standard sociodemographic questions
Geographic	• State or territory, capital city or rest of state, and SEIFA[b] (geographic measure of disadvantage)	• Descriptive variables collected using standard questions about location of residence

Type of variable	Question number and variable	Concepts tested
Lifestyle	- How often they check the internet, how often they post comments or images to social media, how often they post on blogs, forums, or interest groups, early adopter by type, and television viewing by type of viewing	- Variables collected for weighting purposes

[a] Variables in italics were collected from both the Life in Australia and web-based panel samples; all others were collected from the web-based panel alone.

[b] SEIFA: Socio-Economic Indexes for Areas.

Data Collection Processes and Weighting

Data collection occurred between March 16, 2020, and March 29, 2020, with respondents mainly completing the questionnaire on the web.

The AVA-AI study comprises 2 sample components: one obtained from the Life in Australia (LIA) survey [40] with a responding sample size of 2448 and a web-based panel sample with a responding sample size of 2000. Thus, the combined responding sample size was 4448.

The full set of questions was used for the web-based panel sample. For the LIA sample, a subset of sociodemographic variables and all the geographic and lifestyle questions were used. The LIA sample also answered the general support question (B01) and the importance of AI attributes for scenario C01. In Table 1, the variables in italics were collected from both the LIA and web-based panel samples, and all others were collected from the web-based panel alone.

The LIA sample was selected using scientific probability sampling methods, whereas the web-based panel sample was a nonprobability sample. Weights for the LIA sample were calculated using standard methods for a probability sample using generalized regression estimation [41] to adjust for differences in selection probabilities and nonresponse and calibrate to population benchmarks obtained from the population census, current demographic statistics, and the 2017 to 2018 National Health Survey obtained from the Australian Bureau of Statistics. The variables used in the calibration were age by highest education level, country of birth by state, smoking status by state, gender by state, household structure by state, part of state, and state or territory.

A web-based panel allowed us to generate a relatively large sample, enabling a good level of disaggregation into subpopulations, comparisons between groups, and analysis of associations. Such panels can be subject to self-selection biases and coverage issues, reducing the accuracy of population prevalence estimates [42], but may enable the examination of associations and, with adjustments to reduce biases, improve the estimation of population characteristics [43]. The calibration to population benchmarks for major sociodemographic variables may not eliminate these issues. To enhance our adjustment of the web-based panel data in the AVA-AI study, we included 2 substantive questions, a set of behavioral and lifestyle questions, and major sociodemographic variables in both the web-based panel survey and the probability sample–based LIA survey, as indicated in Table 1. This approach was similar to that used in the study by Zhang and Dafoe [26], although our approach for the AVA-AI study went further by adjusting for behavioral and lifestyle variables and 2 substantive variables. The use of behavioral and lifestyle variables in adjusting web surveys, also known as webographic variables, is discussed in the study by Schonlau et al [44], for example.

In the AVA-AI study, questions common to the LIA and web-based panel samples were used to calibrate the web-based panel to the LIA sample, producing weights designed to reduce potential biases owing to the web-based panel sample being nonrandom; the LIA served as a reference survey [35]. The probability of inclusion for the web-based panel respondents was estimated using a propensity score model. This involved combining the LIA and web-based panel samples and fitting a logistic regression model, with the response variable being membership of the web-based panel. In fitting this model, the original LIA weights were used for respondents in that sample, and a weight of 1 was used for the web-based panel respondents. The variables used in the logistic regression were selected using Akaike Information Criterion–based stepwise regression and consisted of age by education, gender, household structure, language spoken at home, self-rated health, early adopter status, and television streaming watching. In a final calibration step, the weights were further adjusted to agree with the population benchmarks for these variables. This approach is described, for example, in the study by Valliant and Dever [45,46] and by Elliot and Valliant [47]. The weighting led to a weighted sample of 1950 for the web-based panel and 2498 for the LIA sample.

Statistical Analysis Methods

Overview

All estimates and analyses were based on a weighted analysis using the largest sample possible. Each respondent had a weight determined by the sample they came from. The weights were scaled so that the sum of the weights for the combined sample was 4448. Two substantive questions (B01 [general support or opposition] and C01 [support or opposition for autonomous machine decision-making in medical testing]) were asked to the combined LIA+web-based panel sample. The remainder of the attitude and value questions was asked only to the web-based panel sample. Any analysis involving questions included in the LIA and web-based panel sample was based on the combined sample and the associated weights. Any analysis involving questions that were only collected from the web-based panel sample was based on the web-based panel sample and the associated weights.

The analyses focused on determining and comparing the distribution of responses to the attitude and value questions. The methods used accounted for the use of weights in calculating estimates and associated 95% CIs and allowed for the testing of statistical significance, assessed when the *P* value of the relevant statistical test was <.05.

Statistical Analysis of Each Question Using Univariate Analyses

All variables concerning attitudes and values had 5 substantive response categories reflecting *support* or *importance.* Univariate analysis calculated the estimated percentage in each response category for each question, with 95% CIs for each estimated percentage. For questions asking for degree of support or opposition, we examined whether there was a majority support and compared across scenarios and between health care and welfare contexts; for questions asking for the importance attached to different attributes or values, we examined whether attributes or values mattered more in some contexts than others.

Weights must be accounted for in the calculation of estimates and in the statistical inference, such as estimates of SEs and the associated CIs obtained from them and *P* values for any statistical tests used. The CIs and *P* values were obtained using *Complex Samples* in SPSS (version 26; IBM Corp), which accounts for the use of weights in producing the estimates. Although the use of weights can reduce bias, there is an associated increase in variances and SEs of the estimates. This is reflected in the design effect, the variance of an estimate accounting for the weights (and complex design if used), compared with the use of simple random sampling and no weighting. The effect is variable specific, but a broad indication can be obtained considering the design effect because of weighting or unequal weighting effect [48,49]. This is 1+*Cw2*, where *Cw* is the coefficient of variation of the weights, which is the SD of the weights divided by their mean. For the combined sample, the design effect because of weighting was 1.83; for the LIA, it was 1.99; and, for the web-based panel, it was 1.61. For any specific estimates or analysis in this study, the SEs estimated from the survey data accounting for the weights were used. The effect on the SE is the square root of the design effect (ie, the design factor [50]) and is the factor by which the CIs are larger than if weights did not have to be used. A design effect of 1.83 implies a design factor of 1.35. In this analysis, the design effects were almost all between 1.50 and 2.00.

For questions using ordinal scales from 1 to 5, we also calculated an overall mean response to each question and the associated 95% CI. These included variables assessing the degree of support (ie, B01, C03-C05, and D03-D05), importance attached to attributes of AI (ie, C01-C02 and D01-D02), and the final question (E01) on trading off machine versus human traits. Mean scores close to the midpoint of the scale (3.00) indicated an overall neutral or balanced response to the question, that is, an equal or symmetric distribution of respondents on the respective scale. For support-or-oppose questions, lower scores indicated support and higher scores indicated opposition; for importance questions, lower scores indicated greater importance and higher scores indicated less importance; for E01, lower scores favored machine traits and higher scores favored human traits. For all questions, we tested the null hypothesis that the mean was 3.00 (ie, a distribution centered at the midpoint of the scale, or a balanced distribution of responses) using a 2-tailed *t* test allowing for weighting.

Statistical Analysis Comparing Responses to Questions Using Bivariate Analyses

To assess differences in the responses to pairs of questions—for example, is the support for the use of AI different when respondents are presented with different scenarios?—we compared the distributions of the responses. This was not to assess whether the responses to the 2 questions were independent, which is unlikely, but whether the percentages in their marginal distributions were the same.

Our goal was to determine what percentage of people changed their response between 2 questions and whether this change was net positive or negative. To examine this issue for any 2 questions, we created a *shift variable* to represent the difference between two variables (variables A and B): (1) if the response to variable A was in a category greater than the response to variable B, the *shift variable* was +1, which corresponded to a more positive attitude toward AI for variable B and, equivalently, a more negative attitude for variable A; (2) if the response to variable B was in a category greater than the response to variable A, the *shift variable* was −1, which corresponded to a more positive attitude toward AI for variable A and, equivalently, a more negative attitude for variable B; and (3) if the responses to variables A and B were identical, the *shift variable* was 0.

We estimated the percentage of respondents where the *shift variable* was 0, indicating no change. For those that changed, we estimated the percentage with a shift variable of −1, corresponding to a more positive attitude for the first variable and a more negative attitude for the second variable, and tested for equal percentages of positive and negative changes. The adjusted Pearson chi-square test in SPSS *Complex Samples* was used, which is a variant of the second-order adjustment proposed by Rao and Scott [51]. These tests allowed us to assess the statistical significance of the differences in responses under different scenarios.

We also tested for equal marginal distributions using the ordinal scores. SPSS uses a paired *t* test using these scores, which is similar to the test for marginal homogeneity described in the study by Agresti [52]. This test was implemented accounting for the weights using *Complex Samples* in SPSS by creating a variable for each person equal to the difference between the scores of the 2 questions and testing that the mean difference was 0. We tested answers to our research questions, that is, to determine whether respondents answered differently when questions tested the same ELSI concept in different settings or when questions tested different ELSI concepts in comparable settings. The estimated mean difference and associated 95% CI and the *P* value for the test that the mean difference was 0 were produced.

Ethical Considerations

This study was approved by the University of Wollongong Social Sciences Human Research Ethics Committee (protocol number 2019/458).

Results

Sample Composition

Table 2 provides a summary of the weighted combined sample and web-based panel sample for the key variables. A full composition of the overall combined sample and the web-based panel, including unweighted and weighted frequencies and proportions for key sociodemographic variables, is provided in Multimedia Appendix 2. The use of weights improved the representation of the combined sample for capital cities, age groups <35 years, men, employed status, nonuniversity as the highest level of education, language other than English spoken at home, those with excellent or very good health, and people who look for information over the internet several times a day. The sample was well spread and had respondents across many different sociodemographic groups.

The web-based panel sample was also well spread across many different sociodemographic groups. The effect of weighting was similar to that in the overall sample, although there was very little effect for age and capital cities. Comparing the weighted percentages between the combined sample and the web-based panel sample, the only appreciable difference is for those employed (2709/4448, 60.9% vs 1061/1950, 54.41%, respectively).

Table 2. Sociodemographic composition of Australian artificial intelligence survey sample (weighted data only).

	Combined sample (n=4448), n (%)	Web-based panel (n=1950), n (%)
Part of state		
Capital city	2957 (66.48)	1300 (66.67)
Rest of state	1481 (33.3)	640 (32.82)
Not stated or unknown	10 (0.22)	10 (0.51)
Age group (years)		
18 to 34	1386 (31.16)	637 (32.67)
35 to 54	1472 (33.09)	660 (33.85)
55 to 74	1166 (26.21)	497 (25.49)
≥75	394 (8.86)	156 (8)
Not stated or unknown	30 (0.67)	0 (0)
Gender		
Men	2180 (49.01)	939 (48.15)
Women	2259 (50.79)	1011 (51.85)
Other	9 (0.2)	1 (0.05)
Not stated or unknown	0 (0)	0 (0)
Employment status		
Employed	2709 (60.9)	1061 (54.41)
Not employed	1735 (39.01)	890 (45.64)
Not stated or unknown	4 (0.09)	0 (0)
Highest education level		
Postgraduate qualification	529 (11.89)	246 (12.62)
Undergraduate or diploma	1393 (31.32)	676 (34.67)
Vocational qualification	937 (21.07)	398 (20.41)
School qualification	1492 (33.54)	626 (32.1)
Not stated or unknown	96 (2.16)	5 (0.26)
Gross weekly household income		
≥Aus $3000 (US $2086.20)	635 (14.28)	211 (10.82)
Aus $1500 to Aus $2999 (US $1043.10 to US $2085.50)	1281 (28.8)	589 (30.21)
Aus $500 to Aus $1499 (US $347.70 to US $1042.40)	1646 (37.01)	793 (40.67)
<Aus $500 (US $347.70)	550 (12.37)	261 (13.38)
None	139 (3.13)	70 (3.59)
Negative income	34 (0.76)	26 (1.33)
Not stated or unknown	162 (3.64)	0 (0)
Other language spoken at home		
Yes	1036 (23.29)	438 (22.46)
No	3411 (76.69)	1513 (77.59)
Not stated or unknown	1 (0.02)	0 (0)
General health		
Excellent	549 (12.34)	236 (12.1)
Very good	1887 (42.42)	837 (42.92)
Good	1302 (29.27)	562 (28.82)
Fair	573 (12.88)	255 (13.08)

	Combined sample (n=4448), n (%)	Web-based panel (n=1950), n (%)
Poor	131 (2.95)	59 (3.03)
Not stated or unknown	6 (0.13)	0 (0)

Support for AI in General and in Specific Scenarios

Background

We first discuss questions focused on support for or opposition to AI. The CIs for questions B01 and C01 tended to be narrower as they were based on the combined sample. However, for all questions, estimates of percentages had margins of error (ie, twice the SE) of <3 percentage points, reflecting the relatively large sample size and the reliability of all estimates.

Respondents Expressed General Support for AI

Figure 1 and Table 3 show the level of support for the development of AI in general—an estimated 60.3% in the *strongly support* or *somewhat support* categories.

Although the estimate for the *support* categories was 60.3%, it was only 13.4% for the *opposed* categories and 26.3% for the *neutral* or *don't know* responses. The on-balance support mean score of 2.35 was statistically significant when tested against the midpoint of 3.00 (*P*<.001). The design effects are consistent with the design effect that was due to a weighting of 1.83.

Table 4 shows the percentage that selected a *support* category after *don't know* responses were excluded and also after don't know and neutral responses were excluded. This allowed for direct comparison of support and opposition and examination of whether there was majority support. We tested whether the resulting percentages were >50% using the adjusted Pearson F test for equal percentages in SPSS, where an estimate of 50% would indicate equal levels of support and opposition. Table 4 clearly demonstrates majority support among those taking a positive or negative position—63.1% when *don't know* responses were excluded and 81.8% when neutral and don't know responses were excluded, with *P* values indicating that both estimates were statistically significantly different from 50%.

For each question in the remaining analyses, the very small proportion of refused and *don't know* responses were not included and were no more than 8 cases for any of these questions.

Figure 1. Responses to question B01: How much do you support or oppose the development of artificial intelligence?

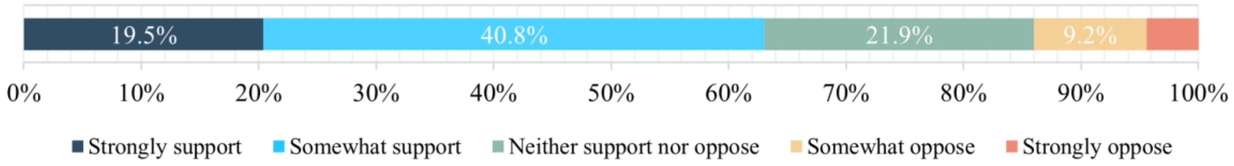


Table 3. Estimated percentages, mean, and 95% CIs for responses to question B01: How much do you support or oppose the development of artificial intelligence?[a,b]

	Estimated percentage (95% CI)	Design effect
Strongly support	19.5 (17.9-21.1)	1.87
Somewhat support	40.8 (38.9-42.8)	1.84
Neither support nor oppose	21.9 (20.3-23.5)	1.74
Somewhat oppose	9.2 (8.1-10.4)	1.87
Strongly oppose	4.2 (3.5-5.1)	1.76
I don't know	4.4 (3.6-5.3)	1.96

[a] Percentages and CIs adjusted for weighting.

[b] The mean score was 2.35 (95% CI 2.31-2.39) with a design effect of 1.83.

Table 4. Percentage of those who strongly support or somewhat support the development of artificial intelligence, 95% CIs, and *P* values for testing against 50%[a].

	Categories deleted	
	"Don't know"	"Don't know and neutral"
Estimated percentage support (95% CI)	63.1 (61.1-65)	81.8 (80-83.5)
P value[b]	<.001	<.001
Design effect	1.80	1.83

[a] Percentages and CIs adjusted for weighting.

[b] *P* value for adjusted Pearson *F* test for equal proportions in *support* and *oppose* categories.

Respondents Showed Less Support for Specific AI Use Scenarios and Supported Some Scenarios More Than Others

Figure 2 shows the estimates of the level of support for AI in specific health care and welfare scenarios, with scenarios presented in increasing order of level of support. Multimedia Appendix 3 shows the related estimates and 95% CIs. Table 5 presents estimates of support in categories 1 and 2 combined for specific scenarios, associated 95% CIs, and *P* values for the test against 50%. For all these specific scenarios, less support was expressed than in the question about AI in general (Figure 1).

Figure 2 shows that the strongest support was expressed for a learning health care system making diagnostic and treatment recommendations, where *over time, patients get different care depending on whether they do, or do not, share their health record with the AI system* (ie, people receive health benefits only at the expense of health data privacy). Overall, the support for this item was 42.3% (Table 5). Regarding social services, the highest level of support was for targeted compliance checking for welfare debt (38.9%). In this scenario, a government department used an algorithm to check groups deemed *high-risk* for welfare overpayment twice as often, which found more welfare debts, saved money, and reduced the number of checks on other people but meant people in high-risk groups were checked more even if they had not done anything wrong. The next highest support was for automated systems to identify parents who required assistance to return to work with limited contestability (34.9%) and employment support recommendation systems that were nonexplainable to employment service workers (31.2%). The least support overall was expressed for AI systems that led to physician deskilling (27% support and 48.3% opposition) and those that made diagnostic and treatment recommendations but were not explainable to physicians (29.1% support and 41.6% opposition).

For the estimates in Table 5, the neutral middle category with a score of 3 was included in the denominator. To directly compare the level of support and opposition and assess whether there was majority support or opposition, we removed the neutral category and recalculated the estimates and tests (Table 6). With the neutral score included, the level of support never reached a majority and ranged from 27% (deskilling physicians) to 42.3% (data sharing for quality care). Once the middle category was excluded, Table 6 shows that, for the nonneutral respondents, there were majorities supporting data sharing and targeted compliance checking; a balance on automated parent support without contestability; and a majority opposed to nonexplainable hospital algorithms, nonexplainable job services, and especially deskilling physicians.

Table 7 uses mean scores to indicate on-balance opposition or support—a score >3.00 indicates on-balance opposition, and a score <3.00 indicates on-balance support, along with *P* values for testing that the mean score was 3 (neither supportive nor opposed on balance). The means of general support for the development of AI were included for comparison. Marginal on-balance support was demonstrated for data sharing for quality care only (this should not be overinterpreted as the mean score was so close to neutral). For targeted compliance checking and noncontestable automated parent support, views were balanced. For both explainability scenarios and clinical deskilling, respondents expressed on-balance opposition at a statistically significant level.

Figure 2. Responses to questions C03 to C05 and D03 to D05: support for or opposition to specific scenarios. AI: artificial intelligence.

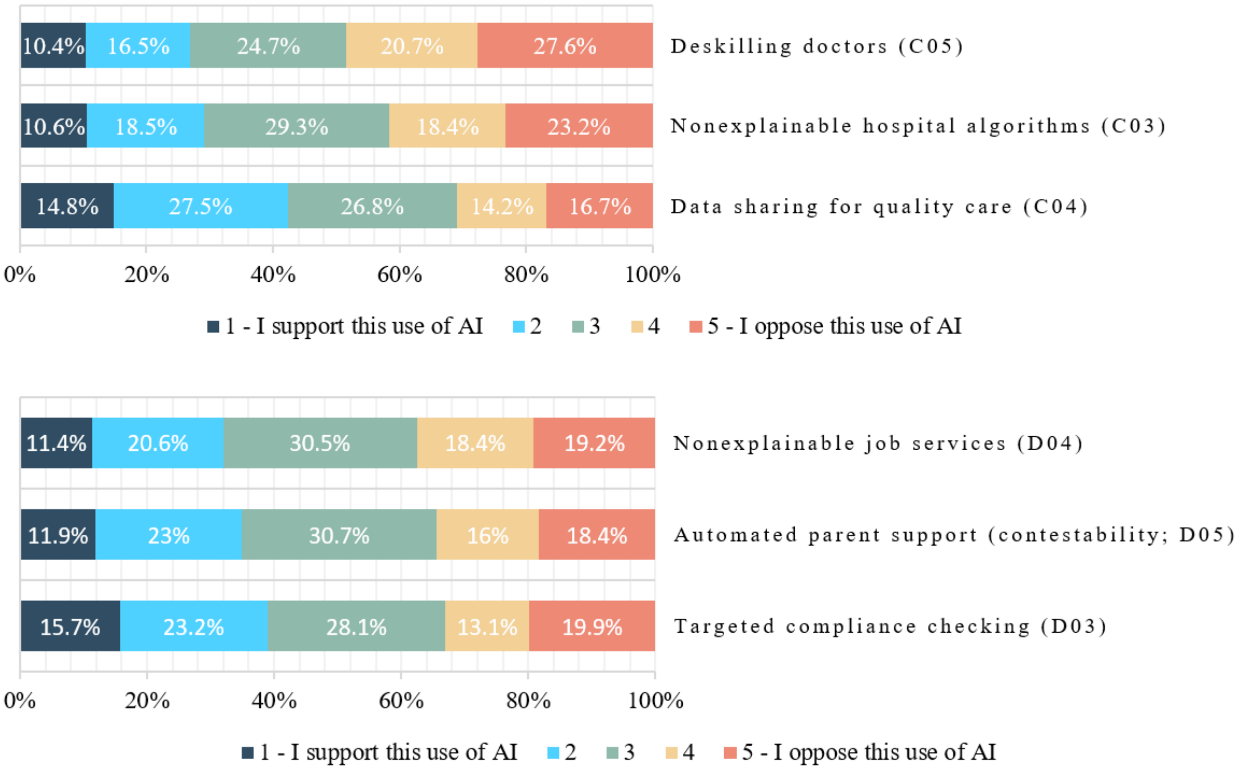


Table 5. Percentage of those supporting artificial intelligence in specific scenarios, 95% CIs, and *P* values for testing against 50%[a].

Domain and scenario	Estimated percentage in "support" or "strongly support" categories (95% CI)	*P* value[b]	Design effect
Health			
Data sharing for quality care (C04[c])	42.3 (39.6-45.1)	<.001	1.62
Nonexplainable hospital algorithms (C03)	29.1 (26.7-31.6)	<.001	1.57
Deskilling physicians (C05)	27 (24.6-29.5)	<.001	1.57
Welfare			
Targeted compliance checking (D03)	38.9 (36.2-41.7)	<.001	1.61
Automated parent support (contestability; D05)	34.9 (32.3-37.6)	<.001	1.59
Nonexplainable job services (D04)	31.2 (28.7-33.8)	<.001	1.56

[a] Percentages and CIs adjusted for weighting.

[b] *P* value for adjusted Pearson *F* test for 50% proportions in categories 1 and 2 combined.

[c] Code in parentheses (eg, C04) indicates question number in instrument.

Table 6. Proportion of respondents supporting artificial intelligence in specific scenarios, associated 95% CIs, and *P* values for testing against 50%; neutral responses deleted[a].

Domain and scenario	Estimated percentage in "support" or "strongly support" categories	*P* value[b]	Design effect
Health			
Data sharing for quality care (C04[c])	57.8 (54.5-61.1)	<.001	1.63
Nonexplainable hospital algorithms (C03)	41.1 (38-44.4)	<.001	1.58
Deskilling physicians (C05)	35.8 (32.8-38.9)	<.001	1.58
Welfare			
Targeted compliance checking (D03)	54.1 (50.9-57.4)	.01	1.58
Automated parent support (contestability; D05)	50.4 (47-53.7)	.82	1.62
Nonexplainable job services (D04)	44.1 (40.8-47.4)	<.001	1.59

[a] Percentages and CIs adjusted for weighting.

[b] *P* value for adjusted Pearson *F* test for 50% proportions in categories 1 and 2 combined.

[c] Code in parentheses (eg, C04) indicates question number in instrument.

Table 7. Analysis of mean support for use of artificial intelligence (AI) in specific scenarios, 95% CIs, and *P* values for testing against a mean of 3. A score <3 represents support, and a score of >3 represents opposition[a].

Domain and scenario	Estimated mean (95% CI)	*P* value[b]	Design effect
General—support for the development of AI (B01[c])	2.35 (2.31-2.39)	<.001	1.83
Health			
Data sharing for quality care (C04)	2.90 (2.83-2.98)	.01	1.65
Nonexplainable hospital algorithms (C03)	3.25 (3.18-3.32)	<.001	1.57
Deskilling physicians (C05)	3.39 (3.31-3.46)	<.001	1.62
Welfare			
Targeted compliance checking (D03)	2.98 (2.91-3.06)	.64	1.62
Automated parent support (contestability; D05)	3.06 (2.99-3.13)	.10	1.60
Nonexplainable job services (D04)	3.19 (3.12-3.26)	<.001	1.59

[a] Means and CIs adjusted for weighting.

[b] *P* value for *t* test that the mean score was 3.0 using complex samples.

[c] Code in parentheses (eg, B01) indicates question number in instrument.

Statistical Significance of Differences Between Support in General and in Specific Scenarios

To further investigate these results, we statistically tested changes in responses between the general question (B01) and the more specific scenario questions (C03-C05 and D03-D05). Table 8 shows the percentage of those who changed between question B01 and each of the more specific scenarios and, of those who changed, what percentage changed to a more negative attitude. The change was tested against 50%, which corresponded to an equal change in a positive and negative direction.

Table 8 shows that the estimated percentage that answered differently between the general and the more specific questions was between 60.2% and 70.6%. Of those who changed, between 70.8% and 83% changed to a more negative response, and all of these changes were statistically significant. There was also a slight increase of 3% to 9% in neutral responses across specific scenarios compared with the general question.

Table 8. Estimated percentage of those who changed their response between the general question on the development of artificial intelligence and the specific scenarios and, of those who changed, the percentage that had a more negative attitude in the specific scenarios, with 95% CIs and the *P* value for the test of equal change in each direction[a].

Domain and scenario	Percentage of those who changed	Percentage of those who changed becoming more negative (95% CI)	*P* value[b]	Design effect
Health				
Data sharing for quality care (C04[c])	60.2	70.8 (67.3-74)	<.001	1.59
Nonexplainable hospital algorithms (C03)	65.6	81.4 (78.6-83.9)	<.001	1.53
Deskilling physicians (C05)	70.6	83 (80.3-85.3)	<.001	1.56
Welfare				
Targeted compliance checking (D03)	63.8	71.9 (68.5-75)	<.001	1.65
Automated parent support (contestability; D05)	65	76.1 (73-78.9)	<.001	1.56
Nonexplainable job services (D04)	66.6	80.3 (77.5-82.9)	<.001	1.50

[a] Percentages and CIs adjusted for weighting.

[b] Adjusted Pearson *F* test for equal proportions changing in each direction.

[c] Code in parentheses (eg, C04) indicates question number in instrument.

Statistical Significance of Differences in Support Between Scenarios

To assess the statistical significance of differences in support for different detailed scenarios, Table 9 shows estimates of the percentage of those who changed in response to pairs of questions and, of those who changed, the percentage expressing a more negative attitude on the second question and the associated test against 50%. Although most comparisons were within the health care or welfare domain, we asked about explainability in both the health care and welfare contexts, allowing us to make direct comparisons between this pair of questions.

As noted, the health care and welfare question blocks were randomized per participant, and the questions were randomized within blocks. As shown in Table 9, respondents did make different judgments in specific scenarios—there were statistically significant changes within all pairs except between the questions regarding explainability in health care and in welfare. Despite 45.7% of people changing their responses between these 2 questions, people changed their minds in both directions in approximately equal proportions. This suggests divided views on the importance of explainability in different scenarios. The differences between all health care scenarios were statistically significant. Answers on nonexplainability and deskilling were significantly different, and most were more negative than those on data sharing; answers on deskilling were significantly different, and most were more negative than those on nonexplainability. In addition, most changed their responses between these questions in the same direction. A similar pattern was seen in the welfare scenarios—a significant proportion of respondents changed their response among targeted compliance checking, automated parent support without contestability, and nonexplainable job services, in all cases to a more negative response. Again, most tended to change their responses among these questions in the same direction.

Comparisons of the general support and support in specific scenarios and between the scenarios were also analyzed using differences in the means, with similar conclusions.

Table 9. Estimated proportion of those who changed their response between 2 scenarios and, of those who changed, the percentage that expressed a more negative attitude in the second question, with 95% CIs and the *P* value for the test of equal change in each direction[a].

Domain and scenarios compared	Percentage of those who changed	Percentage of those who changed becoming more negative (95% CI)	*P* value[b]	Design effect
Health				
C03[c] (explainability) vs C04[d] (data sharing)	38.1	26.7 (22.7-31.1)	<.001	1.77
C03 (explainability) vs C05[e] (deskilling)	43.6	59.2 (55-63.3)	<.001	1.62
C04 (data sharing) vs C05 (deskilling)	45.7	77.9 (74.2-81.2)	<.001	1.69
Welfare				
D03[f] (compliance checking) vs D04[g] (explainability)	41.7	64.2 (60-68.2)	<.001	1.60
D03 (compliance checking) vs D05[h] (contestability)	45.1	55.6 (51.4-59.6)	.008	1.59
D04 (explainability) vs D05 (contestability)	42.3	41.7 (37.6-45.9)	<.001	1.59
Explainability in health vs in welfare—C03 vs D04	45.7	46.1 (42-50.2)	.06	1.64

[a] Percentages and CIs adjusted for weighting.

[b] Adjusted Pearson *F* test for equal proportions changing in each direction.

[c] C03: nonexplainable hospital algorithms.

[d] C04: data sharing for quality care.

[e] C05: deskilling physicians.

[f] D03: targeted compliance checking.

[g] D04: nonexplainable job services.

[h] D05: automated parent support (contestability).

Which Attributes of Health Care and Social Service AIs Were Most Important?

We provided 2 health care scenarios (C01 [machine diagnosis and treatment recommendations] and C02 [machine triage]) and 2 social service scenarios (D01 [automation of unemployment benefit decision-making] and D02 [chatbot advice about carer payments]). We asked respondents to rate the importance of different attributes of the AI system in each one, where the attributes reflected a key ethical, legal, or social dimension of the AI or its use. For health care scenarios, these attributes included responsibility for decision-making as this is central to medicolegal frameworks and professional autonomy. For welfare scenarios, they included personal tailoring as this is a key promise of automation and machine decision-making in welfare contexts.

Figure 3 shows these responses to the health care and welfare scenarios to allow comparisons to be made between the distributions of the responses to any 2 questions assessing the same ethical or social dimension of AI. Multimedia Appendix 4 provides the detailed estimates of the proportions and the associated estimates of 95% CIs on estimated proportions for Figure 3.

Table 10 provides a summary of the importance that respondents ascribed to different attributes using mean scores, 95% CIs, and design effects. The response categories were scored from 1 for *extremely important* to 5 for *not at all important*; thus, lower scores indicate more importance. All means were <3, the midpoint of the scale; t tests against a mean of 3 were statistically significant with *P*<.001, indicating that more of the distribution of responses was in the extremely or very important categories. The attributes in Table 10 are in ascending order of means, that is, from most to least important (where the most important value is presented first).

As shown in Figure 3 and Table 10, there were distinctions between attributes. In all 4 scenarios, accuracy was rated as most important on average (1.49-1.61), and the ability of an AI system to reduce system costs was rated as least important (2.30-2.60), especially in health care. After accuracy, fairness was the second most important attribute in both social service scenarios (1.80 and 1.81) but, in the health care scenarios, it placed lower relative to other attributes (1.87 and 1.94). After accuracy, responsibility and human contact were the next most important in both health care scenarios. Speed was slightly more important in a health care triage scenario (1.90) than in a medical testing scenario (2.08).

Table 11 compares the mean responses to the attribute questions for the 2 health care scenarios (C01 vs C02) and the 2 welfare scenarios (D01 vs D02) to assess whether there were differences in importance in specific scenarios. In these comparisons, a negative estimate of the difference implies more importance for the first listed question, and a positive difference implies more importance for the second listed question. Table 12 provides further analysis, including statistical significance testing, of shifts in responses to the questions. Taken together, these tables show that, among the health care scenarios, the only statistically significant differences were in relation to speed (more important in triage) and reducing costs (more important in decision support). In the social service scenarios, more statistically significant differences were found, with explanation and cost reduction being more important in automating unemployment benefits and human contact, speed, and personal tailoring being more important in receiving automated carer support advice.

Figure 3. Responses to questions C01 to C02 versus D01 to D02: summary and comparison of health (C) and welfare (D) scenarios. Numerical estimates <10% are not given.

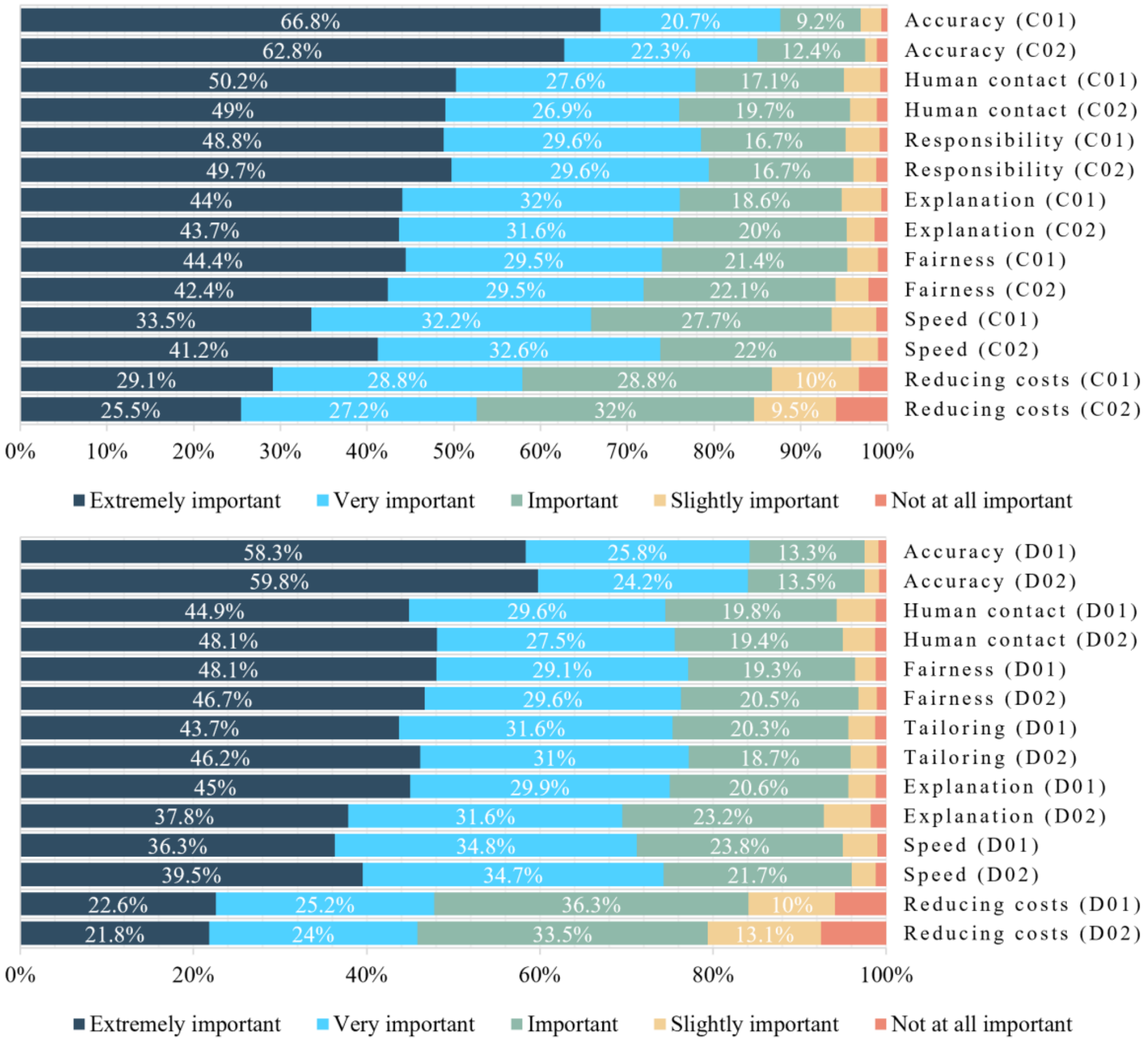


Table 10. Means, 95% CIs, and design effects for importance of values.

	Estimate of the mean[a] (95% CI)	Design effect
C01[b] —machine reads medical test, diagnoses, and recommends treatment		
Accuracy	1.49 (1.46-1.53)	1.98
Human contact	1.78 (1.74-1.81)	1.95
Responsibility	1.78 (1.75-1.82)	1.98
Explanation	1.86 (1.82-1.90)	1.96
Fairness	1.87 (1.83-1.91)	1.91
Speed	2.08 (2.04-2.12)	1.88
Reducing costs	2.30 (2.25-2.34)	1.92
C02—machine triages when you are unwell		
Accuracy	1.56 (1.51-1.61)	1.73
Responsibility	1.76 (1.71-1.81)	1.75
Human contact	1.81 (1.75-1.86)	1.72
Explanation	1.87 (1.82-1.93)	1.76
Speed	1.90 (1.85-1.95)	1.64
Fairness	1.94 (1.88-2.00)	1.81
Reducing costs	2.43 (2.36-2.50)	1.74
D01—machine processes application for unemployment benefits (data sharing required)		
Accuracy	1.61 (1.56-1.65)	1.53
Fairness	1.80 (1.75-1.85)	1.56
Explanation	1.86 (1.80-1.91)	1.61
Personal tailoring	1.87 (1.82-1.92)	1.58
Human contact	1.88 (1.82-1.93)	1.54
Speed	1.99 (1.93-2.04)	1.58
Reducing costs	2.51 (2.45-2.58)	1.59
D02—chatbot advises about carer payments		
Accuracy	1.60 (1.55-1.64)	1.6
Fairness	1.81 (1.76-1.87)	1.68
Personal tailoring	1.82 (1.77-1.87)	1.67
Human contact	1.83 (1.77-1.88)	1.63
Speed	1.91 (1.86-1.97)	1.71
Explanation	2.02 (1.96-2.08)	1.72
Reducing costs	2.60 (2.54-2.67)	1.71

[a] Means and CIs adjusted for weighting.

[b] Code (eg, C01) indicates question number in instrument.

Table 11. Differences in mean responses on importance of attributes between 2 scenarios[a].

Domain and attribute	Mean difference (95% CI)	*P* value[b]	Design effect
Health—C01[c] vs C02[d]			
Explanation	−0.001 (−0.048 to 0.046)	.96	1.89
Speed	0.082 (0.040 to 0.123)	<.001	1.51
Accuracy	−0.009 (−0.052 to 0.033)	.67	1.91
Human contact	−0.012 (−0.060 to 0.036)	.63	2.12
Responsibility	0.007 (−0.035 to 0.050)	.73	1.88
Reducing costs	−0.111 (−0.162 to −0.060)	<.001	1.99
Fairness	−0.035 (−0.081 to 0.011)	.13	1.93
Welfare—D01[e] vs D02[f]			
Explanation	−0.164 (−0.215 to −0.113)	<.001	1.64
Speed	0.070 (0.029 to 0.111)	<.001	1.59
Accuracy	0.012 (−0.023 to 0.048)	.50	1.42
Human contact	0.049 (0.009 to 0.089)	.02	1.48
Personal tailoring	0.048 (0.006 to 0.090)	.02	1.58
Reducing costs	−0.091 (−0.136 to −0.046)	<.001	1.54
Fairness	−0.018 (−0.059 to 0.029)	.38	1.72

[a] Means and CIs adjusted for weighting.

[b] *P* value for *t* test that the mean difference was 0 using complex samples.

[c] C01: machine reads medical test, diagnoses, and recommends treatment.

[d] C02: machine triages when you are unwell.

[e] D01: machine processes application for unemployment benefits (data sharing required).

[f] D02: chatbot advises about carer payments.

Table 12. Estimated percentages of those who changed their responses on importance of values between 2 scenarios and, of those, the percentage that ranked the value to be more important in the first question than in the second question (C01 vs C02 or D01 vs D02), with associated 95% CIs and the *P* value for the test of equal cell proportions[a].

Domain and values	Percentage of those who changed	Percentage ranking the value as more important in C01 (vs C02) or D01 (vs D02) (95% CI)	*P* value[b]	Design effect
Health—C01[c] vs C02[d]				
Explanation	34.3	47.6 (42.8-52.4)	.33	1.68
Speed	34.9	39.5 (35.2-44.1)	<.001	1.52
Accuracy	25.1	49.5 (43.8-55.2)	.86	1.68
Human contact	29.9	50.3 (45-55.5)	.92	1.70
Responsibility	28.3	47.7 (42.5-53)	.40	1.69
Reducing costs	33	59.2 (54.3-63.9)	<.001	1.66
Fairness	29.3	53.7 (48.5-58.8)	.16	1.66
Welfare—D01[e] vs D02[f]				
Explanation	39.6	63.7 (59.4-67.7)	<.001	1.55
Speed	32.7	41.8 (37-46.6)	.001	1.66
Accuracy	26.4	48.4 (43.2-53.7)	.56	1.57
Human contact	30.7	43.9 (39.1-48.8)	.02	1.64
Personal tailoring	33.1	43.9 (39.1-48.8)	.01	1.69
Reducing costs	35.1	58.8 (54.3-63.1)	<.001	1.58
Fairness	27.1	51.7 (46.3-57.1)	.53	1.70

[a] Percentages and CIs adjusted for weighting.

[b] Adjusted Pearson *F* test for equal proportions.

[c] C01: machine reads medical test, diagnoses, and recommends treatment.

[d] C02: machine triages when you are unwell.

[e] D01: machine processes application for unemployment benefits (data sharing required).

[f] D02: chatbot advises about carer payments.

Final Bundled Attribute Trade-off of AI and Human Attributes

Figure 4 shows the estimated percentages for the final bundled trade-off question (E01), where respondents were asked to weigh speed, convenience, and accuracy against human contact and discretion. Table 13 provides the estimated percentages, mean scores, and 95% CIs. These results show that human attributes were generally valued more, as indicated by a mean score >3. The estimated proportion of those who preferred the machine attributes (categories 1 or 2) was 20.3%, whereas, for human attributes (categories 4 or 5), it was 52%; 27.7% selected a middle position.

Figure 4. Responses to question E01: speed, accuracy, and convenience versus human contact and discretion.

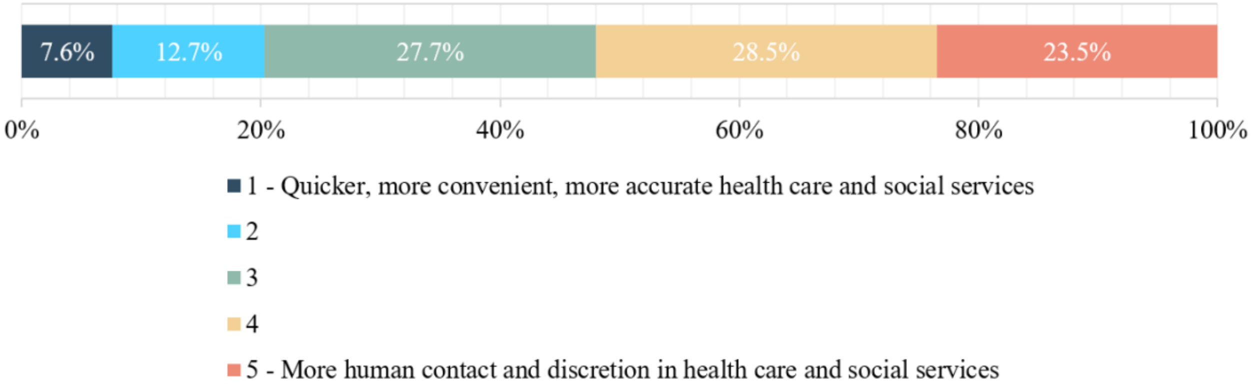


Table 13. Speed, accuracy, and convenience versus human contact and discretion; estimated percentages; and 95% CIs for responses to question E01[a].

	Estimate (95% CI)
1: speed, convenience, and accuracy	7.6 (6.2-9.1)
2	12.7 (11-14.7)
3	27.7 (25.3-30.3)
4	28.5 (26.1-31.1)
5: human contact and discretion	23.5 (21.2-26)
Mean score[b]	3.38 (3.41-3.54)

[a] Percentages and CIs adjusted for weighting.

[b] *P*<.001 for testing that the mean score was 3; design effect=1.602.

Discussion

Principal Findings

Overview

The AVA-AI study has created one of the first large, robust data sets reflecting public views on the potential use of AI in health care and social services, with particular attention to the ELSI of those technologies. Future studies will provide a greater breakdown of the variation in responses among different population subgroups. This analysis focused on answering 3 key questions: how judgments in general compare with judgments in particular, how judgments about use in health care compare with judgments about use in social services, and whether judgments differ when ELSI differ.

General Versus Particular Judgments About AI

Our first general question about support for or opposition to AI was taken from the 2018 survey of the American public by Zhang and Dafoe [26], which included 2000 respondents and used a similar weighting methodology; the Monash Data Futures survey [27] also included this question and surveyed 2019 respondents. Owing to our methodology, we asked this question to 4448 respondents. Table 14 compares these results—as the Monash survey reports combined *all support* and *all oppose* categories only, we have done the same. Both the AVA-AI study and the Monash survey suggest more positive general views in Australia than in the United States, although the results of the AVA-AI study are less positive than those of the Monash survey. Speculative reasons for this difference could include more prominent public discourse regarding harms from AI deployment in the US context (eg, in policing, justice, warfare, and the retail sector) or, more tentatively, that, in the 2 years between the surveys (mid-2018 for the study by Zhang and Dafoe [26] vs March 2020-April 2020 for both the AVA-AI study and the Monash survey), Australians may have had additional positive experiences of the everyday AI described in that question (eg, language translation, spam filters, and streaming content suggestions).

As a minority of AVA-AI study respondents began the survey with negative general views on AI and >60% expressed support, any negative judgments expressed seem likely to be a response to the details of the scenarios presented rather than reflect prejudice against or fear of AI in general. When asked about specific scenarios for AI use, respondents were consistently more negative—the reduction in support between the general question and all 6 specific scenarios was statistically significant, and support expressed in the specific scenarios dropped between 17 and 33 percentage points. The simple opening *support-or-oppose* question presented familiar, helpful everyday examples of AI in use and did not demonstrate any downsides of AI. In contrast, the detailed scenario questions were designed for balance. Each question emphasized that AI could both improve services (eg, make them quicker, more convenient, and more accurate) and have downsides (eg, reduced explainability, contestability, and privacy; unfair burdens on minorities; or human deskilling). On the basis of our findings, we hypothesize that members of the general public may remain broadly unaware of the potential downsides of AI in use and that some of these downsides (eg, deskilling) matter more to them than others (eg, privacy). We did not test the level of awareness of ELSI problems with AI—this is a potential direction for future research. Participants' more negative judgments in the case-specific questions also empirically reinforce what has already been argued in the literature: that the ELSI of AI applications need to be considered in the context of detailed cases.

Table 14. Comparison of findings from the studies by Zhang and Dafoe [26] and the Monash Data Futures Institute [27] and from the Australian Values and Attitudes on Artificial Intelligence (AVA-AI): How much do you support or oppose the development of artificial intelligence?

	Zhang and Dafoe [26] (2018), weighted %	Monash Data Futures Institute [27] (2020), weighted % by age only	AVA-AI (2020), weighted %
Strongly or somewhat support	40.94	62.4	60.3
Neither support nor oppose	27.84	23	21.9
Strongly or somewhat oppose	21.69	10.5	13.4
I don't know	9.54	4.1	4.4

Judgments About Health Care Versus Judgments About Social Services

Respondents had slightly stronger, more diverse, and more negative views on using AI in health care as opposed to in social services. This may be because they themselves have more direct experience of using health care or consider health care more relevant to them; alternatively, respondents may consider health care to be a higher-stakes service for which they are less tolerant of social or ethical wrongs or harms. Again, respondents in the AVA-AI study were less strongly supportive than respondents in the Monash survey, expressing 27% to 43% support for health care scenarios and 31% to 39% support for social service scenarios. In the Monash survey, respondents were asked to rate their support or opposition to *the application of artificial intelligence to social, humanitarian and environmental challenges*. The areas that received the most support—>75% of respondents—were *health* and *medicine*, whereas the areas that received the least support (although still >60%) included *equality and inclusion* and *public and social sector management*.

The differing responses to the 2 surveys may arise from the framing of the questions. The Monash questions were framed optimistically and presented no downsides; the AVA-AI questions presented both benefits and downsides or burdens. In health care, we held effectiveness and health benefits against requirements to share data, nonexplainability, and clinical deskilling. In social services, we held the accuracy and consistency of predictions and decisions against the potential for overtargeting, poor contestability, and nonexplainability. The differences in responses between the 2 surveys may show that the ethical and social risks of AI matter to people and will make a difference in their evaluations.

Do Judgments Differ When ELSI Differ?

The respondents clearly made judgments about the ELSI of AI. Although all ELSI were considered important, this was by degree. Respondents made quite finely graded judgments that intuitively aligned with the characteristics of the scenarios, suggesting both that they took the questions seriously and that different attributes will be differently important in different cases. For example, speed was more important in triage, where time is critical, than in diagnosis. Explanation was more important in automating unemployment benefits than in an information chatbot, which would be consistent with the view that people deserve to know why they do or do not receive payments. Human contact, personal tailoring, and speed were more important for the chatbot than for the benefits system, possibly reflecting that chatbot interactions are short and information-heavy and that people want a human to talk to if the automated system fails.

Two things were consistent: accuracy was always the most highly valued, and reducing costs was always the least highly valued across health care and social services. The lack of any significant difference in the importance of accuracy across scenarios suggests that this is an entry-level requirement for the use of AI (although defining *accuracy* in different contexts is not straightforward). The lower importance given to cost reduction may reflect a general rejection of instrumental decision-making in policy and of cost-based arguments in public services. Contextual factors include Australia's publicly funded health care system being *popular and entrenched* [53] and that, despite holding negative views on welfare recipients, the Australian public remains similarly supportive of the welfare system as a whole [54].

Fairness was more important in social services than in health care. This may reflect the centrality of the concept of procedural fairness—that is, the fairness of the decision-making process—in social service administration, particularly within Australia's bureaucratic and rule-bound welfare system [55]. It may also reflect heightened concern for issues of fairness in light of the public controversy surrounding the robodebt program, which centered on the legality, accuracy, and fairness of the program's debt calculations [23]. Perhaps the most deliverable promise of AI is increased speed, but this was not highly valued by respondents in any of the scenarios presented.

Knowing who is responsible for decisions, especially any mistakes made, was consistently important in health care, suggesting that the regulatory and ethical governance challenges in health care AI will matter to the public. Human contact was also important in health care. Prominent health care AI advocates have suggested that the core benefit of health care AI is its ability to release clinicians from mundane duties, freeing them to engage more deeply in care work [56]. However, the digitization of health care in some contexts has had the opposite effect, overburdening clinicians with data management and system requirements that alienate them from patient care [57]. This will be a key challenge to manage if health care AI is to deliver on its promises. Relatedly, respondents rejected medical deskilling most strongly among our 3 health care scenarios. This resonates with empirical research suggesting that people strongly value the preservation of human oversight for AI decision-making but also suggests the need for more work on what kinds of deskilling matter most as deskilling is highly likely to occur as automation increases. As in other research, participants were weakly supportive of sharing their health data with a learning health system if it delivered better quality care [58], although qualitative and deliberative research suggests that this support is likely to be conditional [59]. Respondents were also weakly supportive of algorithmic targeting of welfare compliance checking to high-risk groups if this saved money and reduced the number of checks on other people, which may reflect an on-balance judgment about proportionality or may simply reflect the aforementioned negative views on welfare recipients.

We asked about explainability in both health care and welfare scenarios and contestability in welfare scenarios. Respondents expressed an on-balance opposition to both health care and welfare AIs that were not explainable to relevant professionals. However, different respondents valued explainability differently in health care and welfare scenarios, suggesting that there may be some divergence in people's views on the domains in which explanation is more important. There was also an on-balance opposition to noncontestability in welfare scenarios, which reinforces support for processes of review and appeal when welfare decision-making is automated.

When asked to make an on-balance judgment about the *bundle* of attributes most commonly associated with machines versus with humans, respondents strongly preferred human attributes. Although they considered attributes such as accuracy to be important if an AI system was to be implemented, they still highly valued human support and connection and were not prepared to give them up in exchange for accuracy (despite the accuracy of AI being highly valued in itself). This suggests the importance of pursuing an augmentation rather than a replacement role for AI in both health care and social services. For all of these findings, further qualitative research is needed to better understand the reasons underpinning people's judgments.

Limitations

To the best of the authors' knowledge, this study is one of the largest and most robust surveys of public attitudes toward health care and welfare AI to date. The methodological approach taken allowed for the collection of detailed information on attitudes for a substantial sample using a relatively low-cost web-based panel while compensating for the potential biases in the creation of such a panel. Although the results suggest that respondents were able to engage with the details of the questions, the relatively low level of knowledge of AI in the community and the speculative nature of the questions mean that people's responses to a direct experience of AI may differ from their responses in this survey. A strength of our design was the use of questions that were deliberately structured to present both the potential benefits and the potential burdens or harms of AI while attempting to maintain neutral sentiment and avoid normative valence in the language used. The survey was conducted before the onset of the COVID-19 pandemic, which initiated the rapid digitization of many health care and social services; it is possible that responses would be different if the survey were repeated today.

Conclusions

Australians support the idea of AI in a general sense, but their support diminishes when considering the details of particular scenarios and the potential harms or burdens that may accompany any promised benefits. Respondents consistently rated the accuracy of performance as the most important attribute in an AI system, but only 1 in 5 valued the speed, accuracy, and convenience of AI systems more than continued human contact and discretion in service provision. Overall, this study suggests that the ethical and social dimensions of AI systems matter to Australians and that Australians want AI systems to augment rather than replace humans in the provision of both health care and social services and to reflect human values. Meaningful engagement and participation of ethicists, social scientists, and the public can highlight what harms and wrongs are most important to avoid in all stages of the development and implementation of AI, including in sensitive and value-laden domains such as health care and social services.

Acknowledgments

This research was funded by the University of Wollongong Global Challenges Program. The authors acknowledge Andrew Ward at the Social Research Centre for weighting the data and Kathleen Prokopovich, who assisted with project management and cognitive testing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey.
[DOCX File , 3112 KB - jmir_v24i8e37611_app1.docx]

Multimedia Appendix 2
Extended sample composition.
[DOCX File , 28 KB - jmir_v24i8e37611_app2.docx]

Multimedia Appendix 3
Support or opposition in specific artificial intelligence scenarios.
[DOCX File , 19 KB - jmir_v24i8e37611_app3.docx]

Multimedia Appendix 4
Importance of health (C) and welfare (D) scenarios.
[DOCX File , 26 KB - jmir_v24i8e37611_app4.docx]

References

1. Walsh T, Levy N, Bell G, Elliott A, Maclaurin J, Mareels I, et al. The effective and ethical development of artificial intelligence: an opportunity to improve our wellbeing. Australian Council of Learned Academies. 2019 Jul 30. URL: https://acola.org/hs4-artificial-intelligence-australia/ [accessed 2022-08-10]

2. Alston P. Report of the special rapporteur on extreme poverty and human rights. United Nations Digital Library. 2015. URL: https://digitallibrary.un.org/record/798707?ln=en [accessed 2022-08-10]
3. Carney T. Artificial intelligence in welfare: striking the vulnerability balance? Monash University Law Rev 2021 Mar 15;46(2). [doi: 10.2139/ssrn.3805329]
4. 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]
5. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. PLoS One 2019;14(2):e0212356 [FREE Full text] [doi: 10.1371/journal.pone.0212356] [Medline: 30779785]
6. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:1-12 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
7. Montreal Declaration for a Responsible AI. Université de Montréal. URL: https://recherche.umontreal.ca/english/strategic-initiatives/montreal-declaration-for-a-responsible-ai/ [accessed 2022-08-10]
8. Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos Trans A Math Phys Eng Sci 2018 Oct 15;376(2133):20180080 [FREE Full text] [doi: 10.1098/rsta.2018.0080] [Medline: 30322996]
9. Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M. Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. SSRN J 2020 Jan 15. [doi: 10.2139/ssrn.3518482]
10. Independent high-level expert group on artificial intelligence. European Commission. 2019 Apr 8. URL: https://www.aepd.es/sites/default/files/2019-12/ai-definition.pdf [accessed 2022-07-28]
11. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harvard Data Sci Rev 2019 Jun 23;1(1). [doi: 10.1162/99608f92.8cd550d1]
12. Crawford K. Atlas of AI. New Haven, Connecticut: Yale University Press; 2022.
13. Vincent J. Google is poisoning its reputation with AI researchers. The Verge. 2021 Apr 13. URL: https://www.theverge.com/2021/4/13/22370158/google-ai-ethics-timnit-gebru-margaret-mitchell-firing-reputation [accessed 2021-10-10]
14. Davis RB. The principlism debate: a critical overview. J Med Philos 1995 Feb;20(1):85-105. [doi: 10.1093/jmp/20.1.85] [Medline: 7738461]
15. Whittlestone J, Nyrup R, Alexandrova A, Dihal K, Cave S. Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research. London: Nuffield Foundation; 2019.
16. Morley J, Machado CC, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. Soc Sci Med 2020 Sep;260:1-14. [doi: 10.1016/j.socscimed.2020.113172] [Medline: 32702587]
17. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. Breast 2020 Feb;49:25-32 [FREE Full text] [doi: 10.1016/j.breast.2019.10.001] [Medline: 31677530]
18. 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):1-17 [FREE Full text] [doi: 10.1186/s12910-021-00577-8] [Medline: 33588803]
19. Goirand M, Austin E, Clay-Williams R. Implementing ethics in healthcare AI-based applications: a scoping review. Sci Eng Ethics 2021 Sep 03;27(5):1-53. [doi: 10.1007/s11948-021-00336-3] [Medline: 34480239]
20. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
21. Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: Picador, St Martin’s Press; 2018.
22. Dencik L, Kaun A. Datafication and the welfare state. Global Perspect 2020;1(1):12912. [doi: 10.1525/gp.2020.12912]
23. Carney T. The new digital future for welfare: debts without legal proofs or moral authority. UNSW Law J Forum 2018 Apr 4:1-16.
24. Ng YF. Institutional adaptation and the administrative state. Melbourne University Law Rev 2021;44(3):889-927.
25. Langford M. Taming the digital leviathan: automated decision-making and international human rights. AJIL Unbound 2020 Apr 27;114:141-146. [doi: 10.1017/aju.2020.31]
26. Zhang B, Dafoe A. Artificial intelligence: American attitudes and trends. SSRN J 2019. [doi: 10.2139/ssrn.3312874]
27. Selwyn N, Cordoba B, Andrejevic M, Campbell L. AI for social good? Australian public attitudes toward AI and society. Monash University. 2020 Aug. URL: https://www.monash.edu/data-futures-institute/news/ai-for-social-good-australian-public-attitudes-toward-ai-and-society [accessed 2022-07-28]
28. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, Utikal JS, Hauschild A, Schadendorf D, et al. Artificial intelligence in skin cancer diagnostics: the patients' perspective. Front Med (Lausanne) 2020;7:1-9 [FREE Full text] [doi: 10.3389/fmed.2020.00233] [Medline: 32671078]
29. Palmisciano P, Jamjoom AA, Taylor D, Stoyanov D, Marcus HJ. Attitudes of patients and their relatives toward artificial intelligence in neurosurgery. World Neurosurg 2020 Jun;138:627-633. [doi: 10.1016/j.wneu.2020.03.029] [Medline: 32179185]

30. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 2020 May 01;156(5):501-512 [FREE Full text] [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
31. Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, MacIsaac RJ, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep 2018 Mar 12;8(1):1-6 [FREE Full text] [doi: 10.1038/s41598-018-22612-2] [Medline: 29531299]
32. Yang K, Zeng Z, Peng H, Jiang Y. Attitudes of Chinese cancer patients toward the clinical use of artificial intelligence. Patient Preference Adherence 2019 Nov 1;13:1867-1875. [doi: 10.2147/ppa.s225952]
33. Schou J, Pors AS. Digital by default? A qualitative study of exclusion in digitalised welfare. Soc Policy Admin 2018 Nov 22;53(3):464-477. [doi: 10.1111/spol.12470]
34. Petersen AC, Christensen LR, Harper R, Hildebrandt T. "We would never write that down": classifications of unemployed and data challenges for AI. Proc ACM Human Comput Interact 2021 Apr 13;5(CSCW1):1-26. [doi: 10.1145/3449176]
35. Raso J. Displacement as regulation: new regulatory technologies and front-line decision-making in Ontario works. Can J Law Soc 2017 Jun 27;32(01):75-95. [doi: 10.1017/cls.2017.6]
36. Zejnilovic L, Lavado S, de Troya M, Sim S, Bell A. Algorithmic long-term unemployment risk assessment in use: counselors' perceptions and use practices. Global Perspective 2020 Jun 22;1(1):1-14. [doi: 10.1525/gp.2020.12908]
37. Karusala N, Wilson J, Vayanos P, Rice E. Street-level realities of data practices in homeless services provision. Proc ACM Human Comput Interact 2019 Nov 07;3(CSCW):1-23. [doi: 10.1145/3359286]
38. Ammitzbøll Flügge A, Hildebrandt T, Møller NH. Street-level algorithms and AI in bureaucratic decision-making. Proc ACM Human Comput Interact 2021 Apr 13;5(CSCW1):1-23. [doi: 10.1145/3449114]
39. Liu K, Tao D. The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. Comput Human Behav 2022 Feb;127:1-11. [doi: 10.1016/j.chb.2021.107026]
40. Kaczmirek L, Phillips B, Pennay D, Lavrakas P, Neiger D. Building a probability-based online panel: life in Australia. CSRM Methods Series 2019(2/2019):1-47.
41. Deville J, Särndal C. Calibration estimators in survey sampling. J Am Stat Assoc 1992 Jun;87(418):376-382. [doi: 10.1080/01621459.1992.10475217]
42. Baker R, Blumberg S, Brick J, Couper M, Courtright M, Dennis M, et al. Report on Online Panels. The American Association for Public Opinion Research. 2010 Jun. URL: https://www.aapor.org/Education-Resources/Reports/Report-on-Online-Panels.aspx [accessed 2022-07-28]
43. DiSogra C, Cobb C, Chan E, Dennis J. Calibrating non-probability internet samples with probability samples using early adopter characteristics. In: Proceedings of the American Statistical Association, Section on Survey Research Joint Statistical Meetings (JSM). 2011 Presented at: American Statistical Association, Section on Survey Research Joint Statistical Meetings (JSM); Jul 30 - Aug 4, 2011; Miami Beach, Florida.
44. Schonlau M, van Soest AH, Kapteyn A. Are 'webographic' or attitudinal questions useful for adjusting estimates from web surveys using propensity scoring? SSRN J 2007 Jun:1-18. [doi: 10.2139/ssrn.1006108]
45. Valliant R, Dever J. Survey Weights A Step-by-step Guide to Calculation. College Station, Texas: Stata Press; 2018.
46. Valliant R, Dever JA. Estimating propensity adjustments for volunteer web surveys. Sociol Method Res 2011 Jan 11;40(1):105-137. [doi: 10.1177/0049124110392533]
47. Elliott MR, Valliant R. Inference for nonprobability samples. Statist Sci 2017 May 1;32(2):249-264. [doi: 10.1214/16-STS598]
48. Kish L. Weighting for unequal Pi. J Official Stat 1992;8(2):183-200.
49. Kish L. Survey Sampling. Hoboken, New Jersey, United States: Wiley; 1965.
50. Kish L. Methods for design effects. J Official Stat 1995;11(1):55-77. [doi: 10.4135/9780857020116.n132]
51. Rao JN, Scott AJ. On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. Ann Statist 1984 Mar;12(1):46-60. [doi: 10.1214/aos/1176346391]
52. Agresti A. Categorical Data Analysis. Hoboken, New Jersey, United States: John Wiley & Sons; 2003.
53. Boxall A. Medicare: the making and consolidation of an Australian institution. In: Successful Public Policy: Lessons from Australia and New Zealand. Canberra, Australia: ANU Press; 2019.
54. Schofield TP, Butterworth P. Patterns of welfare attitudes in the Australian population. PLoS One 2015 Nov 10;10(11):1-14 [FREE Full text] [doi: 10.1371/journal.pone.0142792] [Medline: 26554361]
55. Adler M, Henman P. e-justice: a comparative study of computerization and procedural justice in social security. Int Rev Law Comput Technol 2010 Jul 21;15(2):195-212. [doi: 10.1080/13600860120070510]
56. Topol E. Deep Medicine How Artificial Intelligence Can Make Healthcare Human Again. New York: Basic Books; 2019.
57. Wachter R. The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age. New York: McGraw-Hill Education; 2015.
58. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):1-24 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]

59. Street J, Fabrianesi B, Adams C, Flack F, Smith M, Carter SM, et al. Sharing administrative health data with private industry: a report on two citizens' juries. Health Expect 2021 Aug;24(4):1337-1348 [FREE Full text] [doi: 10.1111/hex.13268] [Medline: 34048624]

Abbreviations

AI: artificial intelligence
AVA-AI: Australian Values and Attitudes on Artificial Intelligence
ELSI: ethical, legal, and social implications
LIA: Life in Australia

Edited by G Eysenbach; submitted 28.02.22; peer-reviewed by L Weinert, FM Calisto; comments to author 08.04.22; revised version received 25.05.22; accepted 19.07.22; published 22.08.22.

Please cite as:
Isbanner S, O'Shaughnessy P, Steel D, Wilcock S, Carter S
The Adoption of Artificial Intelligence in Health Care and Social Services in Australia: Findings From a Methodologically Innovative National Survey of Values and Attitudes (the AVA-AI Study)
J Med Internet Res 2022;24(8):e37611
URL: https://www.jmir.org/2022/8/e37611
doi:10.2196/37611
PMID:35994331

Original Paper

Sociotechnical Intervention for Improved Delivery of Preventive Cardiovascular Care to Rural Communities: Participatory Design Approach

Michelle Partogi[1*], MS; Simon Gaviria-Valencia[1*], MD; Mateo Alzate Aguirre[1], MD; Nancy J Pick[1], BA; Huzefa M Bhopalwala[1], MBBS; Barbara A Barry[1], PhD; Vinod C Kaggal[1], MS; Christopher G Scott[1], MS; Maya E Kessler[1], MPH, MD; Matthew M Moore[1], MFA; Jay D Mitchell[1], MD; Rajeev Chaudhry[1], MBBS, MPH; Robert P Bonacci[1], MD; Adelaide M Arruda-Olson[1], MD, PhD

Mayo Clinic, Rochester, MN, United States

*these authors contributed equally

Corresponding Author:

Adelaide M Arruda-Olson, MD, PhD
Mayo Clinic
200 First Street SW, Rochester, MN, 55905
Rochester, MN,
United States
Phone: 1 507 266 1376
Email: arrudaolson.adelaide@mayo.edu

Abstract

Background: Clinical practice guidelines recommend antiplatelet and statin therapies as well as blood pressure control and tobacco cessation for secondary prevention in patients with established atherosclerotic cardiovascular diseases (ASCVDs). However, these strategies for risk modification are underused, especially in rural communities. Moreover, resources to support the delivery of preventive care to rural patients are fewer than those for their urban counterparts. Transformative interventions for the delivery of tailored preventive cardiovascular care to rural patients are needed.

Objective: A multidisciplinary team developed a rural-specific, team-based model of care intervention assisted by clinical decision support (CDS) technology using participatory design in a sociotechnical conceptual framework. The model of care intervention included redesigned workflows and a novel CDS technology for the coordination and delivery of guideline recommendations by primary care teams in a rural clinic.

Methods: The design of the model of care intervention comprised 3 phases: problem identification, experimentation, and testing. Input from team members (n=35) required 150 hours, including observations of clinical encounters, provider workshops, and interviews with patients and health care professionals. The intervention was prototyped, iteratively refined, and tested with user feedback. In a 3-month pilot trial, 369 patients with ASCVDs were randomized into the control or intervention arm.

Results: New workflows and a novel CDS tool were created to identify patients with ASCVDs who had gaps in preventive care and assign the right care team member for delivery of tailored recommendations. During the pilot, the intervention prototype was iteratively refined and tested. The pilot demonstrated feasibility for successful implementation of the sociotechnical intervention as the proportion of patients who had encounters with advanced practice providers (nurse practitioners and physician assistants), pharmacists, or tobacco cessation coaches for the delivery of guideline recommendations in the intervention arm was greater than that in the control arm.

Conclusions: Participatory design and a sociotechnical conceptual framework enabled the development of a rural-specific, team-based model of care intervention assisted by CDS technology for the transformation of preventive health care delivery for ASCVDs.

(J Med Internet Res 2022;24(8):e27333) doi:10.2196/27333

KEYWORDS

sociotechnical; secondary prevention; atherosclerotic cardiovascular diseases; community health; rural health; participatory design; team-based care

Introduction

Background

Atherosclerotic cardiovascular diseases (ASCVDs) are the leading cause of morbidity and mortality in the United States and exemplify the national urban-rural health disparity [1,2]. Rural populations, which comprise 20% of the US population, have a 40% higher absolute prevalence of ASCVDs than urban dwellers [2]. According to the American Heart Association (AHA), rural residents also have higher rates of uncontrolled cardiovascular risk factors than their urban counterparts [2]. These risk factors include tobacco use, hypertension, and high cholesterol [3-6]. For patients with established ASCVDs, adherence to risk modification strategies prevents adverse events, improves survival, reduces the need for revascularization procedures, and enhances life quality [7]. However, strategies for risk modification are underused by patients with ASCVDs, especially in rural communities [2,8].

Clinical practice guidelines from the AHA and the American College of Cardiology for patients with ASCVDs recommend risk modification strategies, including antiplatelet and statin therapies, blood pressure control, and cessation of the use of tobacco products for secondary prevention in patients with established ASCVDs [7,9-11]. These recommendations are collectively referred to as cardiovascular guideline recommendations *(V4)* and are also endorsed by the “Million Hearts” initiative from the Centers for Disease Control and Prevention [12]. The *V4* recommendations have been designated as Class I, which indicates that the supporting data are strong and treatment is useful and effective and should be administered to most patients under most circumstances [9,13]. The level of evidence that supports these guideline-endorsed recommendations is also considered to be of the highest quality (level of evidence designation “A”) as it is derived from multiple randomized controlled trials [9,14,15].

Objectives

Multiple health care system factors affect the appropriate delivery of cardiovascular risk modification strategies to rural residents [2]. One factor is the inadequate number of physician providers in rural communities, as documented by the World Health Organization, the AHA, and the American Stroke Association [2,16]. A presidential advisory document from the AHA and the American Stroke Association has suggested that new and sustainable rural-specific and team-based care models assisted by technology may be a solution to improve the delivery of care in rural communities [2]. The question of this study was what are the characteristics of a rural-specific, team-based model for the delivery of care assisted by technology that is feasible in “real-world” rural clinics? The study goal was to develop and evaluate the feasibility of a new team-based model for rural practices with the following 2 components: a care model (the socio component) and a technological component (the clinical decision support [CDS] system).

It has been proposed that team-based care involves collaboration between physicians and nonphysician health professionals for the delivery of care instead of the traditional model in which care is delivered by physicians only [17]. A previous study showed that pharmacists working in collaboration with other health professionals in a team-based model improved cardiovascular health [18]. A second study demonstrated that a team-based delivery model using both physicians and advanced practice providers delivered outpatient cardiovascular care of a similar quality compared with a physician-only model [17,19].

For the sustainable adoption of new models for care delivery, it is fundamental that intended users are involved in the design process to ensure integration within redesigned user workflows [20,21]. Previous studies have demonstrated that well-executed participatory design processes support the implementation of health interventions [22-29]. According to Carrol and Rosson [30], participatory design advocates that users be included in the design process, and their input will increase the likelihood of successful design. According to Clemensen et al [31], the main feature of this design approach is the participation of users who work with researchers to produce new technology systems that can be understood and managed in practice.

Previously, in underserved rural settings, participatory design has informed strategies for the development of scalable systems such as mobile technology to disseminate health information for reproductive and child health services [32] and electronic, tablet-based community assessment tools for food and physical activity assessment [33]. When conducted under the sociotechnical theory framework, participatory design promotes the adoption of health care IT systems, including CDS [21,34]. The sociotechnical systems theory encourages the joint design of both the social and technical elements of a system [35]. A purely technocentric approach to system design may be unable to address the complex relationships between human and social factors and technology within the organizational context [36]. Therefore, in this study, participatory design under the sociotechnical system theory framework was used to design a new rural-specific and team-based care model for the coordination and delivery of secondary prevention to patients with ASCVDs assisted by CDS technology.

Methods

Setting and Context

The Office of Management and Budget defines rural counties as those with an urban core of ≥10,000 to <50,000 people [2]. By this definition, Austin, an urban core of 25,000 residents located in Mower County, Minnesota (MN), was identified as the site for the development of the rural-based model of care. Austin is the only urban core area in Mower County. The outpatient primary care clinic located in Austin where this study was conducted is part of the Mayo Clinic Health System (MCHS). The MCHS is a network of community-based health care professionals in primary care clinics located in >60 communities in MN, Iowa, and Wisconsin. Within this care network, patients receive primary care in their own communities. These clinics use the Mayo electronic health record (EHR) with digital medical data stored in a common centralized data warehouse that enables the deployment of CDS populated by EHR data for use in the rural clinics of the MCHS. Austin is located within driving distance (42 miles) of Rochester, MN.

The research team drove to Austin or connected remotely with Austin providers and patients during this study. The design, IT, and clinical informatics teams were from Rochester, whereas the rural primary care teams that participated in the study worked in the MCHS Austin. Primary care providers in the primary care teams included physicians and advanced practice providers such as nurse practitioners (NPs) and physician assistants. Primary care nursing supported the day-to-day work of primary care providers within a team and included registered nurses (RNs) and licensed practical nurses (LPNs). The expanded primary care team supported multiple primary care teams and included pharmacists, tobacco cessation coaches, and other teams of nurses such as care coordinators and complex disease coordinators. Importantly, all teams collaborated in the design of a rural-specific and team-based model for the delivery of care from June 2019 to December 2020.

Ethics Approval

This project was approved by the Mayo Clinic Institutional Review Board (approval numbers 19-011925 and 20-001192).

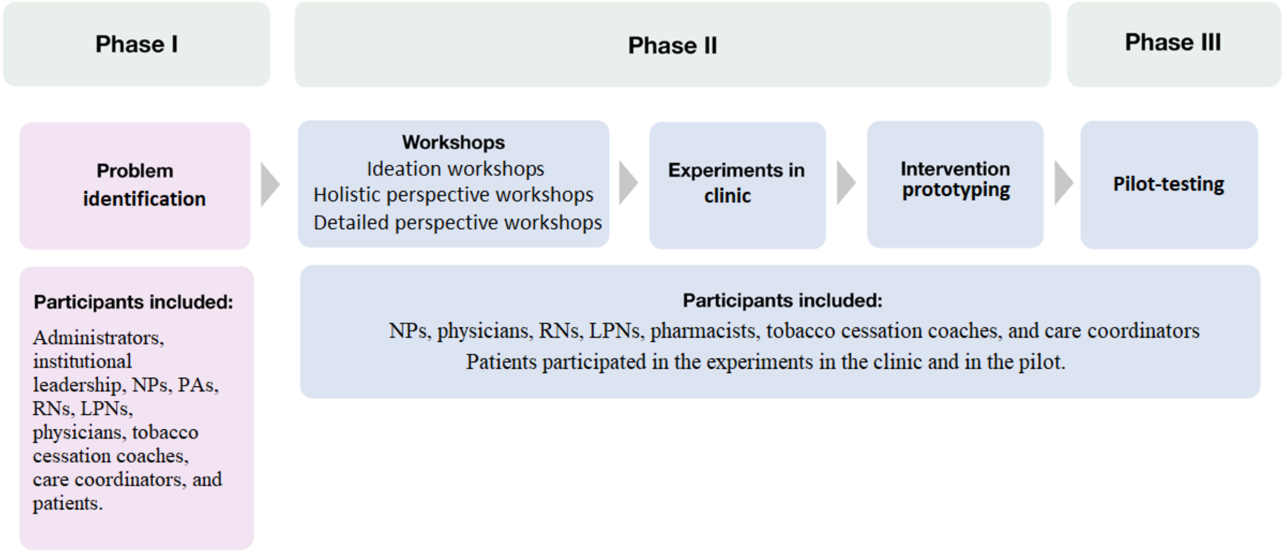
Project Phases

Overview

The project had three distinct phases: (1) problem identification, (2) experimentation, and (3) testing (Figure 1). The first aim of the activities in phase I was to gather information about the current status of ASCVD secondary prevention management. The second aim of this phase was to discuss ways of improving delivery of care to these patients. A cohesive group of participants was established representing different stakeholders (Figure 1). In phase II, the aim was to gain insights into possible solutions for the identified problems, creating and testing prototypes in experiments conducted in clinical practice. In phase III, the intervention prototype was tested and iteratively refined with user feedback in a 3-month pilot trial.

In this study, the participatory design activities of telling, making, and acting were conducted in iterative plan-action-observe-reflect cycles [31,37]. The provider workflows and CDS technology components of the sociotechnical model were developed using design activities and plan-action-observe-reflect cycles. The MCHS Austin leadership identified and recruited a team of local health care professionals to participate in this study. Patients who underwent medical visits in the MCHS Austin were recruited for interviews and observations. For the pilot trial, the patient cohort was identified electronically via a Cohort Knowledge Solution platform, and the patients were randomized into the control or intervention arm.

Figure 1. Project phased-design approach. LPN: licensed practical nurse; NP: nurse practitioner; PA: physician assistant; RN: registered nurse.



Phase I—Problem Identification

Phase I had 2 components: problem identification and ideation workshops.

Problem Identification

- Overview: The purpose of this phase was to gather data about the intervention context, including the setting, barriers, and facilitators of integration of the proposed technology solution into the rural clinic [31,37,38]. This phase focused on telling design activities. Applied ethnographic methods of participant observation and interviews comprised the telling activities [31,37], conducted during 75 hours of workflow observation in the clinic and 18 semistructured interviews with patients and health care professionals. Purposive sampling was used to identify interview participants with relevant roles and expertise, including administrators; institutional leadership; and various clinical health care professionals, including pharmacists, RNs, NPs, tobacco cessation coaches, and physicians. In this phase, 8 patients (mean age 77, SD 10 years; n=6, 75% women and n=2, 25% men) were also interviewed after primary care visits. Research team members (MEK, MP, and NP) performed rapid content analysis on observation field notes and verbatim interview transcriptions to identify key concepts and themes related to user needs for a sociotechnical-driven integration of the

CDS technology [39]. The results of the analysis were synthesized to inform making and prototyping activities.

- Output from problem identification: The insights from problem identification that informed the design of the ideation workshops are outlined in Textbox 1.

Textbox 1. Insights from the problem identification component.

Insights from patients and providers

- For many patients, especially those not meeting cardiovascular guideline (*V4*) recommendations, cardiovascular care was a low priority and, consequently, not addressed during office visits. Patients cited frequent provider turnover as a barrier, expressing uncertainty about who was responsible for managing cardiovascular health. The following are examples from patients:
  - “It is hard to get an appointment.”
  - “It isn’t like you can call in and see your doctor when you don’t feel right.”
  - “I find it extremely difficult to see anyone.”
  - “Our one thing here is, getting a doctor and keeping a doctor.”
- The cost of medications was not considered a barrier, but the cost of visits was a frequent concern. The following are examples from patients:
  - “I have always had really good medical insurance that help cover the cost, it’s never been a problem.”
  - “The cost of all this is just astronomical.”
- Patients who had previous intolerance for medications prescribed for cardiovascular prevention were reluctant to try another medication, especially in the absence of a relationship with a trusted provider. The following are examples from providers:
  - “Patients with elevated LFTs so a barrier to statin therapies and using it comfortably in those types of patients.”
  - “People who are not on statin who are have a cardiovascular event usually have problems with tolerating statins before so then it’s just sort of going down the statin.”
- Rural social networks are tightly knit, indicating that health care would ideally be delivered by local professionals. The following are examples from providers and patients:
  - “You form that connection and they look to you, a familiar face” (provider).
  - “People feel like they can trust us, that it’s a well-established practice” (provider).
  - “After we meet them for the first time, we develop relationships, they can see we can help” (provider).
  - “They become like family” (provider).
  - “They rely on you” (provider).
  - “Just knowing that you are going to be with them on their journey. They feel better about that” (provider).
  - “It depends on how I feel about the person. If I trust them” (patient).
  - “Feeling like you have someone’s undivided attention” (patient).
  - “I think they want to work with us” (patient).

Ideation Workshops

The ideation workshop structure was as follows:

- Overview: The purpose of the ideation workshops was to discuss future ways of organizing the delivery of care for patients with ASCVDs. In total, 2 ideation workshops were conducted. Each workshop drew health care professionals from various roles, including 5 RNs, 4 NPs, 3 LPNs, 2 pharmacists, 3 physicians, 1 tobacco cessation coach, and 1 care coordinator.
- Output from the ideation workshops: The insights from the ideation workshops that informed the design of the phase II experimentation workshops are outlined in Textbox 2.

Textbox 2. Insights from the ideation workshops.

Insights from providers

- Nonphysician care team members were motivated to collaboratively deliver preventive cardiovascular care to patients. However, there were no dedicated workflows and tools to support such initiatives. The following are examples from providers:
    - "I don't know sorting it out by their blood pressures, whether they're elevated, just being on medication or not, obviously smokers, are they on medications? What meds?"
    - "Being able to distinguish the groups like that might be helpful for us in determining where our resources should go."
- Preventive cardiovascular care should be proactively and intentionally delivered. The following are examples from providers:
    - "We don't get referrals like we used to."
    - "They probably need a visit with the provider."
    - "Some of them definitely should have been seen by a provider just because of the length of time they've been seen."
    - "They're not able to take that medication is there something else we can find for them."
- Improving the delivery of preventive cardiovascular care on a systematic level cannot be regarded as a low priority. The following are examples from providers:
    - "They just get through if they've been in the hospital."
    - "Identifying those patients that need more care and making sure that they are getting scheduled every so often, just for checking in, so that its keeping them out of the ED and out of the hospital."
- Existing personnel should be dedicated to the intentional delivery of preventive cardiovascular health care. The following are examples from providers:
    - "Regular appointments rather than waiting for something to happen...and they're probably going to need more time then we can give in fifteen to thirty minute appointments."
    - "Get back into that role of relationship building and connecting with people and then from there we can then take the next step."

Phase II—Experimentation

This phase had 3 components: prototyping workshops (including holistic and detailed perspective workshops), experiments in the clinic, and intervention prototyping.

Prototyping Workshops

The prototyping workshop structure was as follows:

- Overview: The purpose of the prototyping workshops was to design components of the novel model for the delivery of preventive cardiovascular health care for rural communities. The prototyping workshops involved making design activities. In making activities, user workshops were conducted to generate ideas to address issues identified during the telling activities and tailor the intervention to the needs of users and the context of the rural clinic. The multidisciplinary workshop methodology proposed by Scandurra et al [28] was used for the workshops. Prototyping workshops covered holistic and detailed perspectives from the different types of professionals on the care team [28].
- Holistic perspective workshops: There were 2 multidisciplinary interprofessional prototyping workshops in phase II. These workshops covered strategies for co-operation between different professionals [28] and included 4 RNs, 4 NPs, 3 LPNs, 2 pharmacists, 1 tobacco cessation coach, 1 RN care coordinator, and 3 physicians. In the first of these workshops, the staff suggested possible experiments. In the second workshop, participants selected the experiments and built on the proposed experiments in an iterative process. In these workshops, pamphlets summarizing insights and ideas were the discussion-inducing artifacts to facilitate collaborative and iterative idea generation [40].
- Detailed perspective workshops: These workshops included 1 health care professional work category each and focused on details of current and future professional workflows with discussion, feedback, and usability tests. There were workshops for nurses (4 NPs, 3 LPNs, 4 RNs, and 1 RN care coordinator), pharmacists (n=2), and tobacco cessation coaches (n=1). Artifacts for these workshops included system workflow diagrams, CDS user screenshots, deidentified patient information, drafts of templates for clinical notes summarizing encounters, and handouts summarizing insights from experiments.
- Output from prototyping workshops: Health care professionals participating in the prototyping workshops designed the "rooming reminder" and "reaching out" prototypes. Both prototypes were evaluated during experiments in the clinic.

Experiments in the Clinic

The experiments in the clinic were carried out as follows:

- Overview: The purpose of the experiments in the clinic was to explore how designs affect and change practice. This phase focused on acting design activities. For acting, activities were conducted as prototype intervention experiments. The experiments enabled quick testing of the prototypes and evaluation of new ideas through iterative cycles [31,37]. During >37 hours of experimentation, 2 prototypes were evaluated. The experiments evaluated the "rooming reminder" and "reaching out" prototypes.

- Rooming reminder experiment:
  - Overview: Insights from patients and health care professionals indicated that, often during routine medical encounters, other complex medical issues are prioritized, and cardiovascular prevention is not addressed. These insights led to the decision to design a rooming reminder experiment. The purpose of the rooming reminder was to remind clinicians to address cardiovascular prevention during upcoming encounters using handouts that summarize the gaps in preventive cardiovascular care for each patient. Handouts were created and named the "Cardiovascular-Patient Appointment Note" (Figure 2). During the experiment, hard copies of these handouts were given to primary care providers by a desk attendant for rapid review before the encounter. The design team observed the Cardiovascular-Patient Appointment Note impact on provider-patient interactions and the content of the medical visits. After 5 days, the team concluded that the Cardiovascular-Patient Appointment Notes had minimal impact. The notes affected only 21% (3/14) of the visits from patients with ASCVDs from a total of 196 visits during this time frame. Providers were interviewed before starting the experiment and asked follow-up questions after experiment completion. In addition, providers were observed during the rooming reminder experiment.
  - Output from the rooming reminder experiment: In total, 3 main insights were gained from the rooming reminder experiment. First, not all the information in the EHR was up to date. Second, only a small number of patients not meeting V4 metrics came to the clinic weekly, suggesting that focusing on current in-visit care will not have the greatest impact. Third, the experiment made clear that the visit context was a major influence on whether cardiovascular health was evaluated. Clinicians used their judgment to determine whether the visit context was appropriate to discuss individual patient cardiovascular metrics. The Cardiovascular-Patient Appointment Note successfully prompted cardiovascular health conversations when all variables identified in Textbox 3 were met, which was rare. Although prompting discussions on cardiovascular prevention during routine outpatient visits can affect care, there is more opportunity to optimize community health through intentional encounters focused on cardiovascular prevention.

- Reaching out experiment:
  - Overview: Insights that guided the "reaching out" experiment were that cardiovascular prevention is often not addressed during routine medical encounters and that rooming reminders had minimal impact. The purpose of the reaching out experiment was to actively contact patients for intentional delivery of preventive cardiovascular care. This experiment had three phases: (1) verification, (2) care coordination and sorting algorithm, and (3) care output (Figure 3). In the reaching out experiment, a total of 8 workflows and 48 processes were developed and tested. A detailed description of these 3 phases is provided in the following sections.
  - Verification phase: Overview: We learned that clinicians must trust the information used for patient management. However, EHR information is often outdated and should be verified with patients before making decisions. The purpose of the verification phase was to gather updated information on the use of guideline-recommended strategies directly from the patients. Output from the verification phase: The initial verification had 2 stages. First, messages were sent through the Mayo Clinic portal app containing a survey asking patients about their cardiovascular prevention status. A team member called patients who did not reply to portal messages and conducted a scripted telephone interview with the same questions used in the survey sent via the portal app. Textbox 4 shows the patient survey questions. As an initial proof of concept, 89 web-based surveys were sent to patients active on the Mayo Clinic portal app. The response rate was 40% (36/89).

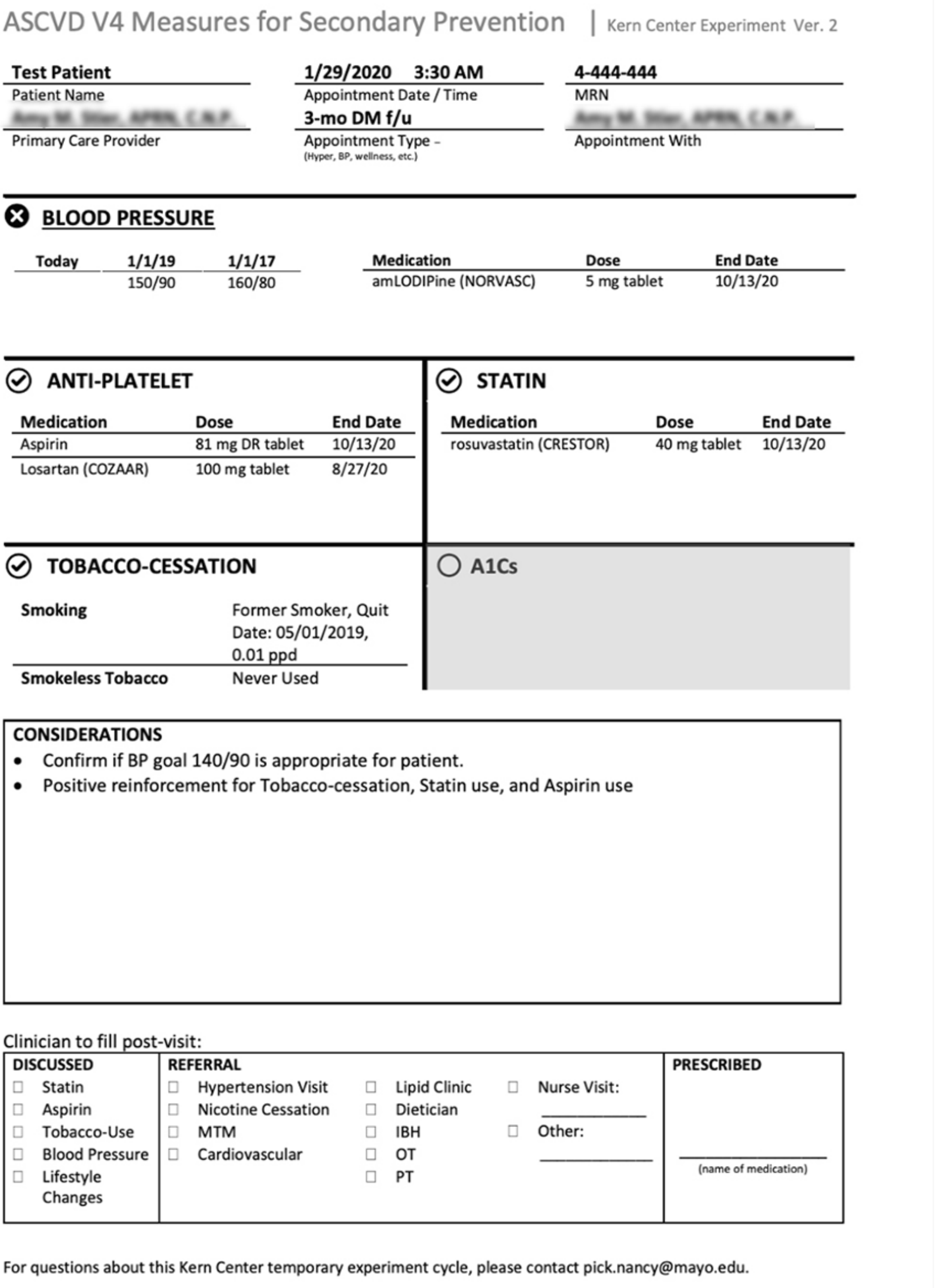
- Care coordination and sorting algorithm phase:
  - Overview: The purpose of this phase was to define criteria to assign the right patient to the right provider. The insight that informed this step was the need to assign the right patient to the right provider. Provider skill sets had to match patient gaps in preventive care such that health care professionals with the appropriate skill set would be assigned to evaluate patients with specific gaps in preventive cardiovascular care. Examples from providers are shown in Textbox 5.
  - Output from the care coordination and sorting algorithm phase: Initially, health care professionals (2 pharmacists, 2 RNs, and 1 advanced practice provider) reviewed the charts of 10 patients with ASCVDs and recommended which care team role should be assigned to each patient. Subsequently, the design team worked with health care professionals to articulate criteria for an automated sorting algorithm assigning patients to the most appropriate team member (both in terms of licensure and specialty) to deliver care plans to each patient tailored to care gaps.

- Care output phase:
  - Overview: The purpose of "care output" was to use patient preference for the selection of the type of encounter for the delivery of cardiovascular prevention by rural providers. The insight that informed the "care output" was that patient preference defined the type of encounter for the delivery of preventive cardiovascular care. The options for encounter types were phone call conversation, telemedicine, or in-person visit. For example, patients with limited access to transportation may prefer either phone call conversations or

telemedicine visits. By contrast, those with access to transportation may prefer an in-person visit. Examples from providers and patients are shown in Textbox 6.

- Output from the care output phase: This experiment created and tested workflows to assign encounter types based on patient preference and 10 templates for clinical notes documenting encounters.

Figure 2. Cardiovascular-Patient Appointment Note handout summarizing the status of use of guideline recommendations by a patient with atherosclerotic cardiovascular disease. Handouts were given to primary care providers before encounters.



ASCVD V4 Measures for Secondary Prevention | Kern Center Experiment Ver. 2

Test Patient	**1/29/2020 3:30 AM**	**4-444-444**
Patient Name	Appointment Date / Time	MRN
[redacted]	**3-mo DM f/u**	[redacted]
Primary Care Provider	Appointment Type – (Hyper, BP, wellness, etc.)	Appointment With

☒ **BLOOD PRESSURE**

Today	1/1/19	1/1/17
	150/90	160/80

Medication	Dose	End Date
amLODIPine (NORVASC)	5 mg tablet	10/13/20

✓ **ANTI-PLATELET**

Medication	Dose	End Date
Aspirin	81 mg DR tablet	10/13/20
Losartan (COZAAR)	100 mg tablet	8/27/20

✓ **STATIN**

Medication	Dose	End Date
rosuvastatin (CRESTOR)	40 mg tablet	10/13/20

✓ **TOBACCO-CESSATION**

Smoking	Former Smoker, Quit Date: 05/01/2019, 0.01 ppd
Smokeless Tobacco	Never Used

○ **A1Cs**

CONSIDERATIONS

- Confirm if BP goal 140/90 is appropriate for patient.
- Positive reinforcement for Tobacco-cessation, Statin use, and Aspirin use

Clinician to fill post-visit:

DISCUSSED

☐ Statin
☐ Aspirin
☐ Tobacco-Use
☐ Blood Pressure
☐ Lifestyle Changes

REFERRAL

☐ Hypertension Visit
☐ Nicotine Cessation
☐ MTM
☐ Cardiovascular
☐ Lipid Clinic
☐ Dietician
☐ IBH
☐ OT
☐ PT
☐ Nurse Visit: ____________
☐ Other: ____________

PRESCRIBED

____
(name of medication)

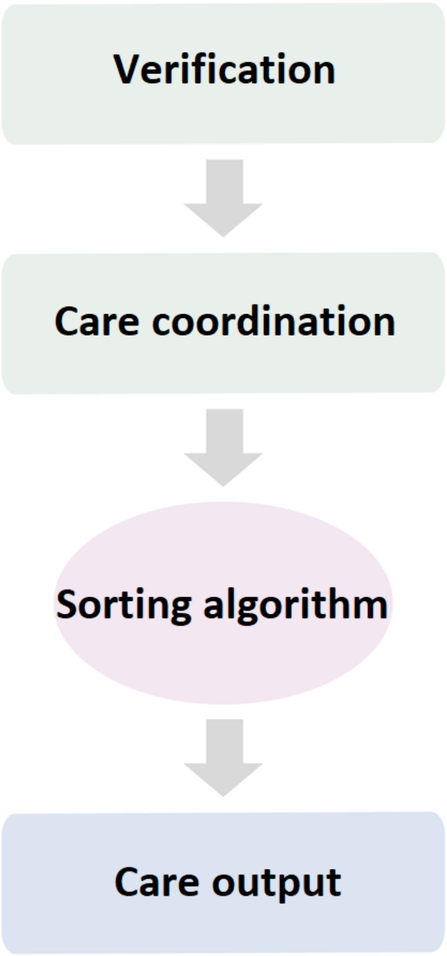
For questions about this Kern Center temporary experiment cycle, please contact pick.nancy@mayo.edu.

Textbox 3. Factors needed before initiating a cardiovascular preventive health conversation with patients.

Factors necessary for starting cardiovascular prevention evaluation

- Visit appropriateness: visits regarding well-managed chronic care more readily transitioned to a conversation on cardiovascular metrics. By contrast, complex or uncontrolled comorbidities became top visit priority and left little time for other discussion.
- Patient appropriateness: provider perception of patient workload to capacity determined whether the provider would address cardiovascular metrics.
- Provider priorities: providers addressed what they viewed as patients' health priorities. Providers needed to believe cardiovascular guideline (*V4*) recommendations were important for patient health for them to address them. Trust between patient and clinician further enabled dialogue and negotiation related to cardiovascular care.
- Trust in information: providers needed to believe that the recommendations were tailored to each patient case and not based on generic guidelines.
- Actionability of information: actionable recommendations reduced the time required for providers to match up information with next steps for patients.

Figure 3. Design of the reaching out experiment.



Textbox 4. Questions in the patient verification survey.

Questions

- Do you currently take a daily aspirin?
- Do you know the dose of aspirin you take daily?
- Do you know another antiplatelet medication you take? Which dose?
- Do you know which statin you take? Which dose?
- Have you tried a statin in the past? Did you experience any adverse reactions because of the statin you took previously?
- Would you be open to our team’s medication expert connecting with you to discuss medication strategies for reducing your risk of future cardiovascular episodes?
- Are you taking daily blood pressure (BP) medications?
- Would you be open to a care team member (a pharmacist or nurses) working with you on custom strategies to lower your BP?
- Do you have a BP monitoring device at home?
- If known, what was your last BP reading from your home BP monitor? What date was it taken?
- Are you currently using tobacco products?
- Would you be open to our tobacco cessation coach calling you to offer information or see if you have questions?
- If your care team wishes to recommend next steps, what would be the best means to contact you? Messages via the Mayo app, phone call, or both?

Textbox 5. Example quotes from providers for the care coordination and sorting algorithm phase.

Example quotes

- “The only thing about blood pressure vs statin and aspirin, statin and aspirin are you, you’re on it or you’re not on it, that’s it, blood pressure we’re adjusting and fine tuning.”
- “Smoking is definitely its own thing.”
- “A lot of different medications, there are certain ones that need lab work, how often am I going to need to monitor you.”
- “Statin and aspirin at a certain level you’re prescribing it now in a perfect world they are taking it as well. But you’re prescribing it whereas smoking and blood pressure are contingent on patient behavior.”
- “They need different tracks.”
- “Different providers own different groups.”

Textbox 6. Example quotes from providers and patients for the care output phase.

Example quotes from providers

- “Assisted living or at home, are they home bound?”
- “Can they get out?”
- “Barriers to even getting here (at the clinic) in the first place.”
- “Where people live and how they get here (at the clinic).”

Example quotes from patients

- “See somebody [health care provider] personally, I prefer that.”
- “I like the good old phone call.”
- “Coming to the clinic, oh it’s nice to be out.”

Results

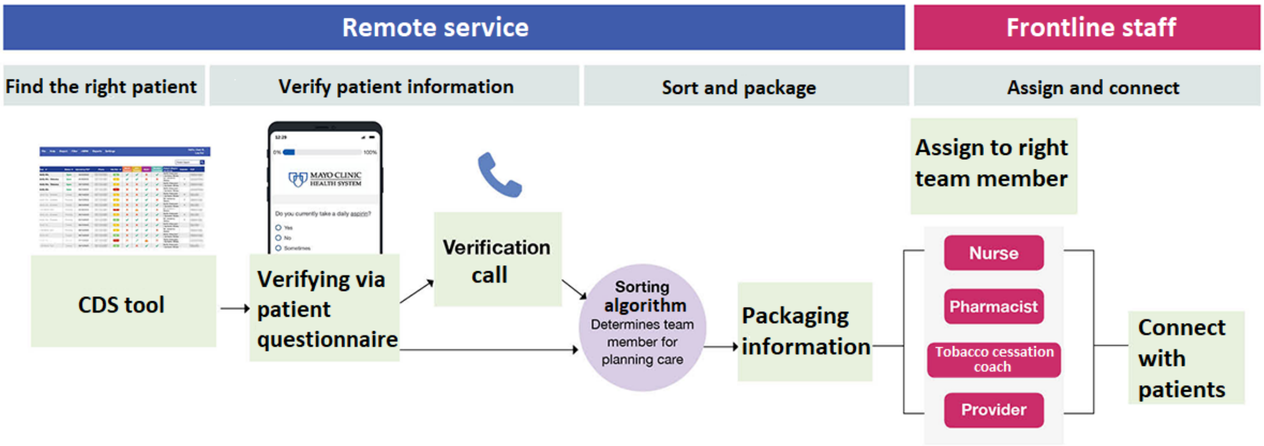
Intervention Prototyping

Overview

The purpose of this phase was to iteratively design an intervention prototype with health care professionals based on the insights from the experimentation phase. The design team delivered handouts for health care professionals during detailed perspective workshops summarizing insights from phase II experiments to facilitate collaborative and iterative idea generation [40].

This process resulted in the design of a sociotechnical intervention prototype with the following components: (1) finding the right patient, (2) verifying patient information, (3) sorting and packaging patient information, and (4) assigning the right provider and connecting with patients (Figure 4). Each component of this intervention is described separately in the following sections.

Figure 4. The final resulting intervention design is a sociotechnical system—a combination of roles, processes, and technology—to enable primary care teams to improve the delivery of cardiovascular prevention strategies. CDS: clinical decision support.



Finding the Right Patient

On the basis of stakeholder feedback, rapid prototyping of the CDS tool was conducted in the testing environment of the web-based technology platform, termed Cohort Knowledge Solution, using the Agile Scrum methodology for software development [41]. The Cohort Knowledge Solution platform is populated by EHR data that are computationally extracted from the institutional data warehouse. Computational phenotyping algorithms were installed in the Cohort Knowledge Solution to automatically identify patients with ASCVDs and identify individual gaps in adherence to *V4* recommendations [42].

ASCVDs include coronary artery disease, peripheral artery disease (PAD), and ischemic stroke [43]. Rule-based billing code algorithms for identification of patients with these conditions were deployed via the Cohort Knowledge Solution platform. Random samples of the retrieved data were manually reviewed by a trained abstractor following the written criteria for standardization to create a reference standard. The processes to support accurate data collection were developed using iterative validation cycles and used Boolean combinations of billing codes [44]. The rule for retrieval of coronary artery disease was a diagnostic code for myocardial infarction or angina pectoris or a procedural code for coronary revascularization procedure (percutaneous or surgical); for ischemic stroke, it required the presence of an International Classification of Diseases, 10th Revision, diagnostic code for ischemic stroke or transient ischemic attack; and, for PAD, a diagnostic code and International Classification of Diseases, 10th Revision, procedural code for lower extremity limb revascularization (endovascular or surgical) was required. Performance metrics were generated by comparison with the reference standards (Table 1).

Given the inferior performance of billing codes for PAD cohort identification, a natural language processing algorithm for the extraction of PAD from clinical narratives was also created and validated with a sensitivity of 91% and a positive predictive value of 92% [45]. This previously validated natural language processing PAD algorithm was also deployed to the Cohort Knowledge Solution to identify cases.

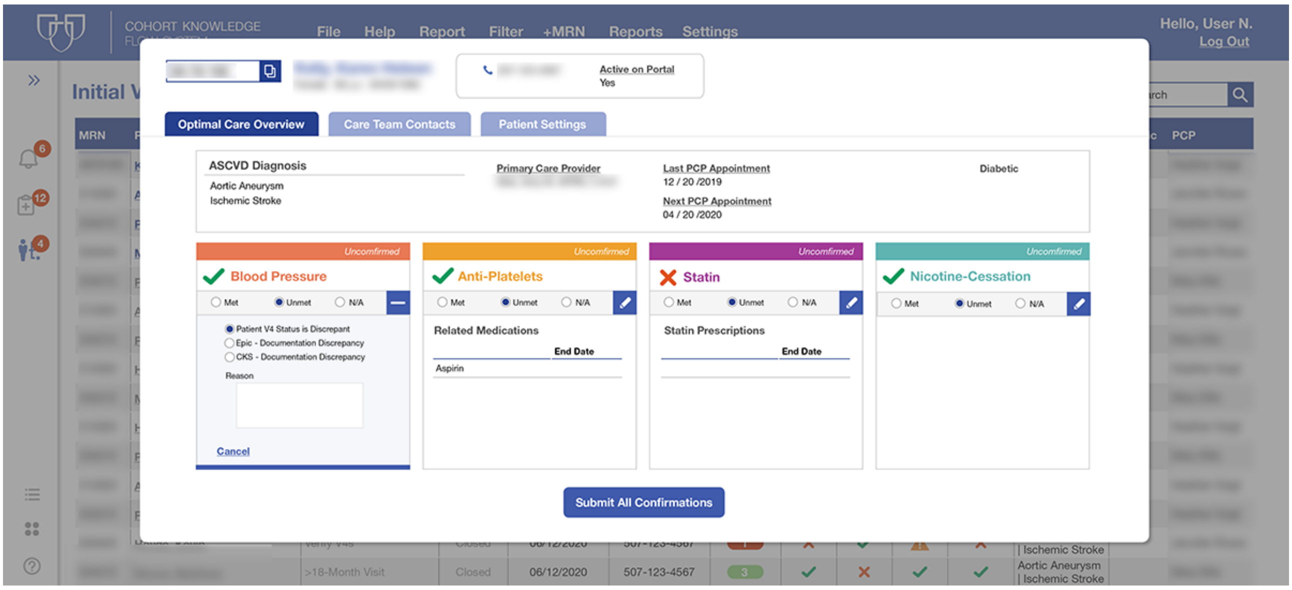
Wireframe usability tests were conducted with care team members to identify which data were most relevant for planning and delivering preventive cardiovascular care. The technology developed also enabled these data to be retrieved from the EHR data and displayed on the same screen with a single mouse click (Figure 5).

Table 1. Performance metrics for billing code algorithms in the Cohort Knowledge Solution.

Type of ASCVD[a]	Charts reviewed, N	Sensitivity, %	Positive predictive value, %	F_1 score, %
Coronary artery disease	189	96	94	95
Peripheral artery disease	140	64	100	78
Stroke	156	98	81	89

[a] ASCVD: atherosclerotic cardiovascular disease.

Figure 5. Cohort Knowledge Solution platform redesigned after user testing with health care professionals.



Active on Portal: Yes

ASCVD Diagnosis: Aortic Aneurysm; Ischemic Stroke

Primary Care Provider

Last PCP Appointment: 12 / 20 /2019

Next PCP Appointment: 04 / 20 /2020

Diabetic

Blood Pressure: ○ Met ● Unmet ○ N/A

● Patient V4 Status is Discrepant
○ Epic - Documentation Discrepancy
○ CKS - Documentation Discrepancy

Reason

Anti-Platelets: ○ Met ● Unmet ○ N/A

Related Medications — End Date
Aspirin

Statin: ○ Met ● Unmet ○ N/A

Statin Prescriptions — End Date

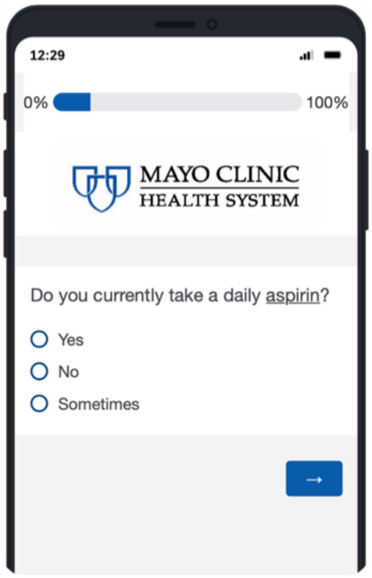
Nicotine-Cessation: ○ Met ● Unmet ○ N/A

>18-Month Visit	Closed	06/12/2020	507-123-4567	3	✓	✗	✓	✓	Aortic Aneurysm Ischemic Stroke

Verifying Patient Information

Health care professionals need the right information to deliver medical care efficiently and effectively. Messages were delivered to patients through the Mayo Clinic portal app with survey questions regarding adherence to *V4* recommendations (Textbox 4 and Figure 6). Patients who did not complete the portal survey, did not answer 3 phone calls from the study team, or declined to participate in this project continued to receive care through the usual model of care. Patients who opted to participate in the intervention were connected with health care professionals with expertise tailored to patient gaps for cardiovascular prevention management. Survey responses were used to update the EHR of each patient, and workflows were designed to support the verification process.

Figure 6. Screenshot of the verification survey sent via the Mayo Clinic patient portal app.



Do you currently take a daily aspirin?

○ Yes
○ No
○ Sometimes

Sorting Algorithm

The design team adapted the sorting algorithm into criteria that the informatics team translated into a sorting computational algorithm for the Cohort Knowledge Solution. Examples of criteria included "if patients were missing statin or anti-platelet, send to pharmacist" or "if a patient reported taking an undocumented anti-platelet, message nurses to update the medical record." The workflows were redesigned by incorporating the new CDS technology.

Packaging Patient Information

The design team originally assumed that the local care team members (eg, pharmacists and nurses) would use the platform. However, during usability testing, the participant health care professionals expressed dissatisfaction with the thought of using another tool for clinical practice. Accordingly, the workflow was redesigned to have a dedicated remote user—analogous to the role of an "air traffic controller." This dedicated remote user took responsibility for care coordination using the Cohort Knowledge Solution in a central hub model. Once patients had filled the verification questionnaire, the dedicated user aggregated relevant patient information within the Cohort Knowledge Solution and electronically assigned it to the right provider.

Assigning the Right Provider

The dedicated user leveraged the Cohort Knowledge Solution sorting algorithm to identify the right providers and sent an in-basket message with the aggregated medical information. The care team members, often nonphysician health care professionals, used the packaged medical information to plan care and "connect with patients" following workflows specifically designed for this phase.

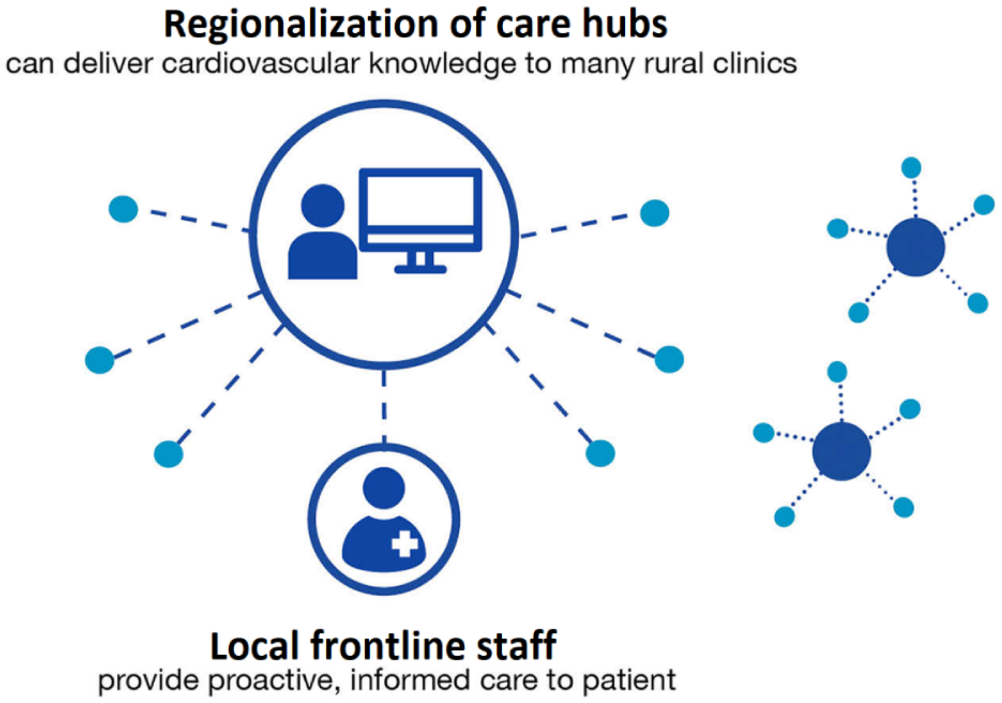
Connecting With Patients

Rather than developing new cardiovascular visits, we focused on routing patients to the existing visit types. Therefore, the redesigned workflows assigned the right patient to the right provider. The provider reviewed the packaged medical information and initiated communication, such as phone calls, video telemedicine, or face-to-face encounters, based on patient preferences and needs.

Centralized Hub Model

To address obstacles stemming from the rural provider shortage, a regionalization of the care hub model was designed (Figure 7). Verifying and packaging patient information took place in a central rural hub with dedicated resources assisted by technology to gather and summarize the patient-specific gaps in preventive care. The information gathered at the central hub was shared with the right rural provider for delivery of care locally. The central hub gathered all the information necessary for the delivery of care, reducing the need for manual chart review by health care professionals and enabling them to focus on delivering tailored care.

Figure 7. The regionalization of the care hub model.



Phase III—Testing

A 3-month prospective randomized pilot trial was conducted in the outpatient primary care clinic in Austin, MN. The goal of this pilot was to test and iteratively refine the sociotechnical intervention prototype with all the components, as shown in Figure 4. Patients with ASCVDs were assigned to the intervention or control arm by stratified randomization with strata based on the number and types of *V4* cardiovascular recommendations in use. The intervention was a rural-specific and sociotechnical model prototype that was proactive and delivered preventive cardiovascular care specific to each patient. The care delivery model for the control arm was usual care, which is a symptom-based and reactive model. In the usual care model, health care professionals rely on patients to contact the system for evaluation of symptoms [46,47]. Patients assigned to the control and intervention arms were evaluated during the same period. A Cohort Knowledge Solution ASCVD registry was used to identify patients with ASCVDs. A total of 369 patients with ASCVDs on ≤3 *V4* guideline recommendations were randomly assigned to the control (n=182, 49.3%) or intervention (n=187, 50.7%) arm. Baseline clinical characteristics were similar in patients assigned to the 2 study arms (Table 2).

Subsequently, the medical records of all patients were reviewed by a trained abstractor before surveying to find extenuating circumstances that would justify not reaching out to a given patient during the pilot. These circumstances included dementia or cognitive impairment, end-stage medical conditions on hospice care, active cancer treatment, hospitalization during the pilot, patient relocation to a different county and no longer receiving medical care from MCHS Austin, upcoming cardiology appointment within 3 months of the pilot, or other medical conditions that did not require the use of *V4* (Figure 8). These reasons were discussed with health care professionals in the detailed perspective workshops and subsequently incorporated into the workflows.

A total of 33 patients crossed over to the usual model of care arm for analysis after this review. After crossing over, 82.4% (154/187) of eligible patients remained in the intervention arm and were surveyed. The patient survey (depicted in Textbox 4) was a component of the intervention and was delivered via portal messages or addressed during scripted telephone interviews. Patients in the intervention arm completed the verification survey and answered the following question: “If your care team wishes to recommend next steps, what would be the best means to contact you? Messages via the Mayo app, phone call, or both?” The patients were contacted using their preferred strategy. Those who did not respond to this question were not contacted. Of the 154 contacted patients, 86 (55.8%) responded to the patient survey via electronic messages or scripted telephone interviews in <3 months. These patients were subsequently assigned to nonphysician health care professionals for the proactive delivery of patient-specific V4 strategies for secondary prevention. There were no differences in the proportion of patients implementing guideline recommendations between those who answered the survey via the patient portal and those who answered the survey by phone call.

The primary outcome of the pilot trial was the proportion of *V4* recommendations delivered, as measured by encounters with nonphysician health care professionals. Encounters with nonphysician health care professionals (nurses, pharmacists, or tobacco cessation coaches) included consults (in person or via telemedicine) or phone conversations. During the pilot, the proportion of patients who had encounters with nonphysician health care professionals for delivery of V4 recommendations in the intervention arm was greater for all types of professionals than in the control arm (Table 3). After exclusion of the 33 patients who crossed over from the intervention arm to the control arm, similar results were observed. In subsequent analysis removing all participants in the control arm who met the criteria for exclusion from the intervention arm (73/369, 19.8%), the results remained unchanged.

During and after the pilot, health care professionals were interviewed, and the information gathered was used to further refine the intervention. The pilot trial demonstrated that this model (1) connected the right health care professional with the right patient, (2) saved time by reducing the need for manual chart review, (3) enabled health care professionals to work to the top of licensure, (4) has potential for expansion to other conditions, and (5) promoted interdisciplinary collaboration to optimize care.

Table 2. Characteristics of patients assigned to the control and intervention arms (N=369).

Clinical variables[a]	Control arm (n=182)	Intervention arm (n=187)
Age (years), mean (SD)	71 (13)	71 (14)
Sex (male), n (%)	92 (50.5)	95 (50.8)
Race (White), n (%)	167 (91.8)	174 (93)
Ethnicity (“not Hispanic or Latino”), n (%)	171 (94)	178 (95.2)
Married, n (%)	94 (51.6)	103 (55.1)
Taking antiplatelet medications, n (%)	120 (65.9)	121 (64.7)
Taking statin medications, n (%)	111 (61)	112 (59.9)
Nonsmokers, n (%)	132 (72.5)	135 (72.2)
Blood pressure at goal, n (%)	87 (47.8)	89 (47.6)

[a] Two-sample 2-tailed *t* tests were used to compare means, and chi-square tests were used for comparison of percentages; all *P* values comparing the control and intervention arms were not significant (*P*>.05).

Figure 8. Pilot trial design. V4: cardiovascular guideline recommendations. *Extenuating circumstances included dementia or cognitive impairment, end-stage medical conditions on hospice care, active cancer treatment, hospitalization during the pilot, patient relocation to a different county and no longer receiving medical care from the Mayo Clinic Health System Austin, upcoming cardiology appointment within 3 months of the pilot, or other medical conditions that did not require the use of V4.

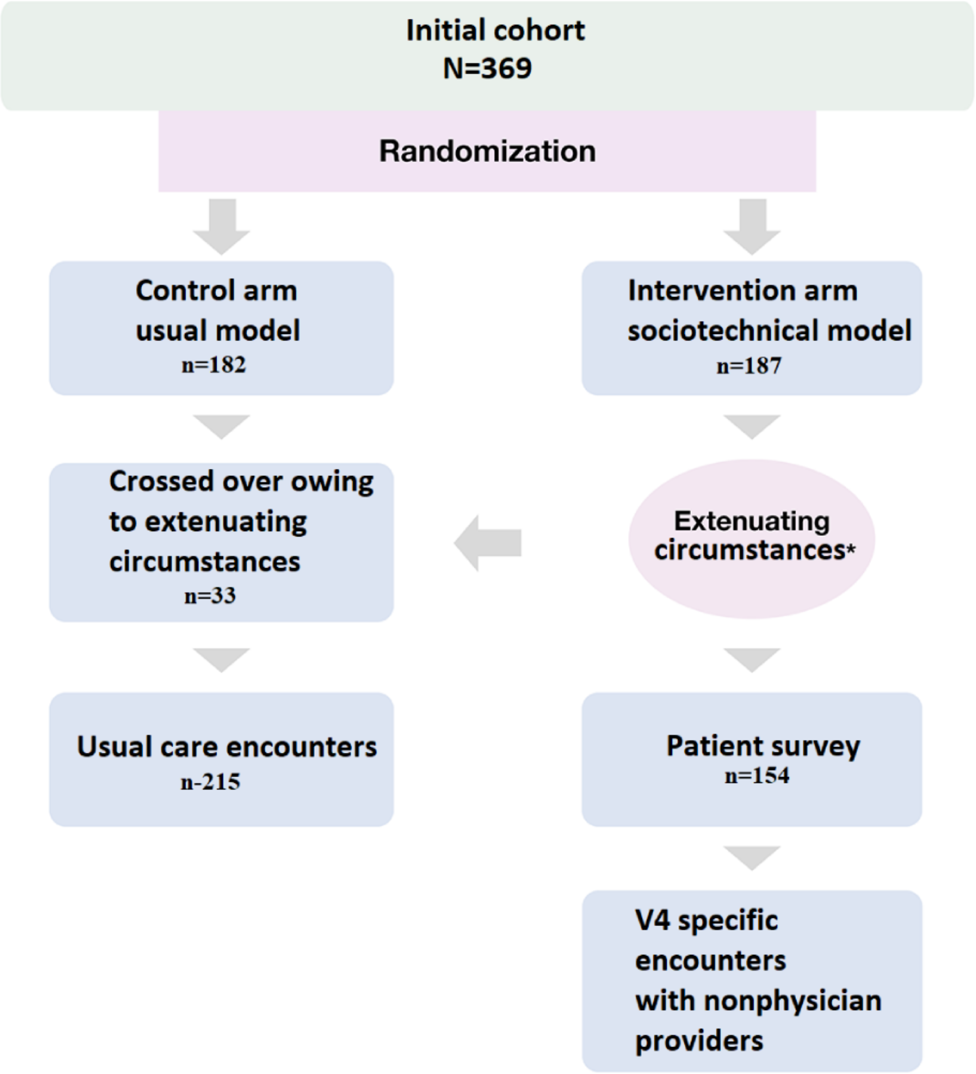


Table 3. Number of encounters for the delivery of cardiovascular guideline recommendations (V4) by nonphysician health care professionals during the pilot (after crossover; N=369).

	Control arm (n=215)[a], n (%)	Intervention arm (n=154)[a], n (%)	*P* value[b]
Nursing encounters for the delivery of *V4* recommendations	48 (22.3)	51 (33.1)	.02
Pharmacist encounters for the delivery of *V4* recommendations	1 (0.5)	31 (20.1)	<.001
Tobacco cessation coach encounters for the delivery of tobacco discontinuation strategies	3 (1.4)	30 (19.5)	<.001

[a] Total number of patients in each arm after crossing over.

[b] Two-sample 2-tailed *t* tests were used to compare means, and chi-square tests were used for comparison of percentages; for all analyses, *P*<.05 was considered significant.

Discussion

Principal Findings

This study used participatory design and the sociotechnical theory framework to create a team-based care model for the coordination and delivery of secondary prevention to rural patients with ASCVDs by nonphysician health care professionals. The new model of care redesigned workflows, integrated health care professional roles, and deployed a novel CDS technology. The subsequent pilot trial demonstrated the feasibility for effective implementation of this new model of care in a rural outpatient clinic. For the next phase of this project, a scalable intervention is planned to be implemented in other rural sites of the MCHS and evaluated in a multisite trial.

The requirements for successful practice transformation include changes in both workflows and technology [48]. Importantly, the optimal use of technology is dependent on the interrelation of the system with skilled and pragmatic work by health care professionals [49]. The design of systems focusing only on technological factors has been a major contributor to their underuse [36,50]. A previous study showed that, for most patients (80% of 5568 patients), providers disregarded the recommendations of a CDS to improve the use of ASCVD secondary prevention at hospital discharge [51]. In that study, the CDS was not part of a sociotechnical system and was not integrated with provider workflows [51]. The low use rate of an EHR-based CDS for cardiovascular risk reduction in community health centers was also reported in a cluster randomized clinical trial that focused on CDS technology [52,53].

This study reported a practical application of the sociotechnical system theory framework to design an intervention to improve the delivery of cardiovascular prevention. The sociotechnical design approach considers both technological and social factors to inform system design [50,54,55]. The use of a sociotechnical theory approach for system design leads to systems that are more acceptable to users and have long-term sustainability [36,56,57]. For these reasons, the sociotechnical design approach was used to design an intervention that will be likely sustainable when implemented in the rural sites of a large medical enterprise.

A previous study in India developed a CDS tool for cardiovascular risk screening used for single-visit encounters [58]. In that study, the CDS was not linked to information from EHRs, and it was not possible to follow patients longitudinally, which is a fundamental requirement for the workflow of providers managing secondary prevention strategies for patients with ASCVDs. By contrast, the CDS of this study was populated with data automatically extracted from EHRs, enabled the delivery of longitudinal care for patients with ASCVDs, and was integrated with provider workflows. A second study reported a CDS linked to the EHR for the primary prevention of patients without established ASCVDs in an urban setting, and a printed copy of the CDS summary screen containing the patient-specific status of use of guideline recommendations was placed on the exam door for rapid review by providers before the encounter [59]. In this study, the rooming reminder experiment used a similar approach. However, this strategy had minimal impact as not all the information in the EHR was up to date, only a small number of patients not meeting V4 metrics came to the clinic weekly, and the visit context had a major influence on whether cardiovascular health was evaluated. On the basis of these insights, the subsequent experiment (reaching out) aimed to actively contact patients for the intentional delivery of preventive cardiovascular care. Insights from the reaching out experiment became a core component of the new model for the intentional delivery of care designed for this study.

Health care professionals work in teams and are commonly assisted by computerized information systems. These systems display the information that different health care professionals need to complete their work. Scandurra et al [28,60] proposed multidisciplinary thematic workshops based on participatory design and computer-supported participatory work theories. This method uses a collaborative design and enables the translation of health care professionals' needs into technical requirements. This study applied this methodology for the development of a team-based model of care, enabling multidisciplinary co-operation among team members for the delivery of preventive care for rural patients with ASCVDs.

A systematic review of the literature by Hardy et al [61] showed that access to the internet, digital literacy, and computer skills are important characteristics for the design of sustainable technology tools for residents of rural areas. This pilot recruited older adult patients from rural communities. Patient recruitment was first conducted using patient portal messages. We observed a low response rate to portal messages and, in the intervention group, 39.6% (61/154) did not have active portal accounts. Other studies have also demonstrated that access to and ability to use technology and the internet are barriers to the use of portals by older adult patients [62-64]. To overcome these barriers, in this study, patients who did not have active portal accounts or did not respond to portal messages were contacted by phone. The same survey questions sent via portal messages were used for scripted telephone conversations. The survey questions were simple and focused on patient-specific gaps in preventive cardiovascular care. The survey response rate of 55.8% (86/154) was superior to the 43.9% response rate of a previous survey of community-dwelling older adults in rural areas [14].

On the basis of our observations, rural providers need to consider patient preference for visit type to mitigate barriers to transportation and limited mobility, which often occur in rural residents, especially older adults, who were a major target group for this intervention. Compared with their urban counterparts, rural citizens are more prone to mobility and transportation barriers that make access to health care difficult [61]. In mitigation, phone conversations and telemedicine became options of visit types connecting patients and providers for the delivery of secondary prevention strategies for patients with ASCVDs. In addition, it is important to underscore that rural society values "neighborliness," which manifests as trust in community members and potential distrust of outsiders [61]. The intervention developed in this study leveraged neighborliness and supported a rural place-based identity. Consequently, the hub model was maintained regionally rather than centrally in Rochester, MN (Mayo Clinic headquarters). The regionalization of the care model promoted the delivery of care by rural providers. During the implementation of this intervention, staff resources will be distributed to serve small rural clinics located in the area covered by a hub. In other rural health care settings, a similar process for the allocation of rural providers may be used.

The novel model of care described in this study assigns the right patient to the right professional for the delivery of preventive care. In addition, the selection of the encounter type was based on patient preference for in-person visits at the clinic, phone conversations, or video encounters via telemedicine. This flexibility facilitated access to health care professionals for patients with limited transportation resources. Furthermore, this model also enabled a patient-centered health connection that goes beyond traditional symptom-based visits. The primary outcome of the pilot trial was the proportion of *V4* recommendations delivered, as measured by encounters with nonphysician health care professionals. These metrics were set a priori following good practice. During the pilot, the proportion of patients who had encounters with nonphysician health care professionals for the delivery of *V4* recommendations in the intervention arm was greater for all types of professionals than in the control arm. Therefore, the pilot trial showed that this model connected the right health care professional with the right patient for the delivery of guideline-recommended strategies for patients with ASCVDs, demonstrating the feasibility of the intervention.

Limitations

The pilot study was not powered to evaluate the differences in the use of specific guideline-recommended strategies. However, we are planning a subsequent prospective randomized implementation trial in the Midwest sites of the MCHS with central hubs spanning multiple rural clinics. This trial will be powered to evaluate the impact of the intervention on the proportion of patients using specific guideline-recommended strategies. In preparation for this implementation, the informatics and IT teams have been building the additional Cohort Knowledge Solution functionalities designed in this study, which are aligned with the new workflows. These functionalities include automation of the sorting algorithm and automated retrieval of extenuating circumstances.

The reason for not including patients in the workshops was that the Cardiovascular-Patient Appointment Note experiment showed that a major obstacle for the delivery of preventive cardiovascular care was the difficulty in reaching out to patients. Strategies for health care professionals to reach out to patients were developed to overcome this barrier. The cost of visits was a concern for patients. However, during this study, there was no additional cost for the patients. Further analysis of the cost of health care and strategies for billing will be performed during the implementation phase. During the pilot trial, health care professionals were asked questions about their experience with the intervention. However, patients were not asked similar questions. During the planned implementation trial, patients will be asked to identify barriers to and facilitators of the implementation of this sociotechnical intervention.

This novel team-based model of care was specifically designed to enable the delivery of care in resource-constrained clinics located in rural communities and promotes teamwork with shared responsibilities among team members. For clinics that do not have certain types of health care professionals (eg, pharmacists), we propose to use the resources of the regional hubs, in which health care professionals from other rural clinics could remotely support the teams where this specific expertise is not available. In addition, during the process of implementation, further information will be gathered about the characteristics of the intervention that are necessary for adaptation to other enterprises. Other implementation studies are needed to evaluate the reproducibility and scalability of this model of care to other enterprises that deliver health care to patients in rural communities.

Conclusions

Participatory design within the sociotechnical theory framework enabled the development of a rural-specific, team-based care intervention assisted by CDS technology for the transformation of preventive health care delivery in rural clinics for patients with ASCVDs. By systematically promoting preventive care, this intervention has the potential to strengthen longitudinal relationships between clinics and their communities—the underlying requirement for secondary prevention for patients with ASCVDs in rural settings.

Acknowledgments

This research project was funded by the Mayo Clinic Practice Transformation Award. The authors thank the pharmacists, tobacco cessation specialists, nurse managers, registered nurses, licensed practical nurses, nurse practitioners, care coordinators, and informatics nurses for their invaluable contribution in developing the intervention. The authors also thank the IT and informatics teams for their work on the Cohort Knowledge Solution tool; the study coordinators for conducting the scripted telephone conversations; the project managers for management of the project; Tessa Flies for secretarial support; Jor Arcila for the picture posted on the Journal's website; and Jill Smith, Jennifer Walter, and Dr Robert C Albright Jr for their invaluable support.

Conflicts of Interest

None declared.

References

1. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics-2019 update: a report from the American heart association. Circulation 2019 Mar 05;139(10):e56-528 [FREE Full text] [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
2. Harrington RA, Califf RM, Balamurugan A, Brown N, Benjamin RM, Braund WE, et al. Call to action: rural health: a presidential advisory from the American heart association and American stroke association. Circulation 2020 Mar 10;141(10):e615-e644 [FREE Full text] [doi: 10.1161/CIR.0000000000000753] [Medline: 32078375]

3. Lundeen EA, Park S, Pan L, O'Toole T, Matthews K, Blanck HM. Obesity prevalence among adults living in metropolitan and nonmetropolitan counties - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Jun 15;67(23):653-658 [FREE Full text] [doi: 10.15585/mmwr.mm6723a1] [Medline: 29902166]
4. Mainous AG, King DE, Garr DR, Pearson WS. Race, rural residence, and control of diabetes and hypertension. Ann Fam Med 2004;2(6):563-568 [FREE Full text] [doi: 10.1370/afm.119] [Medline: 15576542]
5. Matthews KA, Croft JB, Liu Y, Lu H, Kanny D, Wheaton AG, et al. Health-related behaviors by urban-rural county classification - United States, 2013. MMWR Surveill Summ 2017 Feb 03;66(5):1-8 [FREE Full text] [doi: 10.15585/mmwr.ss6605a1] [Medline: 28151923]
6. Patterson PD, Moore CG, Probst JC, Shinogle JA. Obesity and physical inactivity in rural America. J Rural Health 2004;20(2):151-159. [doi: 10.1111/j.1748-0361.2004.tb00022.x] [Medline: 15085629]
7. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American college of cardiology/American heart association task force on clinical practice guidelines. Circulation 2019 Jun 18;139(25):e1082-e1143 [FREE Full text] [doi: 10.1161/CIR.0000000000000625] [Medline: 30586774]
8. Kumbhani DJ, Steg PG, Cannon CP, Eagle KA, Smith SC, Hoffman E, REduction of Atherothrombosis for Continued Health Registry Investigators. Adherence to secondary prevention medications and four-year outcomes in outpatients with atherosclerosis. Am J Med 2013 Aug;126(8):693-700.e1. [doi: 10.1016/j.amjmed.2013.01.033] [Medline: 23800583]
9. Smith SC, Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, et al. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation endorsed by the World Heart Federation and the Preventive Cardiovascular Nurses Association. J Am Coll Cardiol 2011 Nov 29;58(23):2432-2446 [FREE Full text] [doi: 10.1016/j.jacc.2011.10.824] [Medline: 22055990]
10. Jia X, Al Rifai M, Birnbaum Y, Smith SC, Virani SS. The 2018 cholesterol management guidelines: topics in secondary ASCVD prevention clinicians need to know. Curr Atheroscler Rep 2019 Apr 02;21(6):20. [doi: 10.1007/s11883-019-0784-8] [Medline: 30941517]
11. Virani SS, Smith SC, Stone NJ, Grundy SM. Secondary prevention for atherosclerotic cardiovascular disease: comparing recent US and European guidelines on dyslipidemia. Circulation 2020 Apr 07;141(14):1121-1123. [doi: 10.1161/CIRCULATIONAHA.119.044282] [Medline: 32250694]
12. Hsu S, Ton V, Dominique Ashen M, Martin SS, Gluckman TJ, Kohli P, et al. A clinician's guide to the ABCs of cardiovascular disease prevention: the Johns Hopkins Ciccarone Center for the Prevention of Heart Disease and American College of Cardiology Cardiosource Approach to the Million Hearts Initiative. Clin Cardiol 2013 Jul;36(7):383-393 [FREE Full text] [doi: 10.1002/clc.22137] [Medline: 23670948]
13. Halperin JL, Levine GN, Al-Khatib SM, Birtcher KK, Bozkurt B, Brindis RG, et al. Further evolution of the ACC/AHA clinical practice guideline recommendation classification system: a report of the American college of cardiology/American heart association task force on clinical practice guidelines. J Am Coll Cardiol 2016 Apr 05;67(13):1572-1574 [FREE Full text] [doi: 10.1016/j.jacc.2015.09.001] [Medline: 26409257]
14. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992 Nov 04;268(17):2420-2425. [doi: 10.1001/jama.1992.03490170092032] [Medline: 1404801]
15. Fanaroff AC, Califf RM, Windecker S, Smith SC, Lopes RD. Levels of evidence supporting American college of cardiology/American heart association and European society of cardiology guidelines, 2008-2018. JAMA 2019 Mar 19;321(11):1069-1080 [FREE Full text] [doi: 10.1001/jama.2019.1122] [Medline: 30874755]
16. Lopez-Abuin JM. Review of: 'Increasing access to health workers in remote and rural areas through improved retention: global policy recommendations: Global policy recommendations'. Rural Remote Health 2010 Nov 18:1654. [doi: 10.22605/rrh1647]
17. Harrington RA, Heidenreich PA. Team-based care and quality: a move toward evidence-based policy. J Am Coll Cardiol 2015 Oct 20;66(16):1813-1815 [FREE Full text] [doi: 10.1016/j.jacc.2015.08.016] [Medline: 26483106]
18. Giberson SF. Million Hearts(TM): pharmacist-delivered care to improve cardiovascular health. Public Health Rep 2013;128(1):2-6 [FREE Full text] [doi: 10.1177/003335491312800102] [Medline: 23277654]
19. Virani SS, Maddox TM, Chan PS, Tang F, Akeroyd JM, Risch SA, et al. Provider type and quality of outpatient cardiovascular disease care: insights from the NCDR PINNACLE registry. J Am Coll Cardiol 2015 Oct 20;66(16):1803-1812 [FREE Full text] [doi: 10.1016/j.jacc.2015.08.017] [Medline: 26483105]
20. Vink P, Imada AS, Zink KJ. Defining stakeholder involvement in participatory design processes. Appl Ergon 2008 Jul;39(4):519-526. [doi: 10.1016/j.apergo.2008.02.009] [Medline: 18417096]
21. Lawler EH, Hedge A, Pavlovic-Veselinovic S. Cognitive ergonomics, socio-technical systems, and the impact of healthcare information technologies. Int J Ind Ergon 2011 Jul;41(4):336-344 [FREE Full text] [doi: 10.1016/j.ergon.2011.02.006]
22. Jessop AB, Bass SB, Brajuha J, Alhajji M, Burke M, Gashat MT, et al. "Take charge, get cured": pilot testing a targeted mHealth treatment decision support tool for methadone patients with hepatitis c virus for acceptability and promise of efficacy. J Subst Abuse Treat 2020 Feb;109:23-33. [doi: 10.1016/j.jsat.2019.11.001] [Medline: 31856947]

23. Kakar R, Combs R, Ali N, Muvuka B, Porter J. Enhancing the design and utilization of asthma action plans through community-based participatory research in an urban African American community. Patient Educ Couns 2021 Feb;104(2):276-281. [doi: 10.1016/j.pec.2020.08.032] [Medline: 32921519]
24. Lipson-Smith R, White F, White A, Serong L, Cooper G, Price-Bell G, et al. Co-design of a consultation audio-recording mobile app for people with cancer: the SecondEars app. JMIR Form Res 2019 Mar 12;3(1):e11111 [FREE Full text] [doi: 10.2196/11111] [Medline: 30860487]
25. Suen AO, Butler RA, Arnold R, Myers B, Witteman HO, Cox CE, et al. Developing the family support tool: an interactive, web-based tool to help families navigate the complexities of surrogate decision making in ICUs. J Crit Care 2020 Apr;56:132-139 [FREE Full text] [doi: 10.1016/j.jcrc.2019.12.002] [Medline: 31896447]
26. Tran NT, Yameogo WM, Langwana F, Kouanda S, Thieba B, Mashinda D, et al. Birth spacing and informed family planning choices after childbirth in Burkina Faso and the Democratic Republic of Congo: participatory action research to design and evaluate a decision-making tool for providers and their clients. Patient Educ Couns 2018 Oct;101(10):1871-1875. [doi: 10.1016/j.pec.2018.05.004] [Medline: 30001823]
27. Estiri H, Lovins T, Afzalan N, Stephens KA. Applying a participatory design approach to define objectives and properties of a "Data profiling" tool for electronic health data. AMIA Jt Summits Transl Sci Proc 2016;2016:60-67 [FREE Full text] [Medline: 27570651]
28. Scandurra I, Hägglund M, Koch S. From user needs to system specifications: multi-disciplinary thematic seminars as a collaborative design method for development of health information systems. J Biomed Inform 2008 Aug;41(4):557-569 [FREE Full text] [doi: 10.1016/j.jbi.2008.01.012] [Medline: 18394969]
29. 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]
30. Carroll J, Rosson M. Participatory design in community informatics. Design Stud 2007 May;28(3):243-261 [FREE Full text] [doi: 10.1016/j.destud.2007.02.007]
31. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
32. Thobias J, Kiwanuka A. Design and implementation of an m-health data model for improving health information access for reproductive and child health services in low resource settings using a participatory action research approach. BMC Med Inform Decis Mak 2018 Jun 25;18(1):45 [FREE Full text] [doi: 10.1186/s12911-018-0622-x] [Medline: 29941008]
33. Seguin RA, Morgan EH, Connor LM, Garner JA, King AC, Sheats JL, et al. Rural food and physical activity assessment using an electronic tablet-based application, New York, 2013-2014. Prev Chronic Dis 2015 Jul 02;12:E102 [FREE Full text] [doi: 10.5888/pcd12.150147] [Medline: 26133645]
34. Sjöberg C, Timpka T. Participatory design of information systems in health care. J Am Med Inform Assoc 1998;5(2):177-183 [FREE Full text] [doi: 10.1136/jamia.1998.0050177] [Medline: 9524350]
35. Doherty NF. The role of socio-technical principles in leveraging meaningful benefits from IT investments. Appl Ergon 2014 Mar;45(2):181-187. [doi: 10.1016/j.apergo.2012.11.012] [Medline: 23305923]
36. Baxter G, Sommerville I. Socio-technical systems: from design methods to systems engineering. Interact Comput 2011 Jan;23(1):4-17 [FREE Full text] [doi: 10.1016/j.intcom.2010.07.003]
37. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]
38. Hackett P, Hayre C, editors. Handbook of Ethnography in Healthcare Research. London and New York: Routledge; 2021.
39. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516 [FREE Full text] [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
40. Fox NJ. Boundary objects, social meanings and the success of new technologies. Sociol 2011 Feb 04;45(1):70-85. [doi: 10.1177/0038038510387196]
41. Dingsøyr T, Nerur S, Balijepally V, Moe NB. A decade of agile methodologies: towards explaining agile software development. J Syst Softw 2012 Jun 13;85(6):1213-1221 [FREE Full text] [doi: 10.1016/j.jss.2012.02.033]
42. Sundaram DA, Berg T, Devens C, Bicknese L, Thattil B, Ramu P, et al. Improving workload management for reducing readmissions and follow-up visits in outpatient practice using a novel application - Cohort Knowledge Solutions (CKS). In: Proceedings of the 2017 IEEE International Conference on Electro Information Technology (EIT). 2017 Presented at: 2017 IEEE International Conference on Electro Information Technology (EIT); May 14-17, 2017; Lincoln, NE, USA. [doi: 10.1109/eit.2017.8053344]
43. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics-2020 update: a report from the American Heart Association. Circulation 2020 Mar 03;141(9):e139-e596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]

44. Rasmussen LV, Thompson WK, Pacheco JA, Kho AN, Carrell DS, Pathak J, et al. Design patterns for the development of electronic health record-driven phenotype extraction algorithms. J Biomed Inform 2014 Oct;51:280-286 [FREE Full text] [doi: 10.1016/j.jbi.2014.06.007] [Medline: 24960203]
45. Afzal N, Sohn S, Abram S, Scott CG, Chaudhry R, Liu H, et al. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. J Vasc Surg 2017 Jun;65(6):1753-1761 [FREE Full text] [doi: 10.1016/j.jvs.2016.11.031] [Medline: 28189359]
46. Dinan MA, Simmons LA, Snyderman R. Commentary: personalized health planning and the patient protection and affordable care act: an opportunity for academic medicine to lead health care reform. Acad Med 2010 Nov;85(11):1665-1668. [doi: 10.1097/ACM.0b013e3181f4ab3c] [Medline: 20844424]
47. Simmons LA, Drake CD, Gaudet TW, Snyderman R. Personalized health planning in primary care settings. Fed Pract 2016 Jan;33(1):27-34 [FREE Full text] [Medline: 30766135]
48. Change Management in EHR Implementation. Center HITR. URL: https://www.healthit.gov/sites/default/files/tools/nlc_changemanagementprimer.pdf [accessed 2022-07-26]
49. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999 Aug;55(2):87-101. [doi: 10.1016/s1386-5056(99)00011-8] [Medline: 10530825]
50. Wesley DB, Schubel L, Hsiao C, Burn S, Howe J, Kellogg K, et al. A socio-technical systems approach to the use of health IT for patient reported outcomes: patient and healthcare provider perspectives. J Biomed Inform 2019;100S:100048 [FREE Full text] [doi: 10.1016/j.yjbinx.2019.100048] [Medline: 34384570]
51. Vani A, Kan K, Iturrate E, Levy-Lambert D, Smilowitz NR, Saxena A, et al. Leveraging clinical decision support tools to improve guideline-directed medical therapy in patients with atherosclerotic cardiovascular disease at hospital discharge. Cardiol J 2020 Sep 28 (forthcoming) [FREE Full text] [doi: 10.5603/CJ.a2020.0126] [Medline: 32986236]
52. Gold R, Larson AE, Sperl-Hillen JM, Boston D, Sheppler CR, Heintzman J, et al. Effect of clinical decision support at community health centers on the risk of cardiovascular disease: a cluster randomized clinical trial. JAMA Netw Open 2022 Feb 01;5(2):e2146519 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.46519] [Medline: 35119463]
53. Shah NR, Khetpal V, Erqou S. Anticipating and addressing challenges during implementation of clinical decision support systems. JAMA Netw Open 2022 Feb 01;5(2):e2146528 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.46528] [Medline: 35119466]
54. Whitworth B, Ahmad A. The Social Design of Technical Systems Building Technologies for Communities. Aarhus, Denmark: Interaction Design Foundation; 2013.
55. Davis MC, Challenger R, Jayewardene DN, Clegg CW. Advancing socio-technical systems thinking: a call for bravery. Appl Ergon 2014 Mar;45(2):171-180. [doi: 10.1016/j.apergo.2013.02.009] [Medline: 23664481]
56. Braithwaite J, Runciman WB, Merry AF. Towards safer, better healthcare: harnessing the natural properties of complex sociotechnical systems. Qual Saf Health Care 2009 Feb;18(1):37-41 [FREE Full text] [doi: 10.1136/qshc.2007.023317] [Medline: 19204130]
57. Bednar PM, Welch C. Socio-technical perspectives on smart working: creating meaningful and sustainable systems. Inf Syst Front 2019 May 3;22(2):281-298. [doi: 10.1007/s10796-019-09921-1]
58. Praveen D, Patel A, Raghu A, Clifford GD, Maulik PK, Mohammad Abdul A, et al. SMARTHealth India: development and field evaluation of a mobile clinical decision support system for cardiovascular diseases in rural India. JMIR Mhealth Uhealth 2014 Dec 08;2(4):e54 [FREE Full text] [doi: 10.2196/mhealth.3568] [Medline: 25487047]
59. Sperl-Hillen JM, Crain AL, Margolis KL, Ekstrom HL, Appana D, Amundson G, et al. Clinical decision support directed to primary care patients and providers reduces cardiovascular risk: a randomized trial. J Am Med Inform Assoc 2018 Sep 01;25(9):1137-1146 [FREE Full text] [doi: 10.1093/jamia/ocy085] [Medline: 29982627]
60. Scandurra I, Hägglund M, Koch S. Application of the multi-disciplinary thematic seminar method in two homecare cases - a comparative study. Stud Health Technol Inform 2008;136:597-602. [Medline: 18487796]
61. Hardy J, Wyche S, Veinot T. Rural HCI research: definitions, distinctions, methods, and opportunities. Proc ACM Human Comput Interact 2019 Nov 07;3(CSCW):1-33. [doi: 10.1145/3359298]
62. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient portal use and experience among older adults: systematic review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
63. Taha J, Sharit J, Czaja SJ. The impact of numeracy ability and technology skills on older adults' performance of health management tasks using a patient portal. J Appl Gerontol 2014 Jun;33(4):416-436 [FREE Full text] [doi: 10.1177/0733464812447283] [Medline: 24781964]
64. Turner AM, Osterhage K, Hartzler A, Joe J, Lin L, Kanagat N, et al. Use of patient portals for personal health information management: the older adult perspective. AMIA Annu Symp Proc 2015;2015:1234-1241 [FREE Full text] [Medline: 26958263]

Abbreviations

AHA: American Heart Association
ASCVD: atherosclerotic cardiovascular disease

CDS: clinical decision support
EHR: electronic health record
LPN: licensed practical nurse
MCHS: Mayo Clinic Health System
MN: Minnesota
NP: nurse practitioner
PAD: peripheral artery disease
RN: registered nurse
V4: cardiovascular guideline recommendations

Edited by A Mavragani; submitted 19.03.21; peer-reviewed by I Hargraves, T Veinot, E Weissler, K Sellen; comments to author 23.04.21; revised version received 30.12.21; accepted 27.06.22; published 22.08.22.

Please cite as:
Partogi M, Gaviria-Valencia S, Alzate Aguirre M, Pick NJ, Bhopalwala HM, Barry BA, Kaggal VC, Scott CG, Kessler ME, Moore MM, Mitchell JD, Chaudhry R, Bonacci RP, Arruda-Olson AM
Sociotechnical Intervention for Improved Delivery of Preventive Cardiovascular Care to Rural Communities: Participatory Design Approach
J Med Internet Res 2022;24(8):e27333
URL: https://www.jmir.org/2022/8/e27333
doi:10.2196/27333
PMID:35994324

Original Paper

Assessing Social Media Data as a Resource for Firearm Research: Analysis of Tweets Pertaining to Firearm Deaths

Lisa Singh[1], PhD; Carole Roan Gresenz[2], PhD; Yanchen Wang[3], MA; Sonya Hu[3], BSc

[1]Department of Computer Science, Massive Data Institute, Georgetown University, Washington, DC, United States
[2]McCourt School of Public Policy, School of Health, Georgetown University, Washington, DC, United States
[3]Department of Computer Science, Georgetown University, Washington, DC, United States

Corresponding Author:
Lisa Singh, PhD
Department of Computer Science
Massive Data Institute
Georgetown University
3700 O Street, NW
Washington, DC, 20057
United States
Phone: 1 2026879253
Email: lisa.singh@georgetown.edu

Abstract

Background: Historic constraints on research dollars and reliable information have limited firearm research. At the same time, interest in the power and potential of social media analytics, particularly in health contexts, has surged.

Objective: The aim of this study is to contribute toward the goal of establishing a foundation for how social media data may best be used, alone or in conjunction with other data resources, to improve the information base for firearm research.

Methods: We examined the value of social media data for estimating a firearm outcome for which robust benchmark data exist—specifically, firearm mortality, which is captured in the National Vital Statistics System (NVSS). We hand curated tweet data from the Twitter application programming interface spanning January 1, 2017, to December 31, 2018. We developed machine learning classifiers to identify tweets that pertain to firearm deaths and develop estimates of the volume of Twitter firearm discussion by month. We compared within-state variation over time in the volume of tweets pertaining to firearm deaths with within-state trends in NVSS-based estimates of firearm fatalities using Pearson linear correlations.

Results: The correlation between the monthly number of firearm fatalities measured by the NVSS and the monthly volume of tweets pertaining to firearm deaths was weak (median 0.081) and highly dispersed across states (range –0.31 to 0.535). The median correlation between month-to-month changes in firearm fatalities in the NVSS and firearm deaths discussed in tweets was moderate (median 0.30) and exhibited less dispersion among states (range –0.06 to 0.69).

Conclusions: Our findings suggest that Twitter data may hold value for tracking dynamics in firearm-related outcomes, particularly for relatively populous cities that are identifiable through location mentions in tweet content. The data are likely to be particularly valuable for understanding firearm outcomes not currently measured, not measured well, or not measurable through other available means. This research provides an important building block for future work that continues to develop the usefulness of social media data for firearm research.

(J Med Internet Res 2022;24(8):e38319) doi:10.2196/38319

KEYWORDS

firearms; fatalities; Twitter; firearm research; social media data

Introduction

Motivation

Firearm violence is a major and costly public health burden in the United States [1-3], and constraints on the availability of research dollars and reliable information to support firearm research have imposed limits on the ability to gather scientific evidence on effective gun policy [4-7]. At the same time, interest in the power and potential of social media analytics in public health contexts has surged. Several aspects of social media data have heightened their promise as a resource, including the fact that the data are inexpensive to obtain compared with survey data; provide access to continuous, automated, and near–real-time monitoring; and are passively collected in a naturalistic setting as part of an individual's day-to-day life, eliminating biases inherent to sampling procedures, questionnaires, and recall [8-13]. Such data are, of course, not without their own methodological challenges and limitations, and practices for their ethical and meaningful use are evolving [14-16].

To date, such data have been deployed in firearm-related research in several ways, including to record narratives, sentiment, and emotion around shooting events [17-20]; characterize gun advertisements on social media [21]; and reflect opinions on gun policies and gun control [22,23]. In this paper, we take up the question of how social media data may contribute to understanding firearm-related outcomes. We identify methodological approaches, challenges, and limitations associated with using social media data for understanding a specific firearm outcome—firearm mortality—for which a benchmark measure for comparison is available from a traditional data source. The analysis of firearm mortality is intended to serve as a test of the potential utility of social media data for understanding firearm outcomes not currently measured, not measured well, or not measurable through other available means.

Assessing the Usefulness of Twitter Data

Specifically, we assessed the usefulness of Twitter data for understanding firearm mortality. Twitter is an online microblogging platform that has >206 million daily active users worldwide and >77 million daily active users in the United States [24]. A key feature of Twitter is its short format: members can only post messages, known as *tweets*, of up to 280 characters. We developed machine learning (ML) classifiers for identifying tweets that pertain to firearm fatalities and compared measures of firearm-fatality discussion volume to firearm-fatality estimates by state from the National Vital Statistics System (NVSS). The NVSS represents one of the few sources of US health-related data with consistently collected and reliable information on a specific gun outcome measured by geographic area. Our goal was to begin to establish a foundation for how social media data may be used by itself or in conjunction with other data resources, such as through data-blending techniques, to improve the information base on which firearm research relies.

Methods

Ethics Approval

The institutional review board of Georgetown University reviewed our submission, STUDY00002288, and determined the study to be exempt.

Overview

Our overarching approach was to compare—within state over time (by month)—measures of firearm-fatality tweet discussion volume with NVSS estimates of firearm fatalities using Pearson linear correlations. Methodologically, with respect to Twitter data, we used a multistage process as described in detail in the following subsections. We first describe our benchmark data and then describe in detail our approach to analyzing Twitter data.

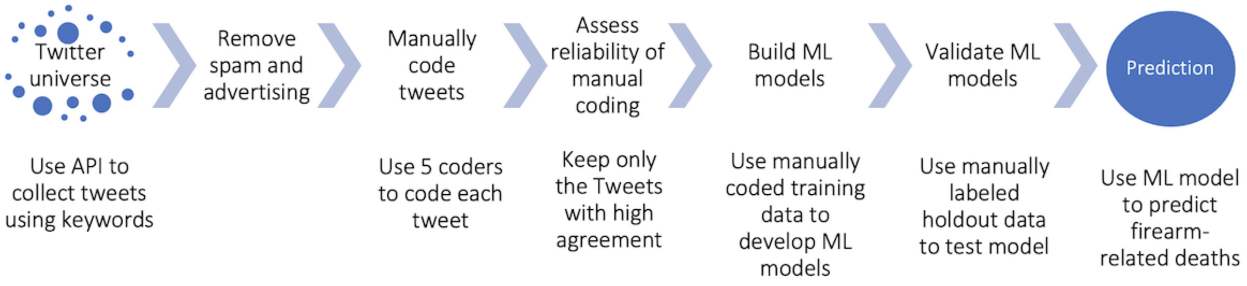
Benchmark Data

Our benchmark data are NVSS estimates of overall firearm fatalities by state and month for 2017 and 2018. Diagnostic (International Classification of Diseases, Tenth Revision) codes in the NVSS identify mortality from accidental firearm discharges, assaults (homicides) by discharge of firearms, and intentional self-harm (suicides) by firearms. Data are collected nationwide using standardized forms and a set of common procedures to ensure comparability of data across locations.

Twitter Data

We developed a Twitter-based gun-related analytic platform based on content culled from the Twitter Enterprise application programming interface (API) for the 2017-2018 time period through the multistage process depicted in Figure 1. The process consists of 4 stages to prepare the data for ML and 3 stages associated with ML analysis.

Figure 1. Construction of the Twitter-based gun-related analytic platform. API: application programming interface; ML: machine learning.



The API allows permitted users to access publicly available Twitter content—including tweets; tweet IDs (a unique identification number generated for each tweet); and Twitter profile information such as display name, username, user bio, and publicly stated location—under a developer agreement. The developer agreement requires that the data are used in ways consistent with people's reasonable expectation of privacy and are not used for developing, creating, or offering commercial services in ways that violate Twitter's policies. To identify relevant tweets, we hand curated a selected set of keywords and hashtags relating to firearms by looking at a random sample of actual tweets and using keywords identified in previous literature. The query we used to collect data from the API included >200 keywords and hashtags (Multimedia Appendix

1). The data retrospectively collected through the API adds a language label to each tweet. In this study, we used tweets labeled as being written in English.

The initial database we derived from the Twitter API using our curated set of firearm-related keywords and hashtags included >2.3 million tweets for 2017 and 2018. More specifically, we obtained 651,466 tweets from 2017 and 1,675,083 tweets from 2018 (with the increase in the number of tweets over time reflecting larger trends in Twitter discussion on the topic). Given that billions of tweets are posted each year in English on Twitter, the discussion of firearms constitutes a relatively modest size.

Next, in *stage 2*, because social media data are subject to the influence of robots, advertisers, and marketers, the data must be classified and filtered to exclude irrelevant data. We used a multistage process to identify and remove spam (advertising, dead links, pornography, etc). We began by detecting spam using a content-based algorithm because spam can be generated by both humans and bots. The content-based algorithm first looks for website URLs related to known advertising, phishing scheme, malware, gambling, and pornography sites. Our spam blacklist contains >2 million website URLs. The second part of the spam classifier looks for content that maps to standard spam content or differs significantly from other content on the tweet stream being collected [25].

In *stage 3*, we randomly sampled tweets from the resulting data for manual labeling—a process of assigning each tweet a set of characteristics, or features, relevant to the study question. We labeled three firearm-related features of tweets: whether the tweet pertains to (1) firearms (2) fatality or fatalities, or (3) a mass shooting. Our analyses focused on firearm fatality (a combination of characteristics 1 and 2) and mass shooting. We also labeled tweets as an advertisement or irrelevant, spam, or noise and used these labels to further improve our spam classifier and remove identified spam tweets from further analysis.

The manual labeling process relied on crowdsourced, distributed labor through Amazon Mechanical Turk (MTurk) [26-28]. We recognize that varying and evolving views exist regarding the use of this platform [29-31] and were attentive to these considerations in our study design, which was vetted and approved by our institutional review board. We applied best practices, creating as clear and streamlined a task as possible and training MTurk coders through a written instruction guide and with labeling examples (Multimedia Appendix 2) [32]. Recent research confirms that MTurk can be a useful resource for quickly gathering reliable data labels for training ML models when best practices are used [33].

We required each tweet to be labeled by 5 different coders, and we calculated the interrater reliability of labeling across coders. The manual labeling process continued until we reached a threshold number (minimum of 400) of tweets that were labeled as positively identifying a particular characteristic. We found that at least 400 tweets for each class in our ML model was reasonable for building a reliable classifier for our learning tasks. The total number of tweets labeled for each characteristic varied because coders may label one or more characteristics for each tweet, rather than all characteristics for each tweet.

As a means of assessing the manual labeling process, we calculated 2 scores for the set of tweets labeled for each characteristic. The first measured task agreement. For each tweet, we assigned the value of the characteristic being measured according to the majority vote (eg, if, of 5 labelers, 3 chose *yes* for firearm-related and 2 chose *no*, we assigned the value of *yes*) and then calculated the percentage of coders who agreed on this value (in this case, 3/5 = 60%). The task agreement is the average across all tweets for a given characteristic of this score. Second, we calculated a worker performance score for each coder in which the denominator was the total number of characteristics a coder labeled, and the numerator was the number of characteristics labeled for which the coder's assigned label aligned with the majority vote. We then calculated the average worker performance score for the set of coders who labeled the set of tweets used for measuring each of the characteristics.

Table 1 summarizes the number of tweets that were manually labeled along with task agreement and worker performance score metrics.

As shown in Table 1, we found high rates of task agreement and worker performance for identifying firearm fatalities (97.14% and 97.19% for task agreement and worker performance, respectively) and mass shooting events (95.42% and 94.96% for task agreement and worker performance, respectively). We noted that 50 tweets that were labeled as being firearm-related were not labeled with a mass shooting characteristic. This occurred in our initial experiment of the labeling task. In this experiment, we labeled tweets as being about a mass shooting, homicide, or suicide. If a tweet was labeled as being about a homicide or suicide, we did not ask the labeler to determine whether the tweet was about a mass shooting. In subsequent experiments, we only focused on capturing firearm-related deaths more broadly and mass shootings explicitly to allow for count adjustments. Therefore, for subsequent experiments (we collected a few hundred labels at a time), we always asked labelers to determine whether a tweet about firearm-related fatality was discussing a mass shooting event.

In *stage 4*, we defined *reliably labeled* tweets as those for which there was manual labeling agreement among ≥3 coders. We dropped tweets that had a reliable label of uncertain or were not reliably labeled from further analysis. This means that our training data did not include ambiguous tweets and, therefore, may undercount our characteristics.

The next three stages (*stages 5*, *6*, and *7*) of the process involved firearm-related ML. In *stage 5*, we divided the subset of reliably labeled tweets into training data—on which we built ML classifiers—and holdout data, which were used to validate the classifiers. We randomly selected 80% of reliably labeled tweets for the training data and 20% for the holdout. When building the ML classifiers, we used 5-fold cross-validation on the training data to measure the reliability of the classifiers. Cross-validation is a resampling procedure that allows researchers to determine whether their ML models are generalizable [34-36]. In 5-fold cross-validation, the data set is partitioned into 5 equal subparts (or *folds*). Of the 5 folds, 4

(80% of the data) are used for training, and 1 (20% of the data) is used for testing. This is repeated 5 times so that each fold is part of the training set 4 times and part of the testing set 1 time, and the final accuracy of the model is determined by taking the mean accuracy of all the created models on the testing set.

We began building ML classifiers to identify tweets pertaining to a firearm fatality and to a mass shooting. We minimally preprocessed the data: lowercased text, removed punctuation and URLs, and removed stopwords. We generated a number of features for the ML classifiers: frequent n-grams, words and phrases, and sentiment. The classifiers we compared were random forest, support vector machine, logistic regression, decision tree, and naïve Bayes. In *stage 6*, we validated the classifiers we developed for firearm fatalities and mass shootings in *stage 5* by further testing them on holdout data. We calculated the sensitivity and specificity of the ML model predictions against those of the manually coded firearm-fatality label.

Table 2 summarizes the best-performing ML classifier for each classification task along with the training and holdout data set sizes and a measure of reliability based on the testing data, using our cross-validation approach, and the holdout data. The F_1-score is a weighted average of sensitivity and specificity (precision and recall) that considers both false positives and false negatives. For firearm-related fatality, we had 6045 labeled tweets. For mass shooting, we had 5842 labeled tweets. Because of heavy skews (imbalance) in the training data, we randomly undersampled from the labeled data of the majority label to balance the training and holdout data sets. Table 2 shows the training and holdout data set sizes after this procedure.

We selected random forest classifiers for both firearm fatalities and mass shooting characteristics. The F_1-scores, as shown in Table 2, are high and comparable for the testing and holdout data, indicating a clear ability of the classifiers to generalize beyond the training data set.

Stage 7 completed the development of our Twitter-based gun-related analytic platform with the third and final piece of the ML analysis. In *stage 7*, we applied the validated classifier to identify firearm-fatality tweets.

Table 1. Manually labeled tweet characteristics.

Tweet label	Firearm-related fatality	Mass shooting
Total number of tweets labeled (yes, no, unsure)	5868 (5528, 330, 10)	5478 (419, 5056, 3)
Task agreement, %	97.14	95.42
Worker performance score, %	97.19	94.96

Table 2. Machine learning (ML) classifier type and reliability.

Prediction task	Firearm-related fatality	Mass shooting
Training data size, n	1142	1038
Holdout data size, n	286	256
Best ML classifier	Random forest	Random forest
F_1-score: cross-validation, test data, mean (SD)	0.91 (0.017)	0.88 (0.012)
F_1-score: holdout data	0.90	0.88

Geographic Area Estimation of Twitter Firearm-Fatality Discussion Volume

The NVSS classifies fatalities according to the geographic jurisdiction in which the fatality occurred. Thus, for comparison with the state-level NVSS estimates, the *location of the fatality* being discussed on Twitter is the location of interest (vs the location of the individual who is tweeting). We relied on the tweet content to identify the location of the fatality because location information from either profile information or tweet geocoding (which some users permit) identifies the location of the user (as opposed to location of the fatality).

Importantly, location mentions in tweets primarily refer to city names. In some cases, state name is also mentioned, whereas in other cases, state can be inferred from the city name. To obtain a reasonably sufficient number of tweets per location for estimating area-level fatality discussion volume, we focused on identifying the larger cities mentioned in tweets. Specifically, we identified tweets in our sample that mentioned any of the 250 most populous cities in 2018 (based on US Census data [35]). A limitation of this approach is that it focuses on fatalities in urban areas rather than in rural areas.

We augmented the list of 250 city names with alternative city names commonly used on social media, such as *nyc*, and with city names that contain no spaces between multiple words, such as *sanfrancisco*. We standardized posts—converting the text to lowercase and removing URLs, user mentions (words prefixed with @), and common phrases that may look as though they are city mentions when they are not. An example of a common phrase we removed is *drag queens* because it may be accidently mapped to Queens, New York City, New York. After standardization, we searched the text for city names that matched our location ontology. The majority of city names among the 250 are associated with, and can thus be reliably mapped to, a single state. For our specific set of tweets, there were no cities mentioned that mapped to multiple states.

We summed tweet discussion volume across the most populous cities within a state to create a state-level measure. We constructed state-level estimates for Arizona, California, Colorado, Florida, Georgia, Illinois, Indiana, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Missouri, Nevada, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Tennessee, Texas, Virginia, Washington, and Wisconsin. We excluded from further analysis those states for which the sample size of tweets was <200 tweets after the mass shooting adjustments (described in the next paragraph) because they are home to only one or only a few of the more populous cities (eg, Idaho, Iowa, Nebraska, and Oregon), and the populous cities in the state are relatively small (eg, Kansas, Alabama, and Arkansas).

The resulting data set, after applying the best ML classifier to identify firearm-fatality tweets and identifying the state of the fatality using location mentions, included 31,747 tweets from 2017 and 44,779 from 2018. We summarized firearm-fatality discussion volume for each state using these data. We then adjusted the state-level estimates of firearm-fatality discussion volume in 3 ways. First, mass shooting events tend to generate disproportionately high levels of discussion, that is, levels of discussion that are far higher than for other less high-profile fatalities. We accounted for the potential distorting influence of mass shooting events on the relationship between a gun fatality and tweet discussion volume by excluding tweets from the location of mass shooting events for a period of 1 week after the event. We based the 1-week exclusion period on observed trends in mass shooting discussion volume. We identified mass shooting events during the time frame of our data using information from the Gun Violence Archive [37], Everytown Research [38], and The Violence Project [39]. Finally, we adjusted our estimates of state-level discussion volume by the percentage of the state-level population that uses Twitter [40].

We tested for serial correlation and found that the NVSS data contained 10 states in our final data set with some moderate serial correlation, and the Twitter data contained 5 states with moderate serial correlation. For this reason, we made each time series stationary by differencing monthly estimates [41]; we refer to this as the *Change* result. For the level correlation, we removed states in which both time series had higher levels of serial correlation because the correlation is valid if one of the time series exhibits serial correlation and the other does not [42]. This issue arose with four states: Georgia, Indiana, Michigan, and North Carolina.

Results

Correlation Analysis

Table 3 shows results from our correlation analysis. We estimated the correlation within state by month between the *level* of firearm-fatality discussion volume and the *level* of NVSS-reported fatalities, as well as the correlation within each state in the monthly *change* in discussion volume versus the monthly *change* in the NVSS fatality rate.

Table 3. Results of correlation analysisa.

	Discussion volume	Discussion volume adjusted for average state-level Twitter use
LEVEL: Correlation, range	–0.293 to 0.535	–0.289 to 0.537
LEVEL: Correlation, mean; median	0.085; 0.091	0.087; 0.093
CHANGE: Correlation, range	–0.057 to 0.682	–0.059 to 0.688
CHANGE: Correlation, mean; median	0.313; 0.303	0.312; 0.301

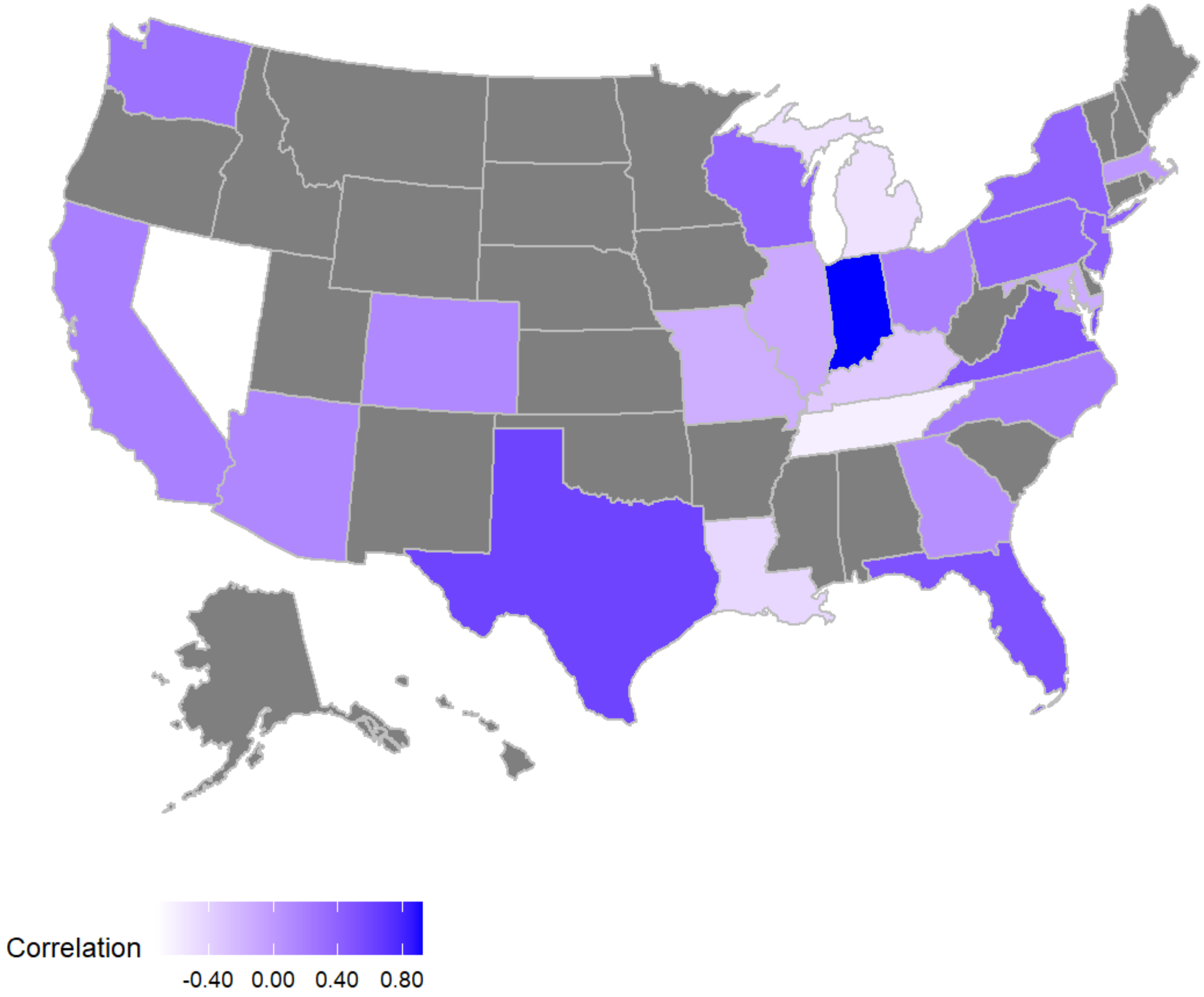
[a] Pearson linear correlations are reported.

The correlation between the monthly level of firearm-fatality tweets and the monthly number of fatalities measured by the NVSS is weak (median 0.081) and widely dispersed across states (range –0.31 to 0.54). The correlation between month-to-month changes in firearm fatalities discussed in tweets versus those estimated in the NVSS is moderate (median 0.30) and exhibited less dispersion among states than the monthly level correlations (range –0.057 to 0.68). For the correlation among month-to-month changes in firearm fatalities, almost half (11/24, 46%) of the states have correlations ranging from 0.1 to 0.4. More than a quarter (7/24, 29%) of the states have correlations below this range, and a quarter (6/24, 25%) have correlations above this range. The results for the adjusted discussion volume (second row of Table 3, discussion volume adjusted for Twitter use in the state) are very similar to the unadjusted results, with negligible differences observed in estimated correlation rates.

Figure 2 provides additional details for the correlation in monthly changes in fatality discussion volume and NVSS-estimated fatalities, with a depiction of state-by-state (adjusted) correlation rates for 2017. White-shaded states have no correlation. The darker the purple shade of a state, the higher the correlation. The gray-shaded states are those for which we were not able to estimate a Twitter fatality discussion rate (refer to the Twitter Data subsection under Methods). Not unexpectedly, the strength of the correlation seems to be related to the percentage of the state's population living in one of the most populous cities that we use in our location ontology; for example, one-third or more of the state population in Texas, New York, and Arizona reside in one of the top 100 most populated cities in the state (34%, 44%, and 51%, respectively). These states exhibit some of the highest correlation rates between monthly fluctuation in firearm-fatality discussion volume and NVSS-based fatality estimates. Likewise, Georgia, Michigan, and Maryland are among the states with both the lowest percentage of their population living in more populous cities (5%, 7% and 10%, respectively) and have some of the lowest rates of correlation among the states studied.

By contrast, 47% of Nevada residents live in one of the most populous cities, but the correlation rate in Nevada falls into a lower tier than the correlation rates in Texas, New York, and Arizona. Although we adjusted for mass shooting discussion volume by removing tweets from the week after such an event, the lower correlation observed in Nevada suggests that the adjustment may have been insufficient for capturing the extent of discussion volume distortion in the wake of the mass shooting event in Las Vegas, given the magnitude of the event. Analyzing the data in more detail shows that discussion of this shooting returns at anniversaries (1 year) and when other larger mass shootings occur in other parts of the country, identifying a need for a more extensive adjustment for historically large mass shootings.

Figure 2. Correlation by state between change in firearm–fatality tweets and change in National Vital Statistics System–estimated firearm fatalities in 2017.



Comparison of Correlations

Furthermore, a comparison of correlations for each state in 2017 versus 2018 shows that states with the largest cities tend to have the most stable correlations; for example, Texas, New York, California, Florida, and Ohio; whereas states with fewer large cities and fewer tweets tend to have higher variation in their correlation estimates; for example, Missouri, Tennessee, South Carolina, and Maryland. An additional factor that is likely to affect the correlation rate is the location within the state of firearm fatalities. To the extent that fatalities within a state are more concentrated in the most populated cities, the correlation between NVSS-estimated fatalities and Twitter discussion volume is expected to be higher.

Discussion

Principal Findings

Among the subset of states studied, we found weak-to-moderate correlation between our measure of the level of firearm-fatality tweets and the NVSS-based estimates of the level of firearm fatalities and higher moderate correlation in measures of the month-to-month changes in firearm-fatality tweets and estimated fatalities. As our ontology for Twitter location mentions relies on identification of the 250 most populous cities, our correlation is higher in states in which more of the state’s population was living in one of these cities. We further expected the correlation to be higher in areas where firearm fatalities were concentrated in the most populated cities and found suggestive evidence regarding this point.

A key limitation of this analysis is that we relied on tweets from more populated cities to develop a state-level estimate of discussion volume. Our approach reflected, dually, the limited availability of firearm-fatality data at the city level and the limited availability of location identifiers for tweets. An important feature of this analysis was the need to identify the location of the event being discussed versus the location of the user. In the case of the latter, geocoding of the user profile is advantageous and can provide a state-level identifier, but the former relies only on location mentions within the tweet.

Even with these limitations, the correlation capturing fluctuation in firearm mortality is moderate. We view this as a promising signal for the potential of social media data to provide meaningful information on gun-related outcomes in the future. More specifically, our findings suggest that Twitter data may hold particular value for tracking dynamics in gun-related outcomes. In addition, for location-specific firearm-related outcomes, the data are most valuable for understanding dynamics in relatively populous cities that are identifiable through location mentions in tweet content. Finally, the data are likely to be particularly valuable for understanding firearm outcomes not currently measured, not measured well, or not measurable through other available means. A key advantage of Twitter data is the continuous, automated, and near–real-time monitoring they provide [13]. Once big data infrastructure has been invested in, the data can be relatively easily processed. The initial cost of big data infrastructure can be high if researchers want to stream data for large periods of time. However, for a single study, researchers who can access a server should be able to conduct the analysis at a low cost. Because of this potentially higher investment, we have developed a text analytic portal that allows researchers to construct variables from our social media data [43], thereby enabling future research with these data without the cost of setting up big data infrastructure.

We recognize the need for additional analyses to continue to adapt and extend upon the approach developed and applied in this research, including, for example, work that assesses the reliability of associations over longer time periods. We also note that, unlike survey data that are sampled to be representative of the underlying population, social media data emanate from those who use a particular platform. Although the use of Twitter in the United States is significant (in 2021, nearly a quarter of adults reported using Twitter, and among those who reported using the platform, nearly half said that they use it once a day or more) [44], it is nonetheless also true that rates of social media use are correlated with age and to some extent with other demographic characteristics [44]. Much of the existing analytic work with social media data does not directly deal with this issue. In our approach, we adjusted our estimates for the percentage of Twitter users in each state. Additional statistical adjustments that more completely account for engagement with the platform are important for future work. Furthermore, social media data include limited sociodemographic information about users. Additional methodological strides toward developing robust methods for demographic imputations represent an important dimension of future efforts.

Usefulness of Social Media Data

The Centers for Disease Control and Prevention [45] describes its public health approach to prevention of violence, including firearm violence, as encompassing four steps: defining and monitoring the problem, identifying risk and protective factors, developing and testing prevention strategies, and assuring widespread adoption [46]. For firearm violence, the first step—building a foundation of information for describing the epidemiology of such violence—requires focused resources and development. In addition to recent developments in survey, administrative, and other data, such as the important efforts by news media and other organizations to track gun violence incidents in significant detail and the advent of data scraping from obituaries [37,47,48], social media data are a promising future source. This research provides an important building block for future work that continues to develop the usefulness of social media data, alone or in conjunction with other data resources, to strengthen the information base on which firearm research relies, and, more generally, contributes to the process of integrating emerging big data algorithms and traditional data sources for behavioral understanding, decision support, and evidence-based public policy.

As we build out the power of social media data for informing public health problems such as firearm violence, several important dimensions need to be kept in mind. The role that social media may play in exacerbating gun violence or spreading trauma related to gun violence cannot be ignored. However, these data can also be used to help target and improve our understanding of those who use guns and allow for new approaches to gun violence–prevention interventions [49]. To use these data to improve public health outcomes and our understanding of human beliefs and behaviors, we must spearhead establishing best practices for using social media data in ethical ways [50-52], as well as understanding representativeness, methodological limitations, and algorithmic biases.

Acknowledgments

This project was conducted at the Massive Data Institute (MDI) at Georgetown University, thanks to funding by the National Collaborative on Gun Violence Research and the MDI. The authors also thank Rebecca Vanarsdall and the MDI technical team for their support.

Authors' Contributions

LS and CRG conceptualized the project, led the team, and wrote the paper. YW and SH conducted the main analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hashtags, keywords, and phrases used to collect data from the Twitter application programming interface.
[PDF File (Adobe PDF File), 32 KB - jmir_v24i8e38319_app1.pdf]

Multimedia Appendix 2
Instructions provided to Amazon Mechanical Turk coders.
[PDF File (Adobe PDF File), 39 KB - jmir_v24i8e38319_app2.pdf]

References

1. Fast Facts: Firearm Violence Prevention. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/violenceprevention/firearms/fastfact.html; [accessed 2022-06-27]
2. Gun violence is a public health crisis [Fact sheet]. American Public Health Association (APHA). 2021. URL: https://www.apha.org/-/media/Files/PDF/factsheets/200221_Gun_Violence_Fact_Sheet.ashx [accessed 2022-06-27]
3. Wintemute GJ. The epidemiology of firearm violence in the twenty-first century United States. Annu Rev Public Health 2015 Mar 18;36:5-19. [doi: 10.1146/annurev-publhealth-031914-122535] [Medline: 25533263]
4. Morrall AR, Ramchand R, Smart R, Gresenz CR, Cherney S, Nicosia N, et al. The Science of Gun Policy: A Critical Synthesis of Research Evidence on the Effects of Gun Policies in the United States. Santa Monica, CA, USA: RAND Corporation; 2018.
5. Smart R, Morral AR, Smucker S, Cherney S, Schell TL, Peterson S, et al. The Science of Gun Policy: A Critical Synthesis of Research Evidence on the Effects of Gun Policies in the United States. 2nd Edition. Santa Monica, CA, USA: RAND Corporation; 2020.
6. The State of Firearms Data in 2019: First Report of the Expert Panel on Firearms Data Infrastructure. NORC at the University of Chicago. 2020 Jan. URL: https://www.norc.org/PDFs/Firearm%20Data%20Infrastructure%20Expert%20Panel/State%20of%20Firearms%20Research%202019.pdf [accessed 2022-06-27]
7. Stark DE, Shah NH. Funding and publication of research on gun violence and other leading causes of death. JAMA 2017 Jan 03;317(1):84-85. [doi: 10.1001/jama.2016.16215] [Medline: 28030692]
8. Mayer-Schönberger V, Cukier K. Big Data: A Revolution that will Transform How We Live, Work, and Think. Boston, MA, USA: Houghton Mifflin Harcourt; 2013.
9. Dredze M. How social media will change public health. IEEE Intell Syst 2012 Jul;27(4):81-84. [doi: 10.1109/mis.2012.76]
10. Stieglitz S, Dang-Xuan L, Bruns A, Neuberger C. Social media analytics. Bus Inf Syst Eng 2014 Feb 15;6(2):89-96. [doi: 10.1007/s12599-014-0315-7]
11. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. Proc Int AAI Conf Web Soc Media 2013;7(1):128-137.
12. Martin S, Singh L. Data analytics and displacement: using big data to forecast mass movement of people. In: Maitland C, editor. Digital Lifeline?: ICTs for Refugees and Displaced Persons. Cambridge, MA, USA: MIT Press; 2018:185-206.
13. Singh L, Traugott M, Bode L, Budak C, Davis-Kean PE, Guha R, et al. Data Blending: Haven’t We Been Doing This for Years? Massice Data Institute, Georgetown University. 2020 Apr. URL: https://mccourt.georgetown.edu/wp-content/uploads/2022/02/MDI-Data-Blending-White-Paper-April2020_ACCESSIBLE.pdf [accessed 2022-06-27]
14. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
15. Ladd J, Ryan R, Singh L, Bode L, Budak C, Conrad F, et al. Measurement considerations for quantitative social science research using social media data. PsyArXiv 2020 Dec. [doi: 10.31234/osf.io/ga6nc]
16. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
17. Wang N, Varghese B, Donnelly PD. A machine learning analysis of Twitter sentiment to the Sandy Hook shootings. In: Proceedings of the IEEE 12th International Conference on e-Science. 2016 Presented at: e-Science '16; October 23-27, 2016; Baltimore, MD, USA p. 303-312. [doi: 10.1109/escience.2016.7870913]
18. Starbird K. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. Proc Int AAI Conf Web Soc Media 2017;11(1):230-239.
19. Lin YR, Chung WT. The dynamics of Twitter users’ gun narratives across major mass shooting events. Humanit Soc Sci Commun 2020 Aug 03;7(1):46. [doi: 10.1057/s41599-020-00533-8]
20. Blankenship M, Graham C. Assessing the social and emotional costs of mass shootings with Twitter data. Brookings. 2021 May 5. URL: https://www.brookings.edu/blog/up-front/2021/05/05/assessing-the-social-and-emotional-costs-of-mass-shootings-with-twitter-data/ [accessed 2022-06-27]

21. Jordan L, Kalin J, Dabrowski C. Characteristics of gun advertisements on social media: systematic search and content analysis of Twitter and YouTube posts. J Med Internet Res 2020 Mar 27;22(3):e15736 [FREE Full text] [doi: 10.2196/15736] [Medline: 32217496]
22. Wasike B. Persuasion in 140 characters: testing issue framing, persuasion and credibility via Twitter and online news articles in the gun control debate. Comput Human Behav 2017 Jan;66:179-190. [doi: 10.1016/j.chb.2016.09.037]
23. Benton A, Hancock B, Coppersmith G, Ayers JW, Dredze M. After Sandy Hook Elementary: a year in the gun control debate on Twitter. arXiv 2016 Oct 6.
24. Leading countries based on number of Twitter users as of October 2021. Statistics. 2022. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 2022-06-27]
25. Kawintiranon K, Singh L, Budak C. Traditional and context-specific spam detection in low resource settings. Mach Learn 2022 Jun 09;111(7):2515-2536. [doi: 10.1007/s10994-022-06176-x]
26. Litman L, Robinson J. Conducting Online Research on Amazon Mechanical Turk and Beyond. Thousand Oaks, CA, USA: Sage Publications; 2020.
27. Schnoebelen T, Kuperman V. Using Amazon Mechanical Turk for linguistic research. Psihologija 2010;43(4):441-464. [doi: 10.2298/psi1004441s]
28. Kees J, Berry C, Burton S, Sheehan K. An analysis of data quality: professional panels, student subject pools, and Amazon's Mechanical Turk. J Advert 2017 Jan 23;46(1):141-155. [doi: 10.1080/00913367.2016.1269304]
29. Semuels A. The Internet Is Enabling a New Kind of Poorly Paid Hell. The Atlantic. 2018 Jan 23. URL: https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/ [accessed 2022-06-27]
30. Newman A. I Found Work on an Amazon Website. I Made 97 Cents an Hour. The New York Times. 2019 Nov 5. URL: https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html [accessed 2022-06-27]
31. Moss AJ, Rosenzweig C, Robinson J, Jaffe SN, LItman L. Is it ethical to use Mechanical Turk for behavioral research? Relevant data from a representative survey of MTurk participants and wages. PsyArXiv 2020 Apr 28. [doi: 10.31234/osf.io/jbc9d]
32. Amazon Mechanical Turk. Requester best practices guide. Amazon Web Services. 2011. URL: https://mturkpublic.s3.amazonaws.com/docs/MTURK_BP.pdf [accessed 2022-06-27]
33. Singh L, Vanarsdall R, Wang Y, Gresenz C. Students or Mechanical Turk: who are the more reliable social media data labelers? In: Proceedings of the 11th International Conference on Data Science, Technology and Applications. 2022 Presented at: DATA '22; July 11-13, 2022; Lisbon, Portugal p. 408-415. [doi: 10.5220/0011278600003269]
34. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edition. New York, NY, USA: Springer; 2009.
35. Metropolitan and micropolitan statistical areas population totals and components of change: 2010-2019. United States Census Bureau. 2021. URL: https://www.census.gov/data/datasets/time-series/demo/popest/2010s-total-metro-and-micro-statistical-areas.html [accessed 2022-06-27]
36. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial intelligence - Volume 2. 1995 Presented at: IJCAI '95; August 20-25, 1995; Montreal, Canada p. 1137-1143.
37. Gun violence archive Database. Gun Violence Archive (GVA). 2022. URL: https://www.gunviolencearchive.org/about [accessed 2022-06-27]
38. Mass Shootings in America. Everytown Research. 2021. URL: https://everytownresearch.org/maps/mass-shootings-in-america/ [accessed 2022-06-27]
39. Peterson JK, Densley JA. The violence project database of mass shootings in the United States, 1966-2019. The Violence Project. 2019. URL: https://www.theviolenceproject.org [accessed 2022-06-27]
40. Jackson M, McKinstry J, McPhee C, Raghunathan T, Singh L, Traugott M, et al. Understanding Who Uses Twitter: State level estimates of those on Twitter. MOSAIC Methods Brief. Measuring Online Social Attitudes and Information Collaborative, University of Michigan. 2021 Nov. URL: https://mosaic.mdi.georgetown.edu/wp-content/uploads/2021/11/MOSAIC-Methods-Brief-State-Level-Twitter-Use_11092021.pdf [accessed 2022-06-27]
41. Sims CA. Bayesian skepticism on unit root econometrics. J Econ Dyn Control 1988 Jun;12(2-3):463-474. [doi: 10.1016/0165-1889(88)90050-4]
42. Dean RT, Dunsmuir WT. Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: the importance of constructing transfer function autoregressive models. Behav Res Methods 2016 Jun;48(2):783-802. [doi: 10.3758/s13428-015-0611-2] [Medline: 26100765]
43. Singh L, Padden C, Davis-Kean P, David R, Marwadi V, Ren Y, et al. Text analytic research portals: supporting large-scale social science research. In: Proceedings of the 2021 IEEE International Conference on Big Data. 2021 Presented at: BigData '21; December 15-18, 2021; Orlando, FL, USA p. 6020-6022. [doi: 10.1109/BigData52589.2021.9671696]
44. Social media fact sheet. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2022-06-27]
45. The public health approach to violence prevention. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/violenceprevention/about/publichealthapproach.html [accessed 2022-06-27]

46. Dahlberg LL, Krug EG. Violence: a global public health problem. In: Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R, editors. World Report on Violence and Health. Geneva, Switzerland: World Health Organization; 2002:1-21.
47. Fatal Force. The Washington Post. 2022. URL: https://www.washingtonpost.com/graphics/investigations/police-shootings-database/ [accessed 2022-06-27]
48. Schnell R, Redlich S. Web scraping online newspaper death notices for the estimation of the local number of deaths. In: Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies. 2019 Presented at: BIOSTEC '19; February 22–24, 2019; Prague, Czech Republic p. 319-325. [doi: 10.5220/0007382603190325]
49. Patton DU, McGregor K, Slutkin G. Youth gun violence prevention in a digital age. Pediatrics 2018 Apr;141(4):e20172438 [FREE Full text] [doi: 10.1542/peds.2017-2438] [Medline: 29496903]
50. Singh L, Polyzou A, Wang Y, Farr J, Gresenz C. Social Media Data - Our Ethical Conundrum. A Quarterly bulletin of the IEEE Computer Society Technical Committee on Database Engineering. National Science Foundation. 2020 Dec. URL: http://sites.computer.org/debull/A20dec/p23.pdf [accessed 2022-06-27]
51. Bowen SA. Using classic social media cases to distill ethical guidelines for digital engagement. J Mass Media Ethics 2013 Apr;28(2):119-133. [doi: 10.1080/08900523.2013.793523]
52. Leonelli S, Lovell R, Wheeler BW, Fleming L, Williams H. From FAIR data to fair data use: methodological data fairness in health-related social media research. Big Data Soc 2021 May 03;8(1):205395172110103. [doi: 10.1177/20539517211010310]

Abbreviations

API: application programming interface
MDI: Massive Data Institute
ML: machine learning
MTurk: Amazon Mechanical Turk
NVSS: National Vital Statistics System

Edited by R Kukafka; submitted 29.03.22; peer-reviewed by Y Kim, B Weichelt, L Jordan; comments to author 15.05.22; revised version received 28.06.22; accepted 15.07.22; published 25.08.22.

Please cite as:
Singh L, Gresenz CR, Wang Y, Hu S
Assessing Social Media Data as a Resource for Firearm Research: Analysis of Tweets Pertaining to Firearm Deaths
J Med Internet Res 2022;24(8):e38319
URL: https://www.jmir.org/2022/8/e38319
doi:10.2196/38319
PMID:36006693

Original Paper

Effects of Video-Based Patient Education and Consultation on Unplanned Health Care Utilization and Early Recovery After Coronary Artery Bypass Surgery (IMPROV-ED): Randomized Controlled Trial

Gijs van Steenbergen[1], MD; Dennis van Veghel[1], PhD; Dideke van Lieshout[2], MA; Merel Sperwer[2], MSc; Joost ter Woorst[1], MD, PhD; Lukas Dekker[1,3], MD, PhD

[1]Cardiothoracic Surgery Department, Catharina Heart Centre, Catharina Hospital, Eindhoven, Netherlands

[2]Dutch Heart Foundation, The Hague, Netherlands

[3]Department of Biomedical Technology, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Gijs van Steenbergen, MD
Cardiothoracic Surgery Department
Catharina Heart Centre
Catharina Hospital
Michaelangelolaan 3
PO Box 1350
Eindhoven, 5602
Netherlands
Phone: 31 040 239 9111
Email: gijs.v.steenbergen@catharinaziekenhuis.nl

Abstract

Background: Health care utilization after coronary artery bypass graft (CABG) surgery is high and is partly of an unplanned nature. eHealth applications have been proposed to reduce care consumption, which involve and assist patients in their recovery. In this way, health care expenses could be reduced and quality of care could be improved.

Objective: The aim of this study was to evaluate if an eHealth program can reduce unplanned health care utilization and improve mental and physical health in the first 6 weeks after CABG surgery.

Methods: A single-blind randomized controlled trial was performed, in which patients scheduled for nonacute CABG surgery were included from a single center in the Netherlands between February 2020 and October 2021. Participants in the intervention group had, alongside standard care, access to an eHealth program consisting of online education videos and video consultations developed in conjunction with the Dutch Heart Foundation. The control group received standard care. The primary outcome was the volume and costs of a composite of unplanned health care utilization, including emergency department visits, outpatient clinic visits, rehospitalization, patient-initiated telephone consultations, and visits to a general practitioner, measured using the Medical Technology Assessment Medical Consumption Questionnaire. Patient-reported anxiety and recovery were also assessed. Intention-to-treat and "users-only" analyses were used.

Results: During the study period, 280 patients were enrolled and randomly allocated at a 1:1 ratio to the intervention or control group. The intention-to-treat analysis consisted of 136 and 135 patients in the intervention and control group, respectively. At 6 weeks, the primary endpoint had occurred in 43 of 136 (31.6%) patients in the intervention group and in 61 of 135 (45.2%) patients in the control group (hazard ratio 0.56, 95% CI 0.34-0.92). Recovery was faster in the intervention group, whereas anxiety was similar between study groups. "Users-only" analysis yielded similar results.

Conclusions: An eHealth strategy comprising educational videos and video consultations can reduce unplanned health care utilization and can aid in faster patient-reported recovery in patients following CABG surgery.

Trial Registration: Netherlands Trial Registry NL8510; https://trialsearch.who.int/Trial2.aspx?TrialID=NL8510

International Registered Report Identifier (IRRID): RR2-10.1007/s12471-020-01508-9

(J Med Internet Res 2022;24(8):e37728) doi:10.2196/37728

KEYWORDS

e-Health; eHealth; digital health; patient education; coronary artery bypass surgery; cardiac surgery; health care utilization; costs; cost; economic; coronary; cardiology; heart; surgery; bypass; RCT; randomized controlled trial; video consultation; telehealth; telemedicine; patient-reported; recovery; expense

Introduction

Coronary artery bypass graft (CABG) surgery is one of the most frequently performed cardiac surgeries in the world, which is generally performed with good outcomes and relatively low 30-day mortality (~1.5%) [1]. In more recent years, the care chain for patients undergoing CABG surgery has been demonstrated to increase efficiency and reduce costs. As a result, the duration of hospitalization has decreased substantially, with patients discharged on the 7th postoperative day (mean). These efficiency-driven early discharge protocols require more self-management skills among patients. Early discharge reduces the time physicians can spend with their patients in the direct postoperative phase in spite of the well-known benefit of patient counseling and guidance through recovery [2,3].

After discharge, patients commonly experience anxiety or uncertainty about symptoms or appropriate physical exercise [4]. These issues are typically addressed during hospitalization; however, after discharge, patients' recall of information is often incomplete and they do not always know who to address with questions [4]. The advantages of a shortened hospital stay might therefore be counterbalanced by preventable unplanned health care utilization, especially since planned care is not initiated until several weeks after discharge. At present, nearly 1 in 7 patients are readmitted in the first 30 days after discharge for noncardiac causes and roughly 15% of patients visit the emergency department within 1 month after CABG surgery [5-8]. It was estimated that potentially preventable readmissions following CABG surgery cost Medicare US $151 million in 2005, placing a significant burden on society [7]. With the expected increase in the number of future patients undergoing CABG surgery, this is a pressing issue urging evaluation and a potential redesign of postoperative follow-up.

eHealth is defined by the World Health Organization as "the cost-effective and secure use of information and communication technologies in support of health and health-related fields," which encompasses multiple digital interventions that can aid in the delivery of patient-centered care and postoperative patient guidance, thereby potentially reducing unplanned health care utilization [9]. eHealth strategies have been successfully applied in postoperative follow-up in various forms, which have been shown to improve patient outcomes, speed recovery, and reduce health care utilization in various surgical populations [10]. In addition, eHealth has proven to be of value for patients to enhance their self-management through better understanding of their disease, increased independence, and improved acceptance to adhere to lifestyle advice [3,11]. However, experience with eHealth in patients following CABG surgery is limited, and it remains unclear if eHealth strategies would be effective in this population.

The objective of this trial was to fill this knowledge and experience gap. We hypothesized that restructuring the postoperative period with an eHealth strategy will reduce unplanned health care utilization through improved mental and physical health and faster recovery.

Methods

Trial Design

The IMPROV-ED trial was a randomized controlled trial (RCT) performed between February 2020 and December 2021 at Catharina Hospital in the Netherlands. A detailed study protocol was published prior to enrollment of the first study participant [11]. No changes were made to the study protocol between publication and initiation of the trial. The trial is reported using the CONSORT (Consolidated Standards of Reporting Trials) checklist for RCTs [12].

Ethics Considerations

The study was approved by the medical ethics committee (R19.100) and was registered in the Netherlands Trial Registry (NL8510). Written informed consent was obtained from all patients who met the inclusion criteria and were willing to participate.

Participants

To minimize selection bias, all patients on the waiting list for isolated CABG surgery over 18 years of age were contacted by telephone and informed about the study by one of the investigators. Patients were eligible for participation if they had access to a computer/tablet/smartphone with internet connection and a webcam/built-in camera; had sufficient knowledge of the use of internet and email (assistance was allowed); and were able to speak, read, and interpret the Dutch language. The eHealth strategy would not be applicable to patients who did not comply with these inclusion criteria and they were therefore not eligible for participation. At inclusion, patients were randomized 1:1 to the intervention or control group using a block size of 4. A certified program was used for sequence generation and randomization (Research Manager). When a patient was randomized but no longer qualified for the inclusion criteria or was lost to follow-up, the patient was excluded from further follow-up and analysis.

Interventions

Patients randomized to the control group received standard postoperative care, comprising planned outpatient follow-up by their cardiothoracic surgeon at 6-8 weeks after discharge and a cardiac rehabilitation program supervised by cardiologists with outpatient follow-up starting between 4 and 8 weeks after surgery. As a result of the COVID-19 health crisis and the measures taken by the government, most of these contacts were telephone consultations (TCs) rather than physical consultations.

Patients randomized to the intervention group had access to the eHealth strategy in addition to standard care. The eHealth strategy comprised web-based educational videos developed by the Dutch Heart Foundation and two postoperative video consultations (VCs) with a physician from the department of cardiothoracic surgery at 1 and 3 weeks after discharge.

Upon randomization, patients in the intervention group received access to the educational videos via a link sent by email. The same link was sent via email again at discharge. By clicking the link, patients were referred to a hidden (for nonparticipants and the control group) part of the website from the Dutch Heart Foundation that contained the educational videos. The content of the educational videos was constructed and validated by physicians and patient representatives prior to the trial. Based on these evaluations and a scoping review of the literature on delivery of information to patients with varying degrees of health literacy, the full content was delivered to patients at inclusion instead of by fragmentized access to videos applicable to the patient's situation [13]. Nevertheless, to prevent cognitive overload in patients with low health literacy, educational videos were categorized in three categories: treatment (10 videos with information on the surgery and how to prepare for admission), recovery (6 videos about what to expect in the postoperative course and when to contact a physician), and healthy living (2 videos on cardiovascular risk management, including smoking cessation, weight reduction, cholesterol management, and exercise). The videos were delivered in spoken text supported by animations for optimal health communication to patients with low and adequate health literacy [13]. Usage data were extracted from the web log for evaluation purposes. Educational videos were available to patients in the intervention group throughout the trial (ie, not only when the link was sent). See the published study protocol for an illustrative overview of the educational videos [11].

VCs were conducted by a nurse practitioner or junior doctor using Microsoft Teams. The dates for VCs were sent to patients by email at discharge. On the day of the VC, patients received an email with a link providing access to the VC. The VC was not scheduled on the same day as routine outpatient follow-up. During the VCs, patients were questioned about their recovery and physical and mental complaints. The sternotomy wound was visually inspected. Patients who required physical examination or diagnostic tests based on the VC were instructed to visit the general practitioner or emergency department, or were scheduled for early outpatient follow-up (within 1 week) at discretion of the physician. The nurse practitioner/junior doctor who conducted the VCs was blinded to the study's objectives and outcomes. Study participants were not blinded. If a VC was unexpectedly not possible (eg, due to unforeseen connection errors, problems with hardware, technical issues), the VC was replaced by a TC. Reasons for replaced VC were reported.

Outcomes

The primary outcomes of the IMPROV-ED trial were the volume and costs of unplanned health care utilization defined by a composite of all emergency department visits, outpatient clinic visits, rehospitalization, patient-initiated TCs, and visits to a general practitioner, as measured by the Institute for Medical Technology Assessment Medical Consumption Questionnaire (iMCQ) at the 6-week follow-up [14]. Cross-validation with the patients' reported health care utilization was performed by contacting their health care providers. The secondary outcomes were the individual unplanned health care activities, and a composite of planned and unplanned in-hospital care (emergency department visits, outpatient clinic visits, rehospitalization, and patient-initiated TCs) and planned and unplanned primary care (consultations with a general practitioner, allied health professionals, psychologists) at 6 weeks. The other secondary outcomes were the patients' self-reported physical and mental health, as measured with the Hospital Anxiety and Depression Scale (HADS) and Recovery Index-10 (RI-10) questionnaires [15,16].

Data Collection

All patients received questionnaires at inclusion (anxiety subscale of the HADS), at discharge (HADS and RI-10), 1 week after discharge (HADS and RI-10), 2 weeks after discharge (HADS and RI-10), and 6 weeks after discharge (HADS, RI-10, and iMCQ). Only the anxiety subscale from the HADS was used. A higher score indicated more symptoms of anxiety (HADS maximum score 21) or favorable progress of recovery (RI-10 maximum score 50). The iMCQ resulted in absolute frequencies of visits for the questioned care activities. Patients in the intervention group also received a self-made questionnaire to evaluate the eHealth strategy and to question them about the use of the education videos (see Figures S1 and S2 in Multimedia Appendix 1). If patients had not returned the iMCQ by 8 weeks postdischarge, the questionnaire was conducted over the telephone. If patients had not returned 2 subsequent questionnaires, a research nurse called patients with a reminder. Questionnaires that were not returned or collected otherwise were considered missing.

Statistical Analysis

We calculated the sample size needed for the study based on the expected effect of the intervention on the primary outcome. Previous studies using a comparable eHealth strategy in CABG patients with health care utilization measured with the iMCQ were not available. In a study with abdominal surgery patients, total health care utilization was estimated at a mean of 0.88 (SD 0.15) per patient [17]. In a systematic review by van der Meij et al [10], the effect of an eHealth strategy in surgical patients was not consistent. Therefore, a small to medium effect (d=0.35) was expected from our intervention. Combined with an α of .05 and a power of 0.80, a total sample size of 260 patients was required. We aimed for 280 participants to account for loss to follow-up and nonadherence to the intervention and return of questionnaires (attrition rate 5%, rounded up to a whole number). Demographic data of randomized patients were collected using definitions in line with the Netherlands Heart Registration [18]. Education was grouped into three levels (low, medium, and high) according to the general definition by Statistics Netherlands (see Multimedia Appendix 1 for the full definition).

The main analysis was performed according to the intention-to-treat (ITT) principle. Because patients in the intervention group were not obliged to use the educational videos and VCs might not be possible due to technical errors, per-protocol analysis was also performed, which included only patients who used the intervention strategy as intended (defined as having at least one VC or TC and accessed the educational videos at least once).

Planned subgroup analyses of the primary outcome were performed according to age (<65 years vs ≥65 years), sex, recent myocardial infarction, left ventricular function, diabetes, type of CABG (on-pump vs off-pump), log EuroScore, and highest level of education.

Continuous variables and outcomes are expressed as mean (SD) in cases of a normal distribution and as median (IQR) in cases of a nonnormal distribution. The Kolmogorov-Smirnov test and Q-Q plots were used to test for normality of the data distribution. Categorical data are summarized as absolute and relative frequencies. The updated Dutch Manual for Cost Analysis in Health Care Research was used as the source for cost prices per health care activity if available [19]. Other tariffs were calculated using top-down microcosting as described by Tan and Hakkaart-van Roijen et al [20, 21] (see Multimedia Appendix 1 for details). Each consumed health care activity was multiplied by the cost price and total costs were calculated by summing these multiplications. The HADS and RI-10 questionnaire scores at each interval were compared between study groups. *P*<.05 was considered significant. Primary and secondary outcomes are presented with effect-size estimates and 95% CIs using the Cox proportional hazards model. The proportional hazard assumption was assessed by log (–log) plots. Analyses were performed using SPSS 25 and RStudio.

Results

Study Population

In total, 280 patients were included in the study between February 2020 and December 2021, and subsequently randomized yielding 140 patients in each study group. One patient in the intervention group and two patients in the control group were excluded after randomization because they underwent percutaneous coronary intervention instead of CABG surgery. In the intervention group, three patients were lost to follow-up (1 withdrew consent, 1 had an early readmission due to a complication, and 1 died). In the control group, three patients were lost to follow-up (1 withdrew consent and 2 died). The ITT analysis therefore consisted of 136 and 135 patients in the intervention and control group, respectively. Weblog and planning data revealed that 8 patients did not use the intervention as intended, whereby 128 patients were included in the intervention group in the per-protocol analysis (Figure 1).

Baseline characteristics of patients were similar in the two groups (Table 1), with a median age of 67.9 and 69.6 years for the intervention and control group, respectively. The majority of patients were male in both groups. At the time of surgery, 25% of patients had an urgent indication and the remainder underwent surgery in the elective setting. In the majority of patients, on-pump CABG was performed using 3 distal anastomoses. The left or right internal mammary artery was used in >98% of patients. Duration of admission was also similar in the two groups (Table 1).

Figure 1. Enrollment overview. CABG: coronary artery bypass graft; PCI: percutaneous coronary intervention.

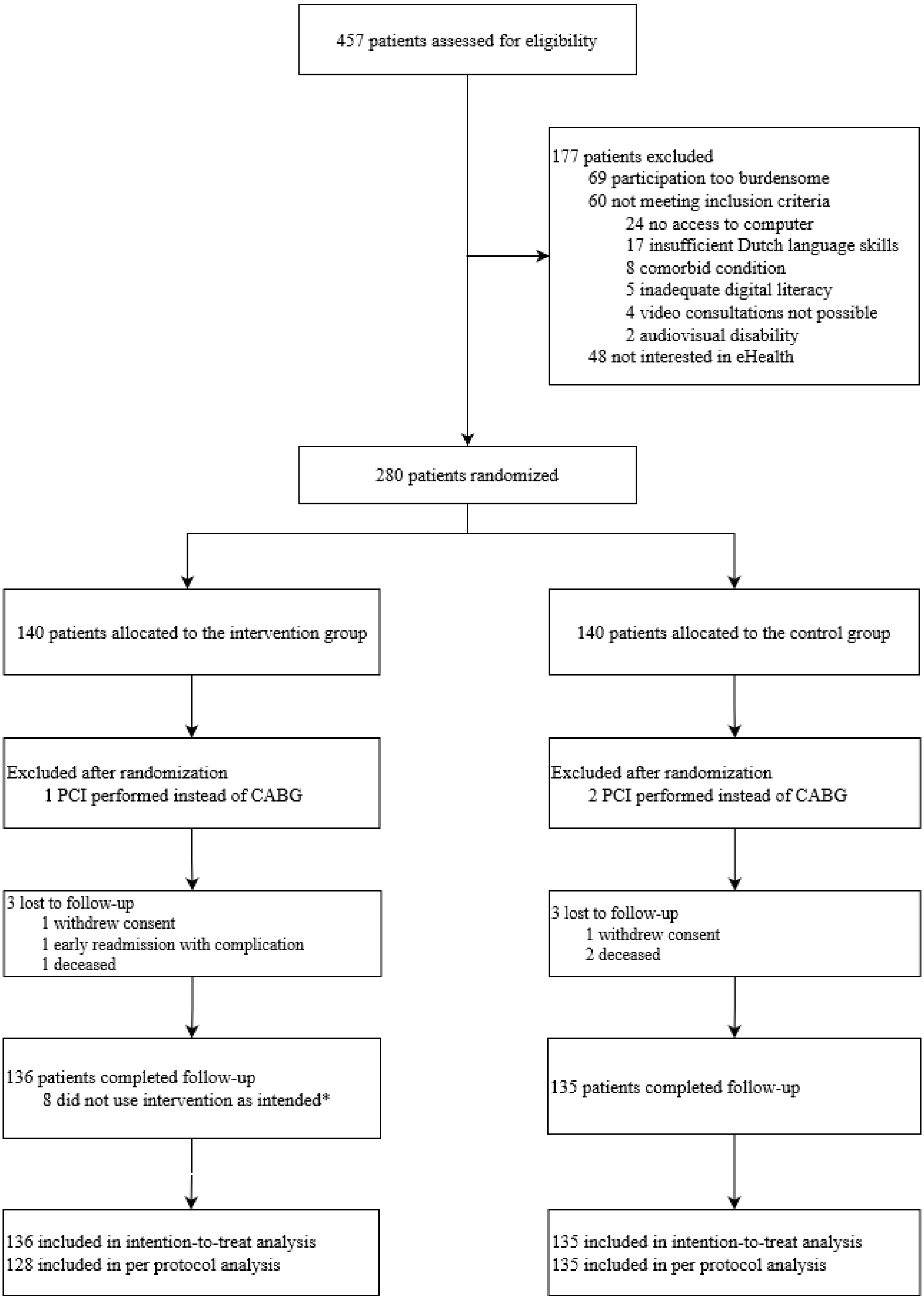


Table 1. Baseline characteristics and procedural data (intention-to-treat analysis).

Characteristics	eHealth group (n=136)	Standard care (n=135)
Age (years), median (IQR)	67.9 (61.5-73.3)	69.6 (65.2-74.1)
Male, n (%)	121 (89.6)	113 (83.1)
BMI, median (IQR)	27.7 (25.1-30.6)	27.2 (25.2-30.3)
Medical history, n (%)		
Diabetes mellitus	45 (33.3)	33 (24.3)
Chronic pulmonary disease	7 (5.2)	15 (11.0)
Atrial fibrillation	9 (6.7)	6 (4.4)
Multivessel disease	117 (86.7)	121 (89.0)
Peripheral vascular disease	17 (12.5)	17 (12.6)
Renal impairment (MDRD[a]<60 mL/min/1.73 m^2)	10 (7.4)	1 (8.1)
Previous stroke	3 (2.2)	5 (3.7)
Recent MI[b] (90 days)	45 (33.3)	46 (33.8)
Previous PCI[c]	36 (26.5)	31 (22.9)
Left ventricular ejection fraction, median (IQR)	55 (50-55)	55 (50-55)
Ejection fraction≤30%, n (%)	0 (0)	3 (2.2)
NYHA[d] class>II, n (%)	3 (2.2)	7 (5.2)
Current health status		
SF-36[e] physical score, median (IQR)	51 (43-56)	48 (40-51)
SF-36 mental score, median (IQR)	58 (55-63)	59 (55-64)
HADS[f], median (IQR)	3 (1-7)	3 (1-6)
Level of education, n (%)		
Low	36 (26.5)	42 (31.1)
Intermediate	53 (39.0)	55 (40.7)
High	47 (34.6)	38 (28.1)
Procedural data		
EuroSCORE log, median (IQR)	2.40 (1.82-4.06)	2.87 (2.01-4.28)
EuroSCORE II, median (IQR)	1.41 (1.05-2.04)	1.32 (0.78-2.43)
Use of ECC[g], n (%)	110 (81.5)	101 (74.8)
ECC duration in users (min), median (IQR)	74 (60-91)	76 (64-91)
Number of distal anastomoses, median (IQR)	3 (2-4)	3 (2-4)
Hospital stay (days), median (IQR)	6 (5-7)	6 (5-7)

[a] MDRD: Modification of Diet in Renal Disease.

[b] MI: myocardial infarction.

[c] PCI: percutaneous coronary intervention.

[d] NYHA: New York Heart Association.

[e] SF-36: Short Item-36.

[f] HADS: Hospital Anxiety and Depression Scale.

[g] ECC: extracorporeal circulation.

Outcomes

At 6 weeks, care was consumed by less patients in the intervention group than in the control group (Table 2). The benefit of the eHealth strategy was most noticeable in patients over 65 years of age, those of male sex, those with recent myocardial infarction, or with a EuroScore>2 (see Figure S3 in Multimedia Appendix 1). Reduction in individual care activities was significantly different between groups for TCs and was borderline significant for general practitioner visits (Table 2).

Costs related to the primary outcome were significantly higher in the standard care group compared with those in the eHealth group (*P*<.001, Table 2), which was attributed to the higher volume of care consumption in the control group (see Table S1 in Multimedia Appendix 1).

A composite of unplanned in-hospital care, a composite of planned and unplanned in-hospital care after discharge, and use of planned and unplanned primary care were all higher in the control group than the intervention group (Table 2). The volume of consumed care was also higher in the control group (Table S1 of Multimedia Appendix 1).

The RI-10 score, indicating patient-reported recovery, was significantly higher in the intervention group in the 3rd and 6th weeks after discharge (Figure 2). Anxiety was not significantly different between study groups (Figure 2). Per-protocol analysis revealed similar findings (see Tables S2 and S3 in Multimedia Appendix 1).

Table 2. Outcomes at 6 weeks.

Outcomes	eHealth group (n=136)	Standard care (n=135)	Hazard ratio (95% CI)	*P* value
Primary outcomes				
Composite outcome[a], n (%)	43 (31.6)	61 (45.2)	0.56 (0.34-0.92)	.02
Cost (Euro[b]), Median (IQR)	0 (0-95)	66 (0-215)	N/A[c]	<.001
Cost (Euro), mean (SD)	183 (515)	285 (777)	N/A	<.001
Secondary outcomes, n (%)				
Composite unplanned in-hospital care	36 (26.5)	53 (39.3)	0.56 (0.33-0.93)	.03
Emergency department visits	14 (10.3)	23 (17.0)	0.56 (0.27-1.14)	.11
Readmissions	7 (5.1)	9 (6.7)	0.76 (0.28-2.10)	.59
Outpatient clinic visits	11 (8.1)	10 (7.4)	1.10 (0.45-2.68)	.83
Telephone consultations	29 (21.3)	47 (34.8)	0.51 (0.29-0.87)	.01
General practitioner visits (unplanned)	28 (20.6)	41 (30.4)	0.59 (0.34-1.04)	.07
Composite of all in-hospital care[d]	69 (50.7)	97 (71.9)	0.40 (0.24-0.67)	<.001
Composite of all primary care[e]	82 (60.3)	101 (74.8)	0.58 (0.36-0.97)	.04

[a] Composite of unplanned health care utilization (ie, emergency department visits, readmissions, outpatient clinic visits, telephone consultations, or general practitioner visits).

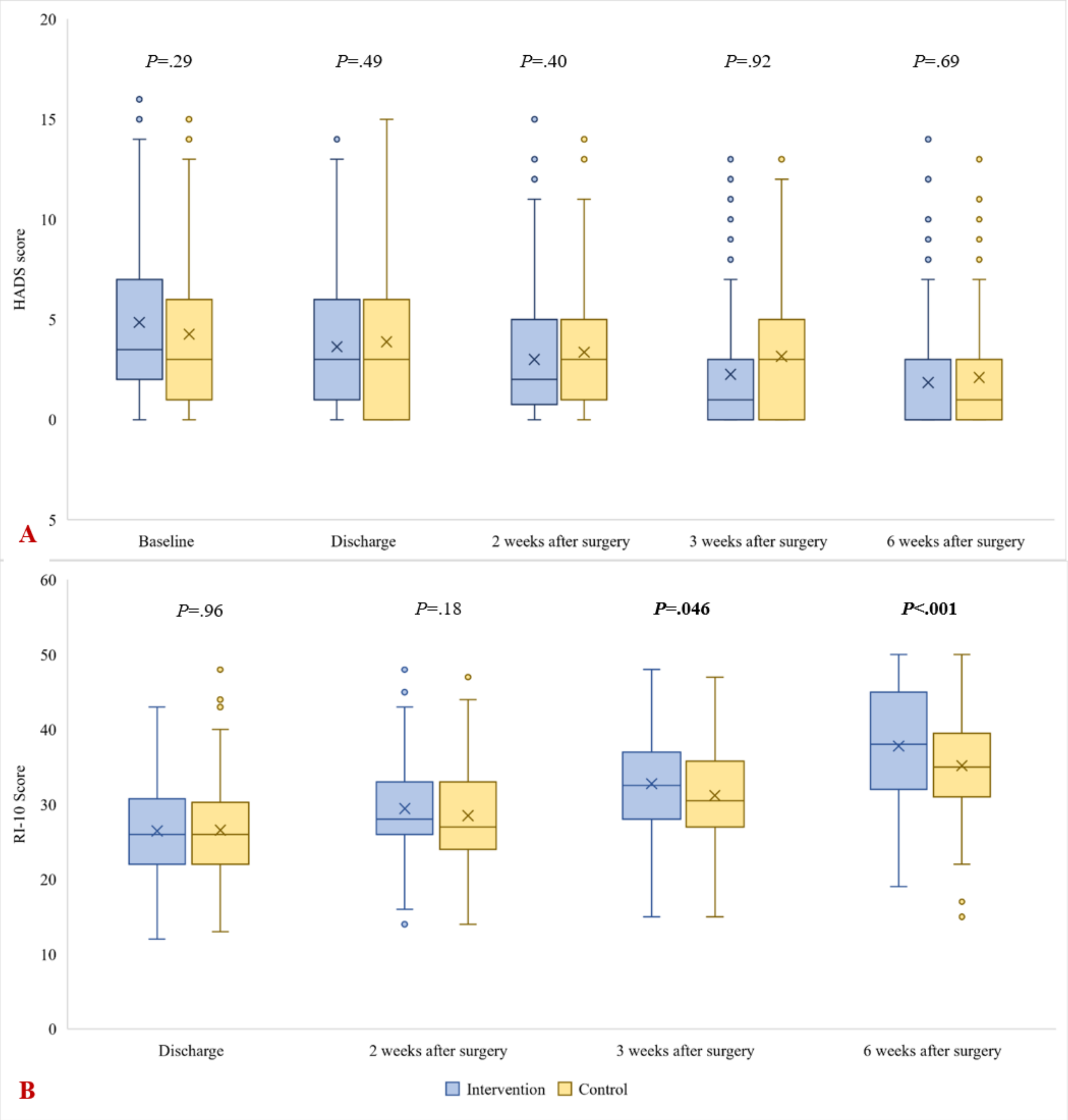
[b] 1 Euro=US $1.13.

[c] N/A: not applicable.

[d] Composite of in-hospital care comprising planned and unplanned emergency department visits, readmissions, outpatient clinic visits, and telephone consultations.

[e] Composite of primary care comprising planned and unplanned visits to the general practitioner, visits to allied health professionals (physical therapists, dieticians, speech therapists, exercise therapy, social workers), and psychologist visits.

Figure 2. Anxiety level measured with the Hospital Anxiety and Depression Scale (HADS) anxiety subscale (A) and progress of recovery measured using the recovery index-10 (RI-10) questionnaire (B).



Process Evaluation

Among patients who were provided access to the educational videos, 95% accessed the videos at least once. A total of 248 of the 272 planned VCs were conducted. Eight patients did not use the intervention as intended and did not receive VCs (n=16 VCs). The other VCs that did not take place were substituted with a TC due to technical errors (n=8). The median duration of VCs was 10 minutes (IQR 8-11) for the first VC and was 8 minutes (IQR 7-9) for the second VC. Patients reported positive attitudes toward the education videos and the VC (Figures S1 and S2 in Multimedia Appendix 1). Most notably, patients reported a fairly positive attitude toward substitution of a face-to-face contact with a VC, but patients also reported that the VC with a physician (not the surgeon) or nurse practitioner should not substitute the consultation with the surgeon.

Discussion

Principal Findings

The principal finding of the IMPROV-ED trial is that an eHealth strategy comprising educational videos and VCs results in a reduction of unplanned care and costs. In addition, the eHealth strategy is associated with faster patient-reported recovery. These findings are of medical and societal importance given the increasing interest in digital health and the need for value-based alongside evidence-based care. Our study is the first to provide robust evidence that an eHealth intervention can aid in reduction of health care utilization and associated costs. This effect appears applicable to both in-hospital care as well as primary care. One of the most pressing concerns from health care insurance companies and decision-makers toward eHealth is the great investment that is required for development of content and the necessary infrastructure and issues that arise after implementation due to lack of reimbursement options [22]. Our findings refute these concerns by showing positive effects on costs. Furthermore, the eHealth strategy did not only contribute to less patients consuming care (Table 2) but also reduced the care consumed per patient (Table S1 of Multimedia Appendix 1), which underlines the high potential of eHealth strategies for this patient population to also positively influence the burden on health care personnel. With an aging population, a vast increase in health care consumption is expected in the near future. Based on the results of our study, an eHealth program is proven to aid in the sustainment of health care systems.

The findings of our study shine new light on previous studies comparing an eHealth program with standard care because it is the first study to use health care utilization as a primary outcome [10]. Very limited studies are available that use health care utilization as an outcome, and those that have considered care consumption as a secondary or tertiary outcome. Previous studies were also not adequately powered to draw reliable conclusions on the use of eHealth in reduction of care consumption and, consequently, these studies reported mixed outcomes. For example, Keeping-Burke et al [23] incorporated health care use as a tertiary outcome in an RCT of patients after CABG surgery using postoperative VCs, and concluded that patients in the telehealth group had fewer physician contacts. Zahlmann et al [24] used telecommunication in the postoperative period after cataract surgery (n=62) and also concluded that care in the intervention group was lower than that in the control group. Conversely, Barnason et al [25,26] conducted two RCTs in 232 and 50 CABG patients, respectively, using a supportive telehealth program and concluded that both groups had similar health care use at 6-month follow-up. Barnason et al [25] and Keeping-Burke et al [23] both reported no differences in emergency department visits and readmissions between study groups. Readmission was also similar in a study by Gandsas et al [27] after laparoscopic gastric bypass using robotic telerounding during admission.

Another major strength of the current eHealth program is that it provides patients of various degrees of socioeconomic status and health literacy with information on the procedure and their medical condition from a reliable source that is endorsed by their surgeon. The Dutch Heart Foundation is a respected organization that is dedicated to providing information on cardiovascular health, advocating patient interests, and conducting research [28]. The educational videos are developed in conjunction with patient representatives and physicians. In the VCs, additional questions are answered and uncertainties are addressed. The impact of educational videos and VCs is presumably in improvement of self-management skills and reduction of fear and anxiety. Recall of information on information provided preoperatively or at discharge is often incomplete, and patients might not know what physical activity is allowed after discharge or who to contact in case of complaints. Patients can turn to the internet for information; however, this information is uncontrolled, sometimes inaccurate, and is not tailored to the care processes of their provider. Because planned care is not initiated until 6 weeks after surgery (and sometimes later in practice), conflicting advice can induce insecurity, which will lead to use of care and will hamper recovery. The results of our study are consistent with this hypothesis. Nevertheless, the anxiety symptoms measured with the HADS questionnaire relate to anxiety in a narrow sense, whereas the anxiety experienced by patients after CABG surgery is likely to be more subtle in nature, which may have contributed to the nonsignificant difference in measured anxiety found in this study.

However, health care utilization is the resultant of a multifactorial behavioral model that attributes a combination of predisposing factors (eg, patient characteristics such as age, sex, sociodemographic parameters, or health literacy and attitude toward health), enabling factors (eg, income, health insurance status, health care organization), and need factors (eg, experience with health care) to health care utilization [29]. The eHealth strategy used in the IMPROV-ED trial has a positive influence on some of these attributes but not all. Interestingly, subgroup analysis showed that the eHealth program had a greater benefit in more vulnerable patients (EuroScore≥2) and revealed a trend toward more benefit in patients with a low level of education. By contrast, a small group of patients who provided informed consent did not use the educational videos or VCs that were part of the eHealth strategy. These patients reported to have received sufficient information from their physician, nurse, or paramedic during admission, or that they found the relevant information online themselves. It might therefore be reasonable to consider adding different modes of digital health delivery to the currently used eHealth strategy (eg, mobile apps, live chat, home monitoring, telerehabilitation) to manage more attributes of health utilization and to offer a more individualized approach tailored to the patients' needs. Combining different modes of digital care might thereby further reduce health care utilization and potentially also improve clinical outcomes [22].

Learning Points and Limitations

Even though the IMPROV-ED study yielded positive results toward the primary outcome (Table 2) and patients were generally positive about the eHealth strategy (Figures S2 and S3 in Multimedia Appendix 1), several learning points and limitations should be taken into account for future eHealth programs.

First, the IMPROVE-ED trial is designed for patients who consume care as a result of insecurity, anxiety, lack of medical knowledge, and/or inadequate discharge counseling. As can be concluded from Figure 1, a relevant number of patients who were invited to participate in the trial did not provide informed consent due to the general burden of having to undergo cardiac surgery (patients used terms such as "stressful," "anxiety," and "insecurity") in conjunction with study obligations. The effect of the eHealth strategy may be underestimated because this group of patients might have been part of the target population in which the eHealth strategy would have incremental value.

Due to ethical constraints (patients did not provide informed consent for participation and thus for data collection), these patients were not further analyzed for the study outcomes.

In this study, standard care was not replaced by digital alternatives, and yet the costs of the intervention group were still lower than those of the control group receiving only standard care. Because VCs were used as an add-on to standard care, there are potentially more opportunities to reduce costs further. The fact that eHealth is being implemented *on top* of current health care services is, in addition to cost concerns, one of the challenges identified by the European Society of Cardiology as hampering the introduction of eHealth into everyday clinical practice [22]. Future endeavors should focus on investigating the potential of substitution of standard physical care with digital alternatives, especially since the patients' attitude was generally positive toward the (hypothetical) substitution of a physical contact with a VC in this study (Figure S1 of Multimedia Appendix 1). Previous studies also stated that it is feasible to obtain the same effective communication and interaction with VCs as with face-to-face care [30].

The majority of patients included in the IMPROV-ED trial were included during the COVID-19 pandemic. The results of the study might therefore be an underrepresentation of care consumption because patients feared transmission in the hospital setting [31]. Nevertheless, the randomized design balances this influence between the study groups.

Conclusion

An eHealth strategy comprising educational videos and VCs can reduce unplanned in-hospital and primary health care utilization and costs, and can aid in faster patient-reported recovery following CABG surgery.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions, tariffs for consumed care, and supplementary data (Tables S1-S3, Figures S1-S3).
[DOCX File , 307 KB - jmir_v24i8e37728_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 116 KB - jmir_v24i8e37728_app2.pdf]

References

1. Weisse AB. Cardiac surgery: a century of progress. Tex Heart Inst J 2011;38(5):486-490 [FREE Full text] [Medline: 22163121]
2. Vonk Noordegraaf A, Anema JR, Louwerse MD, Heymans MW, van Mechelen W, Brölmann HAM, et al. Prediction of time to return to work after gynaecological surgery: a prospective cohort study in the Netherlands. BJOG 2014 Mar;121(4):487-497. [doi: 10.1111/1471-0528.12494] [Medline: 24245993]
3. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
4. Lie I, Bunch EH, Smeby NA, Arnesen H, Hamilton G. Patients' experiences with symptoms and needs in the early rehabilitation phase after coronary artery bypass grafting. Eur J Cardiovasc Nurs 2012 Mar;11(1):14-24. [doi: 10.1016/j.ejcnurse.2010.09.004] [Medline: 21030311]
5. Fox JP, Suter LG, Wang K, Wang Y, Krumholz HM, Ross JS. Hospital-based, acute care use among patients within 30 days of discharge after coronary artery bypass surgery. Ann Thorac Surg 2013 Jul;96(1):96-104 [FREE Full text] [doi: 10.1016/j.athoracsur.2013.03.091] [Medline: 23702228]
6. Shawon MSR, Odutola M, Falster MO, Jorm LR. Patient and hospital factors associated with 30-day readmissions after coronary artery bypass graft (CABG) surgery: a systematic review and meta-analysis. J Cardiothorac Surg 2021 Jun 10;16(1):172 [FREE Full text] [doi: 10.1186/s13019-021-01556-1] [Medline: 34112216]
7. Report to the Congress: promoting greater efficiency in Medicare. Medicare Payment Advisory Commission. 2007. URL: https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/Jun07_ExecSummary.pdf [accessed 2021-10-03]
8. Saab S, Noureddine S, Dumit NY. Readmission rates and emergency department visits after coronary artery bypass graft surgery and related factors. J Med Liban 2013;61(3):155-160. [doi: 10.12816/0001444] [Medline: 24422366]
9. Fifty-eighth World Health Assembly, Geneva, 16-25 May 2005: resolutions and decisions: annex. World Health Organization. 2005. URL: https://apps.who.int/iris/handle/10665/20398 [accessed 2020-02-01]
10. van der Meij E, Anema JR, Otten RHJ, Huirne JAF, Schaafsma FG. The effect of perioperative e-health interventions on the postoperative course: a systematic review of randomised and non-randomised controlled trials. PLoS One 2016;11(7):e0158612 [FREE Full text] [doi: 10.1371/journal.pone.0158612] [Medline: 27383239]

11. van Steenbergen GJ, van Veghel D, Ter Woorst J, van Lieshout D, Dekker L. IMPROV-ED trial: eHealth programme for faster recovery and reduced healthcare utilisation after CABG. Neth Heart J 2021 Feb;29(2):80-87 [FREE Full text] [doi: 10.1007/s12471-020-01508-9] [Medline: 33141398]
12. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
13. Meppelink C. Designing digital health information in a health literacy context. PhD Thesis. University of Amsterdam. 2016. URL: https://dare.uva.nl/search?identifier=220fb9ef-1a51-4325-9019-61cfdaf8633f [accessed 2022-07-21]
14. Manual iMTA Medical Cost Questionnaire (iMCQ). iMTA Productivity Health Research Group. URL: https://www.imta.nl/questionnaires/imcq/ [accessed 2022-07-21]
15. Kluivers KB, Hendriks JCM, Mol BWJ, Bongers MY, Vierhout ME, Brölmann HAM, et al. Clinimetric properties of 3 instruments measuring postoperative recovery in a gynecologic surgical population. Surgery 2008 Jul;144(1):12-21. [doi: 10.1016/j.surg.2008.03.027] [Medline: 18571580]
16. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychol Med 1997 Mar;27(2):363-370. [doi: 10.1017/s0033291796004382] [Medline: 9089829]
17. Pattamatta M, Smeets BJJ, Evers SMAA, Peters EG, Luyer MDP, Hiligsmann M. Quality of life and costs of patients prior to colorectal surgery. Expert Rev Pharmacoecon Outcomes Res 2020 Apr;20(2):193-198. [doi: 10.1080/14737167.2019.1628641] [Medline: 31190575]
18. De Nederlandse Hart Registratie (Netherlands Heart Registration). 2020. URL: https://nederlandsehartregistratie.nl/handboeken/ [accessed 2020-02-01]
19. Kanters TA, Bouwmans CAM, van der Linden N, Tan SS, Hakkaart-van Roijen L. Update of the Dutch manual for costing studies in health care. PLoS One 2017;12(11):e0187477 [FREE Full text] [doi: 10.1371/journal.pone.0187477] [Medline: 29121647]
20. Tan SS. Microcosting in economic evaluations: issues of accuracy, feasibility, consistency and generalisability. Erasmus University. 2009. URL: http://hdl.handle.net/1765/17354 [accessed 2020-02-01]
21. Hakkaart-van Roijen L, Van der Linden N, Bouwmans C, Kanters T, Tan SS. Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg In opdracht van Zorginstituut Nederland Geactualiseerde versie. Kostenhandleiding 2015.
22. Frederix I, Caiani EG, Dendale P, Anker S, Bax J, Böhm A, et al. ESC e-Cardiology Working Group Position Paper: overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiol 2019 Jul;26(11):1166-1177. [doi: 10.1177/2047487319832394] [Medline: 30917695]
23. Keeping-Burke L, Purden M, Frasure-Smith N, Cossette S, McCarthy F, Amsel R. Bridging the transition from hospital to home: effects of the VITAL telehealth program on recovery for CABG surgery patients and their caregivers. Res Nurs Health 2013 Dec;36(6):540-553. [doi: 10.1002/nur.21571] [Medline: 24242195]
24. Zahlmann G, Mertz M, Fabian E, Holle R, Kaatz H, Neubauer L, et al. Perioperative cataract OP management by means of teleconsultation. Graefes Arch Clin Exp Ophthalmol 2002 Jan;240(1):17-20. [doi: 10.1007/s00417-001-0396-0] [Medline: 11954775]
25. 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]
26. Barnason S, Zimmerman L, Schulz P, Tu C. Influence of an early recovery telehealth intervention on physical activity and functioning after coronary artery bypass surgery among older adults with high disease burden. Heart Lung 2009;38(6):459-468 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.01.010] [Medline: 19944870]
27. Gandsas A, Parekh M, Bleech MM, Tong DA. Robotic telepresence: profit analysis in reducing length of stay after laparoscopic gastric bypass. J Am Coll Surg 2007 Jul;205(1):72-77. [doi: 10.1016/j.jamcollsurg.2007.01.070] [Medline: 17617335]
28. Nederlandse Hartstichting (Dutch Heart Foundation). URL: https://www.hartstichting.nl/ [accessed 2020-02-06]
29. Babitsch B, Gohl D, von Lengerke T. Re-revisiting Andersen's Behavioral Model of Health Services Use: a systematic review of studies from 1998-2011. Psychosoc Med 2012;9:Doc11. [doi: 10.3205/psm000089] [Medline: 23133505]
30. Jiménez-Rodríguez D, Ruiz-Salvador D, Rodríguez Salvador MDM, Pérez-Heredia M, Muñoz Ronda FJ, Arrogante O. Consensus on criteria for good practices in video consultation: a Delphi study. Int J Environ Res Public Health 2020 Jul 27;17(15):5396 [FREE Full text] [doi: 10.3390/ijerph17155396] [Medline: 32727042]
31. Derks L, Sturkenboom HN, Zaal M, Houterman S, Woudstra P, Tio RA, et al. Association between public media and trends in new acute coronary syndrome presentations during the first COVID-19 wave in the Netherlands. Neth Heart J 2021 Nov;29(11):577-583 [FREE Full text] [doi: 10.1007/s12471-021-01603-5] [Medline: 34327671]

Abbreviations

CABG: coronary artery bypass graft

CONSORT: Consolidated Standards of Reporting Trials
HADS: Hospital Anxiety and Depression Scale
iMCQ: Institute for Medical Technology Assessment Medical Consumption Questionnaire
ITT: intention to treat
RCT: randomized controlled trial
RI-10: Recovery Index-10
TC: telephone consultation
VC: video consultation

Edited by G Eysenbach; submitted 04.03.22; peer-reviewed by E van der Velde, K Matthias, M Nomali; comments to author 06.05.22; revised version received 04.07.22; accepted 18.07.22; published 26.08.22.

Please cite as:
van Steenbergen G, van Veghel D, van Lieshout D, Sperwer M, ter Woorst J, Dekker L
Effects of Video-Based Patient Education and Consultation on Unplanned Health Care Utilization and Early Recovery After Coronary Artery Bypass Surgery (IMPROV-ED): Randomized Controlled Trial
J Med Internet Res 2022;24(8):e37728
URL: https://www.jmir.org/2022/8/e37728
doi:10.2196/37728
PMID:36018625

Original Paper

Health Information Seeking From an Intelligent Web-Based Symptom Checker: Cross-sectional Questionnaire Study

Kimberly Arellano Carmona[1], MPH; Deepti Chittamuru[1], PhD; Richard L Kravitz[2], MD, MSPH; Steven Ramondt[3,4], PhD; A Susana Ramírez[1], MPH, PhD

[1]School of Social Sciences, Humanities and Arts, University of California, Merced, CA, United States

[2]University of California, Davis, Sacramento, CA, United States

[3]Department of Donor Medicine Research, Sanquin Research, Amsterdam, Netherlands

[4]Department of Communication Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Corresponding Author:
A Susana Ramírez, MPH, PhD
School of Social Sciences, Humanities and Arts
University of California
5200 Lake Rd
Merced, CA, 95343
United States
Phone: 1 209 228 4400
Email: sramirez37@ucmerced.edu

Abstract

Background: The ever-growing amount of health information available on the web is increasing the demand for tools providing personalized and actionable health information. Such tools include symptom checkers that provide users with a potential diagnosis after responding to a set of probes about their symptoms. Although the potential for their utility is great, little is known about such tools' actual use and effects.

Objective: We aimed to understand who uses a web-based artificial intelligence–powered symptom checker and its purposes, how they evaluate the experience of the web-based interview and quality of the information, what they intend to do with the recommendation, and predictors of future use.

Methods: Cross-sectional survey of web-based health information seekers following the completion of a symptom checker visit (N=2437). Measures of comprehensibility, confidence, usefulness, health-related anxiety, empowerment, and intention to use in the future were assessed. ANOVAs and the Wilcoxon rank sum test examined mean outcome differences in racial, ethnic, and sex groups. The relationship between perceptions of the symptom checker and intention to follow recommended actions was assessed using multilevel logistic regression.

Results: Buoy users were well-educated (1384/1704, 81.22% college or higher), primarily White (1227/1693, 72.47%), and female (2069/2437, 84.89%). Most had insurance (1449/1630, 88.89%), a regular health care provider (1307/1709, 76.48%), and reported good health (1000/1703, 58.72%). Three types of symptoms—pain (855/2437, 35.08%), gynecological issues (293/2437, 12.02%), and masses or lumps (204/2437, 8.37%)—accounted for almost half (1352/2437, 55.48%) of site visits. Buoy's top three primary recommendations split across less-serious triage categories: primary care physician in 2 weeks (754/2141, 35.22%), self-treatment (452/2141, 21.11%), and primary care in 1 to 2 days (373/2141, 17.42%). Common diagnoses were musculoskeletal (303/2437, 12.43%), gynecological (304/2437, 12.47%) and skin conditions (297/2437, 12.19%), and infectious diseases (300/2437, 12.31%). Users generally reported high confidence in Buoy, found it useful and easy to understand, and said that Buoy made them feel less anxious and more empowered to seek medical help. Users for whom Buoy recommended "Waiting/Watching" or "Self-Treatment" had strongest intentions to comply, whereas those advised to seek primary care had weaker intentions. Compared with White users, Latino and Black users had significantly more confidence in Buoy ($P<.05$), and the former also found it significantly more useful ($P<.05$). Latino (odds ratio 1.96, 95% CI 1.22-3.25) and Black (odds ratio 2.37, 95% CI 1.57-3.66) users also had stronger intentions to discuss recommendations with a provider than White users.

Conclusions: Results demonstrate the potential utility of a web-based health information tool to empower people to seek care and reduce health-related anxiety. However, despite encouraging results suggesting the tool may fulfill unmet health information needs among women and Black and Latino adults, analyses of the user base illustrate persistent second-level digital divide effects.

(J Med Internet Res 2022;24(8):e36322) doi:10.2196/36322

KEYWORDS

health information seeking; health information; information seeking; information seeker; information behavior; artificial intelligence; medical information system; digital divide; information inequality; digital epidemiology; symptom checker; digital health; eHealth; online health information; user demographic; health information resource; health information tool; digital health assistant

Introduction

Background

The ever-growing amount of health information available on the web is increasing the demand for tools that provide personalized and actionable health information. In addition, patients avidly seek information to inform their own health care decisions, either directly or by verifying information discussed during professional consultations. The broad scope of web-based health information includes generic information obtained through web-based searches and decision aids and tools that deliver personalized advice based on information specific to users. Such tools include symptom checkers that provide users with a potential diagnosis after responding to a set of probes about their symptoms.

Web-based symptom checkers are becoming increasingly popular, and the emergence of the COVID-19 pandemic has increased interest in these tools [1]. However, only a few studies have examined how and why they are used [2-4]. The limited research on symptom checkers has found generally positive effects of their use; technologically sophisticated web-based triage systems may help reduce unnecessary visits to emergency rooms and overuse of antibiotics [4], make health care accessible in low-resource settings [5], and increase patient engagement [6]. However, although the potential for their utility is great, more research is needed on the actual use and effects of such tools.

Some studies have raised concerns about the potential of web-based health information systems to spread disinformation and inaccurate diagnostic information [2,7,8]. For example, a study evaluating the diagnostic and triage accuracy of 23 web-based symptom checkers found that physicians performed better than the symptom checker algorithms [4]. However, physicians made incorrect diagnoses in 15% of the cases. Although research suggests that symptom checkers may be less effective than physicians in terms of diagnostic accuracy, it might be more critical that symptom checkers provide recommended actions (eg, whether symptoms warrant a trip to the hospital). Therefore, it is important to understand the impact of symptom checkers on how patients seek care and respond to health care advice.

A significant potential contribution of web-based symptom checkers as triage systems may be to reduce the negative effects of the current overwhelming health information environment, such as the health information overload experienced by web-based health information seekers and their struggle to discern reliable information from misinformation. A web-based medical information system that addresses the abovementioned problems can help people better understand the potential causes of the symptoms they are experiencing, empower them to seek the right kind of help, and potentially reduce anxiety caused by the symptoms they are experiencing.

Users must be able to trust and follow their recommendations for web-based symptom checkers to make meaningful contributions. If web-based symptom checkers are not trusted, they are less likely to be adopted by users, thereby limiting their potential [9]. Moreover, users may be unclear about the technology behind web-based symptom checkers. Research suggests that web-based symptom checkers' artificial intelligence (AI) systems are neither transparent nor comprehensible to users, which may undermine trust in such tools [10]. Nevertheless, despite hesitancy and concerns regarding the accuracy, AI-powered symptom checkers have been perceived as useful for diagnosis by users [11].

A large body of research on information seeking grounded in the uses and gratification frameworks [12] has examined how people use different media to fulfill or gratify various needs. Research in this tradition has characterized health information–seeking behaviors by sources (ie, web-based vs offline seeking [13,14]) or objective (ie, seeking for themselves vs others [15-17]). Multiple studies have confirmed that active information seekers from nonclinical sources, including the internet, are more likely to be White, female, and have relatively high levels of education and income [18-22]. Racial differences in health information–seeking, as well as confidence in information and trust in various sources, have been well documented. There may be different levels of trust and use of sources by racial groups, which can lead to disparities if inaccurate sources are used [23].

Research based on self-reported media use has established that deliberate information seeking from media, including the internet, has been associated with better health outcomes [24], increased engagement in prevention behaviors [25], and more positive patient-clinician interactions [26-28] and has also assisted individuals in coping with uncertainty [20]. Web-based health information–seeking before presenting to an emergency physician also has the potential to improve patient-provider interaction without negatively affecting adherence to treatment [29].

Despite this extensive body of research on information seeking and the importance of the internet and other "new" media as sources of medical information, the quality of the evidence for the effects has been limited. Most previous studies examining information seeking from nonclinical sources, including nearly all internet-based health information–seeking studies, are limited by their reliance on self-reports of individuals' information-seeking behaviors and behavioral or psychosocial outcomes. Furthermore, most studies rely on generalized, non–time-bound health information–seeking behaviors (ie, "Have you ever looked for information about [a topic] from [a source]”), or ask about information seeking within a specific timeframe, but do not examine the content of the information retrieved or the recommendation provided. Thus, the next frontier in this line of research is one that links objective measures of information seeking—both sources and content—with clinical and psychosocial outcomes to understand how people use the information they seek and find from nontraditional sources.

Objectives

This study aimed to address the methodological limitations of prior information-seeking research and examine who seeks information from an intelligent web-based symptom checker and for what purpose, how users experience the tool, what they intend to do with the information, and predictors of intentions to follow the tool recommendations. The following research questions (RQs) guided this study:

1. RQ1: Who uses a web-based symptom checker?
2. RQ2: What drives users to use a web-based symptom checker?
3. RQ3: What were the web-based symptom checker’s recommendations?
4. RQ4: How do users perceive the web-based symptom checker?
5. RQ5: What is the relationship between perceptions of a web-based symptom checker and intention to follow recommended actions?

Methods

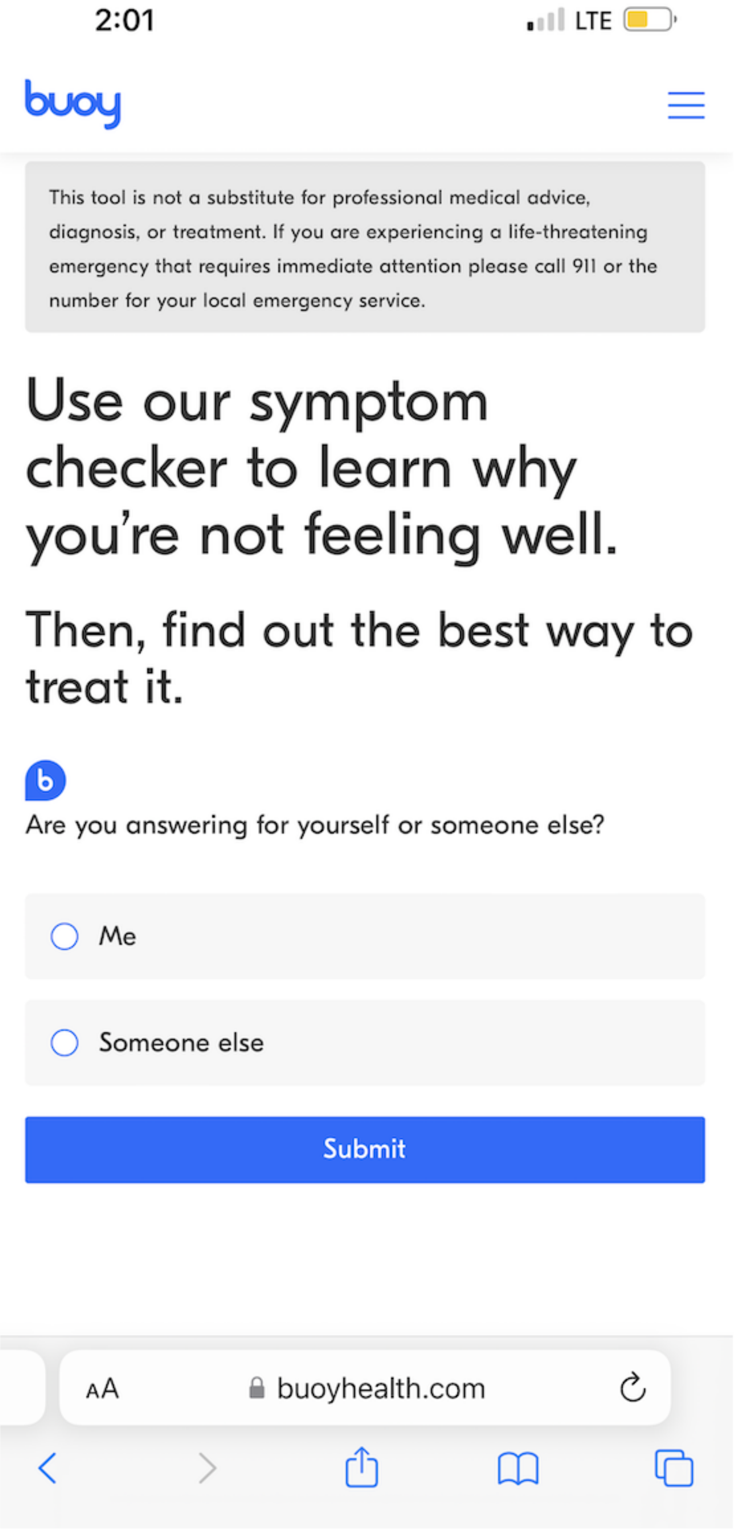
We conducted a cross-sectional survey of web-based health information seekers immediately following the completion of a visit to a web-based intelligent symptom checker, Buoy Health (Buoy Health, Inc [30]; N=2437).

Buoy Health: an AI-Powered Web-Based Symptom Checker

This cross-sectional study used data from patient encounters using Buoy Health, an AI-powered web-based symptom checker, between January 14, 2019, and February 28, 2019. Founded in 2014 by a team of physicians and researchers, the tool is based on conversational medical interviewing, mirroring a conversation with a provider. At the time of writing, Buoy’s symptom checker remains accessible for free on the web or through an app to any internet-connected person. The AI-powered tool uses a progressive series of health questions communicated via a chatbot to assess user symptoms (Figure 1). Buoy’s triage or diagnostic system by design offers health information customized for the user.

Buoy’s proprietary algorithm sources data from >18,000 clinical research studies [31]. Users receive 3 possible diagnoses and recommendations for appropriate levels of care (Figure 2). According to Buoy, the tool’s diagnostic accuracy is 90% [32]. Thus, tools such as Buoy—and other intelligent symptom checkers—have the potential to cut through the clutter of too much and contradictory information to provide personalized, science-based recommendations. A study examining how patients’ use of Buoy affected their plans for seeking care found that Buoy decreased uncertainty among users [33]. Buoy also lowered the level of urgency in patients associated with their condition. This study suggests tools such as Buoy are associated with users’ intended behavior when seeking care based on triage questions. Accordingly, our study adds to the growing literature that seeks to understand how patients use tools such as Buoy together with their providers to manage their health.

Figure 1. Screenshot of the patient-facing, artificial intelligence–assisted Buoy Health symptom checker.



This tool is not a substitute for professional medical advice, diagnosis, or treatment. If you are experiencing a life-threatening emergency that requires immediate attention please call 911 or the number for your local emergency service.

Use our symptom checker to learn why you’re not feeling well.

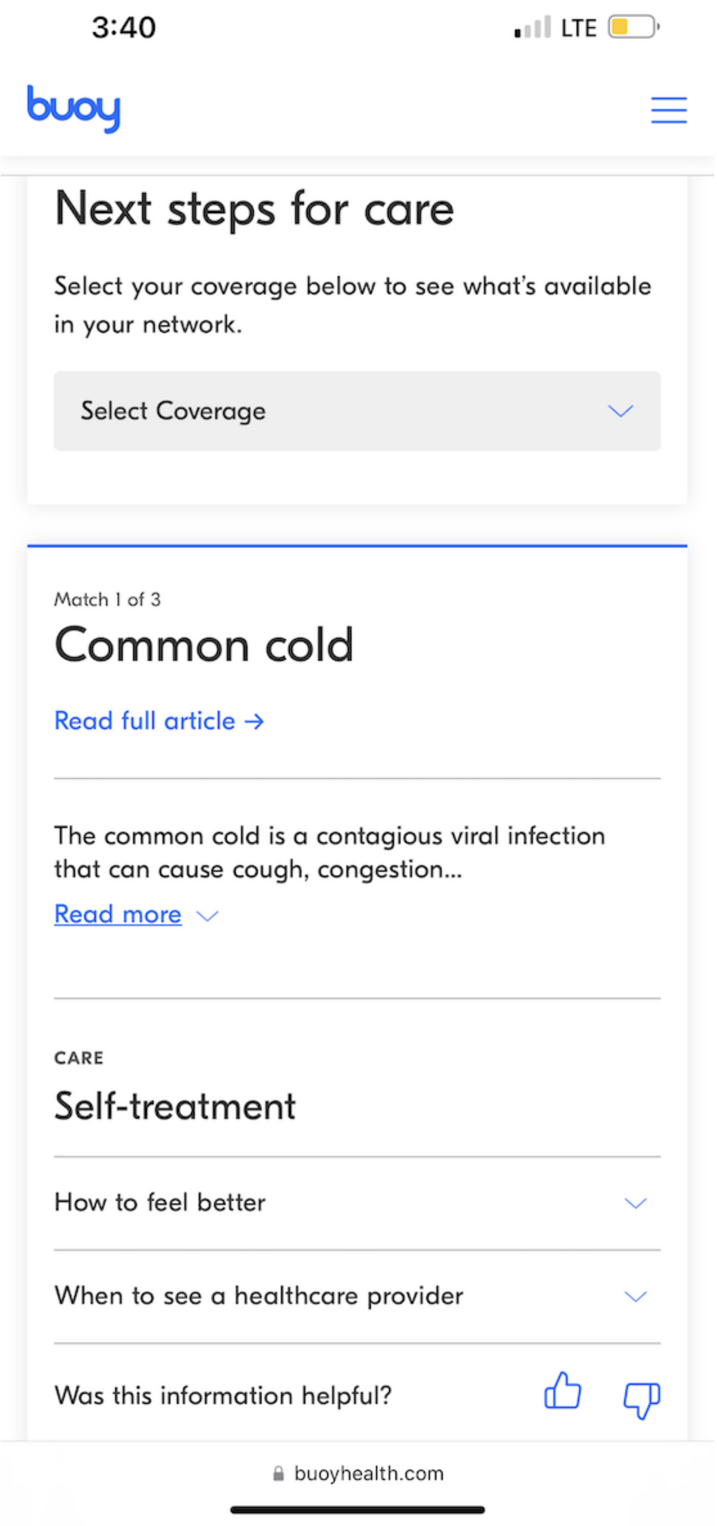
Then, find out the best way to treat it.

Are you answering for yourself or someone else?

○ Me

○ Someone else

Figure 2. Screenshot of a Buoy Health symptom checker recommendation.



3:40 LTE

buoy

Next steps for care

Select your coverage below to see what's available in your network.

Select Coverage

Match 1 of 3

Common cold

Read full article →

The common cold is a contagious viral infection that can cause cough, congestion...

Read more

CARE

Self-treatment

How to feel better

When to see a healthcare provider

Was this information helpful?

buoyhealth.com

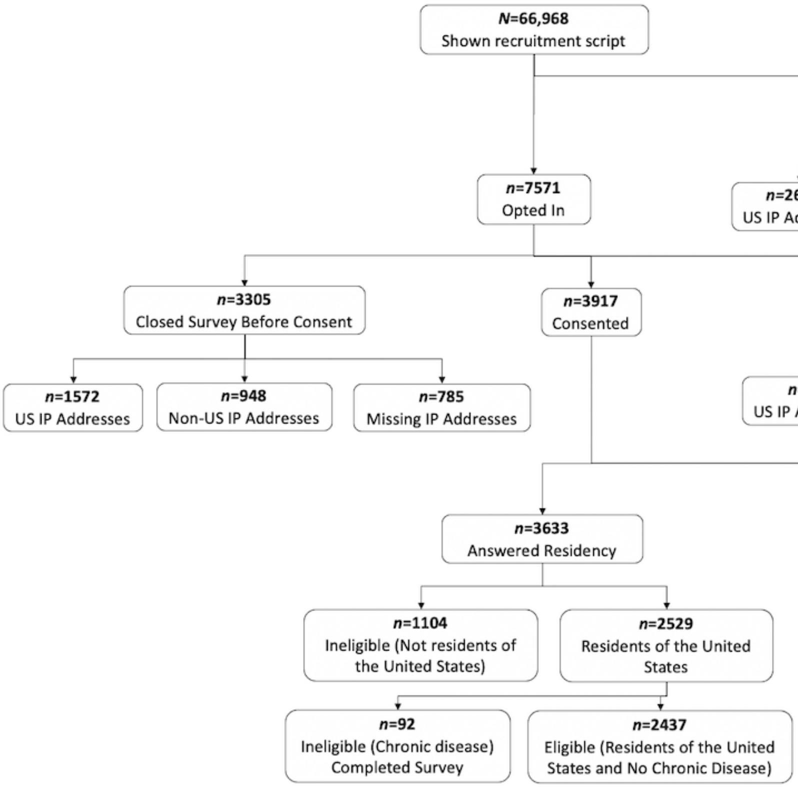
Sampling and Procedure

A recruitment script was shown to Buoy users, assumed to be web-based health information seekers, who met the inclusion criteria via a pop-up window immediately following their Buoy session. Inclusion criteria included completion of the Buoy interview to the recommendation stage in ≤10 minutes, being aged ≥18 years, and residency in the United States (although not necessarily physically in the United States at the time of seeking).

In addition, potential participants must have completed the Buoy interview for themselves; that is, they were seeking information about their symptoms. As 95% of Buoy users complete the diagnostic interview within 10 minutes, users who took longer were not representative of the typical user and thus were not invited to participate in our study to avoid other unanticipated ways in which they might differ from the typical user. Using similar logic, we excluded people who had a pre-existing serious or chronic condition [34] as they may not be representative of the typical Buoy user either. It is expected that their health information–seeking habits and use patterns of Buoy would be different from all other Buoy users. Finally, for ethical reasons, we excluded users who Buoy advised to seek immediate medical care from eligibility, including immediate medical care via 911 or in the emergency department. Figure 3 shows attrition at each stage.

Participants received a US $5 electronic gift card in appreciation of their time following completion of the survey, which had a mean time to completion of 8.61 (SD 6.78) minutes. The gift cards were delivered to an email address that was also used for follow-up. Participants were informed that they would receive another incentive (US $10) following the completion of a second survey. A follow-up assessment was administered 2 weeks after the initial survey to those who chose to provide an email address; however, this study reports only the baseline data.

Figure 3. Flow diagram showing attrition of participants.



Constructs and Measures

Overview

This study followed the tradition of uses and gratifications research [12]. We sought to understand who uses Buoy, perceptions of the user experience, and what they intend to do with the information they obtain. The survey was guided by the integrative model of behavior change [35]. The key constructs and measures are described in the following sections, and the complete survey instrument is available upon request from the corresponding author.

Reasons for Using Buoy

Patients could select ≥1 of 5 reasons for using Buoy; the list was based on the review of internet use for appraisal of symptoms of physical illness by Mueller et al [3]. Options included not being confident that the health care provider provided the correct diagnosis, symptoms not serious, sensitive or embarrassing symptoms, new symptoms, and persistent symptoms. An open-ended response was also provided, and the results were interpreted by 2 coders to map to original close-ended or new codes (access, anxiety, curiosity, and triage).

Trust in Health Information Sources

Trust in a variety of health information sources was assessed using a single Likert item, with responses ranging from “Not at all” (1) to “A lot” (4), adapted from the National Cancer Institute’s Health Information National Trends Survey [36]. The question stem was “In general, how much would you trust information about health or medical topics from each of the following?” The list of sources was randomized: physician/family or friends/newspapers or magazines/radio/internet news/television/government health agencies/social media (such as Facebook or Twitter)/Google/blogs/Buoy.

Prior research has demonstrated that the usability of the health information website affects trust in and credibility of the health information found on the site [37]. Thus, we assessed confidence, comprehensibility, perceived utility, and the emotional effects of using Buoy.

Confidence in Smart Symptom Checkers

A total of 2 items with 5-point response values from “Not at all confident” to “Very Confident” were adapted from Sivakumar et al [38] and combined as a scale where higher values represented greater confidence.

Comprehensibility of Smart Symptom Checkers

We assessed the extent to which the language on the website was easy to understand and the website was understandable and easy to read using 3 response values of 5 points (strongly disagree to strongly agree). Items were based on a scale by van Weert et al [39], with higher values representing greater comprehensibility of Buoy.

Perceived Utility of Smart Symptom Checkers

A total of 3 items with 5-point response values (strongly disagree to strongly agree) and combined as a scale by Davis [40] were used to assess the extent to which the website made the diagnosis of symptoms quicker and easier and the website’s overall usefulness.

Emotional Effects of Using Smart Symptom Checkers

The emotional effect was informed by White and Horvitz [41] and was measured using a scale of 2 items of 5 points (strongly disagree to strongly agree). The items assessed anxiousness about a perceived medical condition and the extent to which the website encouraged help seeking. Higher values represented more positive emotional effects of using the website.

Recommended Action

At the conclusion of the Buoy diagnostic interview, users were given at least one and up to 3 of 8 possible recommendations for the next steps (triage level) that correspond to their cluster of symptoms and potential diagnosis: (1) wait and watch, (2) self-treatment, (3) phone call or in-person visit in the next 3 days, (4) primary care physician in 2 weeks, (5) primary care physician in 1 to 2 days, (6) in-person visit that day or as soon as possible, (7) hospital emergency room, and (8) emergency medical service. Buoy users who received the 2 most urgent recommended actions were not included in our study for ethical reasons. A complete description of the recommendations is available in Multimedia Appendix 1. Buoy provided the research team with the actual recommendations shown to all eligible users. In addition, we asked participants to indicate which of the 6 possible recommendations they had received from Buoy. We compared participants’ self-reports with the Buoy-reported recommendations as a manipulation check. The comparison matched survey responses with at least one of the Buoy recommendations. Most self-reported recommendations matched at least one recommendation, as reported by Buoy (1595/2141, 74.49%).

Intention to Follow Recommended Action

The reasoned action approach informed this measure [35]. Intention was measured for all 6 included recommended actions and intentions to discuss Buoy’s recommendation with a physician or other health professional. The response values ranged from 1 (strongly disagree) to 5 (strongly agree). Examples of the statement are as follows, “I intend to [follow Buoy’s recommendation]” and “I intend to discuss the information I got from Buoy with my doctor or other health care professional.” Higher values on the item represented a stronger intention to follow Buoy’s recommendations or discuss the same with the physician. The recommended action was assessed as a binary variable. Users who scored 4 or 5 on intention (high) were classified as having medical intention, whereas those who scored 1, 2, or 3 were classified as having no medical intention. Intention to follow “Wait and Watch” and “Self-Treatment” were combined on a “No Medical Intention” scale. The intentions “Phone Call or In-Person Visit in the next 3 Days,” “Primary Care Doctor in 1-2 days,” and “In-Person Visit Today or ASAP” were combined in a “Medical Intention” scale.

Coding of Symptoms and Diagnoses

Users’ self-reported symptoms resulting from the Buoy interview were coded into 13 categories using the Centers for Disease Control and Prevention National Ambulatory Medical Care Survey (NAMCS) coding protocol [42]. During the interview, the users were prompted to enter up to 5 presenting symptoms. We report only on the first as that was the primary issue driving the use of the web-based symptom checker. Using an iterative coding process, we generated a set of unique symptoms (N=2040) and unique diagnoses (N=938) from all Buoy data sets.

A total of 2 coders independently coded the first symptom. Coder 1 was part of the codebook development process. Coder 2 was introduced into the study once the codebook was finalized. Disagreements were resolved by discussion. The second author resolved disagreements when consensus was not possible. Cohen κ was run to determine interrater reliability between the 2 coders’ assignment of NAMCS codes for the 2040 unique symptoms; there was substantial agreement between the 2 coders (κ=0.73; [43]). We further categorized whether the first symptoms were serious and likely to require medical attention based on Shapiro et al [44] (chest pain that is heart related, bleeding, loss of consciousness, shortness of breath, and weight loss).

Users are provided with up to 3 possible diagnoses or display names at the completion of their interview, ranked and weighted according to Buoy’s proprietary algorithm, along with recommendations for subsequent actions. Diagnoses were coded into 25 categories comprising major systems, disorders, and conditions, in line with the NAMCS. We report the first diagnosis display name as the algorithm had the highest confidence in it. In addition, the first diagnosis display name had no missing data.

Analytic Approach

For this descriptive analysis, we computed frequencies and percentages to summarize participant characteristics and experiences, overall and by sex and ethnicity where relevant, and to assess intentions to comply with Buoy recommendations. ANOVAs with Bonferroni correction examined the mean outcome differences between racial or ethnic groups on user experience and recommendations. Nonparametric tests in the form of the Wilcoxon rank sum test were performed to test the mean differences between sexes in user experience and recommendations. The relationship between perceptions of the symptom checker and intention to follow recommended actions was assessed using logistic regression. Logistic regression models examined the factors affecting confidence in recommendations and intention to follow these. Analyses were conducted using R (version 4.0.3; R Foundation for Statistical Computing).

Ethics Approval

The University of California, Merced Institutional Review Board approved this study (approval number: UCM2018-124).

Results

Users of a Web-Based Symptom Checker

Consistent with prior studies on health information seekers, Buoy users were well-educated (1384/1704, 81.22% some college or more), mostly White (1227/1693, 72.47%), and female (2069/2437, 84.89%). The mean age of the users was 39.4 (SD 14.7) years. Users were similar to other users of web-based symptom tools, and a prior study of web-based symptom checkers found that users were predominantly female with a mean age of 40 years [33]. Findings from other studies further indicate an age, sex, and socioeconomic divide among adults' web-based health information–seeking behaviors [45]. The sampled users were also relatively privileged in terms of health care access; most had insurance (1449/1630, 89%) and a regular health care provider (1307/1709, 76%). They were generally in good health; 59% (1000/1703) reported their health as good, very good, or excellent. Table 1 shows the additional demographic details.

Table 1. Sample characteristics and comparison with all users of an intelligent web-based symptom checker.

Characteristics	Analytic sample[a] (N=2437)	Eligible opt-outs (N=27,816)
Age (years)		
Values, mean (SD)	39.35 (14.43)	36.92 (14.13)
Values, range	18-87	18-89
Ethnicity (N=1693), n (%)		
White	1227 (72.47)	—[b]
Black or African American	189 (11.16)	—
Latino or Hispanic	139 (8.21)	—
Asian or Pacific Islander	86 (5.08)	—
Other	52 (3.07)	—
Education (N=1704), highest level completed, n (%)		
High school or less	320 (18.78)	—
Some college	689 (40.43)	—
College	695 (40.79)	—
Household income (US $; N=1654), n (%)		
<20,000	304 (18.38)	
20,000-34,999	226 (13.66)	—
35,000-49,999	232 (14.03)	—
50,000-74,999	316 (19.11)	—
75,000-99,999	237 (14.33)	—
≥100,000	339 (20.50)	—
General health status (self-reported; N=1703), n (%)		
Excellent	63 (3.70)	—
Very good	288 (16.91)	—
Good	649 (38.11)	—
Fair	532 (31.24)	—
Poor	171 (10.04)	—
Have regular health care provider (N=1709), n (%)	1307 (76.48)	—
Have insurance (N=1630), n (%)	1449 (88.90)	—

[a] The number of Buoy users in the analytic sample was 2437; during the period of the study, there were a total of 27,816 potentially eligible users (aged ≥18 years, US IP address, those seeking for themselves, and who completed the Buoy interview in <10 minutes) who opted not to participate.

[b] Data not available.

Drivers for Users to Use a Web-Based Symptom Checker

Users selected ≥1 of the 5 stated reasons for using Buoy, as well as open-ended responses, which were coded into 5 new categories. Over one-third (839/2437, 34.43%) of the users indicated persistent symptoms as a reason for using Buoy, followed by new symptoms (767/2437, 31.47%), symptoms not serious (545/2437, 22.36%), sensitive or embarrassing symptoms (269/2437, 11.04%), and not confident that health care provider provided correct diagnosis (220/2437, 9.03%). Less common reasons included new categories/codes: curiosity (66/2437, 2.71%), access (36/2437, 1.48%), anxiety (16/2437, 0.66%), triage (29/2437, 1.19%), and other (30/2437, 1.23%; data not shown).

Recommendations of the Web-Based Symptom Checker

We report the patterns in symptoms and diagnoses in 2 ways. First, we report the frequencies of symptoms and diagnoses organized by the NAMCS Biological Systems associated with them (Multimedia Appendix 2 [42,44]). Second, we report the top 10 symptoms and diagnosis categories, overall and by sex and race/ethnicity (Table 2).

Table 2. Top 10 symptoms and diagnoses (sorted into major categories), overall and by sex and ethnicity (N=2437).

Symptoms and diagnoses	Overall	Female (n=2069)	Male (n=368)	White (n=1227)	Latino (n=139)	Black (n=189)	Asian or Pacific Islander (n=86)
Primary symptom, proportion							
Musculoskeletal pain; headache; other pain	0.35	0.34	0.42	0.40	0.31	0.21	0.21
Axial musculoskeletal pain	0.05	0.05	0.06	0.06	0.02	0.02	0.01
Muscle pain	0.06	0.06	0.08	0.07	0.06	0.03	0.03
Joint pain	0.08	0.08	0.09	0.09	0.07	0.03	0.07
Headache	0.03	0.03	0.02	0.03	0.02	0.02	0
Chest pain	0.02	0.02	0.03	0.02	0.03	0.02	0.01
Other pain	0.11	0.10	0.13	0.11	0.10	0.08	0.08
Gynecological problems	0.12	0.14	0.00	0.08	0.16	0.22	0.26
All masses, lumps, and tumors	0.08	0.07	0.13	0.09	0.07	0.10	0.09
Edema	0.05	0.05	0.05	0.05	0.06	0.05	0.03
Skin issues	0.05	0.04	0.09	0.05	0.06	0.08	0.08
Gastrointestinal problems	0.05	0.05	0.04	0.06	0.10	0.03	0.06
Impaired sensation	0.04	0.03	0.06	0.04	0.01	0.04	0.03
Urinary tract problems	0.03	0.03	0.02	0.03	0.03	0.04	0.03
Acute upper respiratory tract symptoms	0.03	0.04	0.01	0.03	0.05	0.02	0.03
Other	0.18	0.19	0.17	0.19	0.15	0.22	0.16
Primary diagnosis, proportion							
Musculoskeletal conditions	0.12	0.12	0.17	0.14	0.08	0.07	0.07
Musculoskeletal injuries	0.05	0.05	0.05	0.06	0.04	0.04	0.02
Gynecological conditions	0.12	0.15	0	0.09	0.17	0.22	0.20
Skin problems	0.12	0.11	0.17	0.12	0.09	0.15	0.14
Infectious diseases	0.12	0.13	0.10	0.13	0.19	0.09	0.13
Digestive conditions	0.07	0.07	0.08	0.08	0.07	0.05	0.06
Neurological conditions	0.07	0.07	0.08	0.09	0.08	0.05	0.10
Cancer and benign growths	0.05	0.04	0.05	0.04	0	0.02	0.01
Urination problems	0.03	0.04	0.02	0.03	0.04	0.04	0.05
Endocrinal problems and conditions	0.03	0.03	0.02	0.03	0.01	0.04	0.03
Heart related issues	0.02	0.02	0.03	0.02	0.01	0.03	0.01
Other	0.17	0.18	0.22	0.18	0.24	0.19	0.17

Only 3 types of symptoms—pain (855/2437, 35.08%), gynecological issues (293/2437, 12.02%), and masses or lumps (204/2437, 8.37%)—accounted for almost half (1352/2437, 55.48%) of the site visits. The top 3 symptoms entered by men included pain (154/368, 41.8%), masses or lumps (49/368, 13.3%), and skin issues (33/368, 8.9%), whereas the top 3 symptoms in women included pain (701/2069, 33.88%), gynecological issues (293/2069, 14.16%), and masses or lumps (155/2069, 7.49%). Pain, gynecological issues, and masses or lumps were also reported as the top 3 symptoms for White, Black, and Asian or Pacific Islander users. The top 3 symptoms in Latino users were pain (43/139, 30.9%), gynecological issues (22/139, 15.8%), and gastrointestinal problems (14/139, 10.1%).

In comparison, Native Americans, who represented <1% of users, only entered five symptoms: pain (4/13, 31%), gynecological issues (4/13, 31%), skin issues (1/13, 8%), gastrointestinal problems (1/13, 8%), and impaired sensation (1/13, 8%).

Among the entire sample, major diagnoses were musculoskeletal (303/2437, 12.43%), gynecological (304/2437, 12.47%) and skin conditions (297/2437, 12.19%), and infectious diseases (300/2437, 12.31%). Comparably, the top 3 diagnoses reported by Buoy for men included musculoskeletal conditions (63/368, 17.1%) and skin conditions (62/368, 16.8%) and infectious diseases (37/368, 10%). The top 3 diagnoses for women included gynecological conditions (304/2069, 14.69%), infectious diseases (263/2069, 12.7%), and musculoskeletal conditions (240/2069, 11.59%). The diagnoses based on race or ethnicity followed a similar pattern. White users also reported musculoskeletal conditions (177/1227, 14.42%), infectious diseases (163/1227, 13.28%), and skin conditions (148/1227, 12.06%) as the top 3 diagnoses. Latino, Black, and Asian or Pacific Islander users reported gynecological conditions, skin conditions, and infectious diseases as the top 3 diagnoses. Gynecological conditions were reported as the top diagnosis category by Black (42/189, 22.2%) and Asian or Pacific Islander (17/86, 20%) users, whereas Latino users reported infectious diseases (26/139, 18.7%) as the top diagnosis category.

Buoy's primary recommendation was more evenly split across the less-serious triage categories. Users self-reported primary care physicians in 2 weeks (754/2141, 35.22%), self-treatment (452/2141, 21.11%), and primary care in 1 to 2 days (373/2141, 17.42%) as the top 3 recommendations provided by Buoy, followed by wait and watch (339/2141, 15.83%). Only 5.74% (123/2141) and 4.67% (100/2141) of users self-reported phone calls or in-person visits in the next 3 days and in-person visits that day or as soon as possible, respectively. The recommendations reported by Buoy closely matched primary care physicians in 2 weeks (924/2437, 37.91%), self-treatment (552/2437, 22.65%), and primary care in 1 to 2 days (456/2437, 18.71%). Most users (2098/2437, 86.09%) had 2 recommendations. Approximately 71.85% (1751/2437) had 3 recommendations, as reported by Buoy.

Users' Perceptions of the Web-Based Symptom Checking Experience

Users generally reported high levels of confidence in Buoy (mean 3.47, SD 0.97), found it useful (mean 4.18, SD 0.81) and easy to understand (mean 4.64, SD 0.53), and said that Buoy made them feel less anxious (mean 3.60, SD 1.05) and more empowered to seek medical help (mean 3.75, SD 0.96) Compared with White users, Latino and Black users had significantly more confidence in Buoy (*P*<.05), and the former also found it significantly more useful (*P*<.05; Table 3). Consistent with prior studies on trust in web-based health information sources [46-48], physicians were the most trusted source. However, Buoy was trusted more (mean 3.68, SD 0.61) than any other nonmedical source, including government agencies (mean 2.85, SD 0.95), family (mean 2.64, SD 0.76), and Google (mean 2.52, SD 0.79).

Table 3. Buoy user experience and recommendations (N=2437).

Item	Overall	Male (n=368)	Female (n=2069)	White (n=1227)	Latino (n=139)	Black (n=189)	Asian or Pacific Islander (n=86)
Comprehensibility of Buoy, mean (SD)	4.64 (0.53)	4.61 (0.49)	4.65 (0.53)	4.67 (0.50)	4.68 (0.55)	4.67 (0.53)	4.60 (0.45)
Buoy website was understandable, mean (SD)	4.60 (0.61)	4.57 (0.54)	4.60 (0.62)	4.63 (0.57)	4.63 (0.67)	4.60 (0.63)	4.57 (0.50)
Buoy website was easy to read, mean (SD)	4.66 (0.56)	4.62 (0.52)	4.67 (0.57)[a]	4.68 (0.54)	4.69 (0.59)	4.71 (0.55)	4.64 (0.48)
Language used on the Buoy website was easy to understand, mean (SD)	4.68 (0.55)	4.65 (0.51)	4.68 (0.55)	4.70 (0.51)	4.71 (0.58)	4.70 (0.54)	4.59 (0.49)
Confidence in Buoy, mean (SD)	3.47 (0.96)	3.39 (0.89)	3.49 (0.99)	3.44 (0.96)[b]	3.69 (0.92)[b]	3.63 (1.04)	3.48 (0.88)
Confidence in diagnoses, mean (SD)	3.34 (1.05)	3.27 (0.97)	3.36 (1.06)	3.29 (1.05)[b,c]	3.58 (0.99)[b]	3.53 (1.11)[c]	3.35 (0.96)
Confidence in the recommendation, mean (SD)	3.60 (1.02)	3.52 (0.95)	3.62 (1.03)	3.60 (1.01)	3.79 (0.98)	3.73 (1.09)	3.60 (0.91)
Perceived utility of Buoy, mean (SD)	4.18 (0.81)	4.14 (0.77)	4.19 (0.82)	4.16 (0.80)[d]	4.43 (0.73)[d]	4.25 (0.86)	4.20 (0.76)
Buoy enabled me to diagnose my symptoms more quickly, mean (SD)	4.15 (0.85)	4.11 (0.81)	4.16 (0.86)	4.12 (0.84)[e]	4.45 (0.75)[e,f]	4.20 (0.92)[f]	4.19 (0.80)
Using Buoy made the diagnosis of my symptoms easier, mean (SD)	4.16 (0.86)	4.12 (0.81)	4.16 (0.87)	4.13 (0.85)[b]	4.38 (0.79)[b]	4.23 (0.91)	4.14 (0.81)
Overall, I found Buoy useful to diagnose my symptoms, mean (SD)	4.23 (0.86)	4.19 (0.83)	4.24 (0.86)	4.22 (0.85)[b]	4.47 (0.75)[b]	4.31 (0.89)	4.27 (0.77)
Emotional consequences of using Buoy, mean (SD)	3.68 (0.90)	3.65 (0.79)	3.68 (0.91)	3.65 (0.88)	3.76 (1.02)	3.72 (1.00)	3.76 (0.66)
Less anxious, mean (SD)	3.60 (1.05)	3.56 (0.94)	3.61 (1.07)	3.58 (1.04)	3.70 (1.15)	3.59 (1.16)	3.67 (0.79)
Encouraged to seek help, mean (SD)	3.75 (0.96)	3.74 (0.88)	3.76 (0.98)	3.73 (0.95)	3.83 (1.11)	3.86 (1.05)	3.84 (0.76)

[a] Significant difference between sex (*P*<.05).

[b] Significant difference between White and Latino users (*P*<.05).

[c] Significant difference between White and Black users (*P*<.05).

[d] Significant difference between White and Latino users (*P*<.001).

[e] Significant difference between White and Latino users (*P*<.001).

[f] Significant difference between Latino and Black users (*P*<.05).

Relationship Between Perceptions of a Web-Based Symptom Checker and Intention to Follow Recommended Actions

Overall, most users reported intentions to follow Buoy's recommendations (1428/1886, 75.71%) and discuss Buoy's recommendations with a physician or health care professional (1198/1830, 65.44%; Table 4). Users reported the strongest intention to follow Buoy's wait and watch recommendation (mean 4.38, SD 0.90), followed by self-treatment (mean 4.33, SD 0.93), in-person visit that day or as soon as possible (mean 4.17, SD 1.01), phone call or in-person visit in the next 3 days (mean 4.05, SD 1.05), primary care physician in 2 weeks (mean 3.92, SD 1.19), and primary care physician in 1 to 2 days (mean 3.68, SD 1.26).

Intention to discuss Buoy's recommendations was positively associated with having a regular provider (odds ratio [OR] 1.37, 95% CI 1.04-1.82), and an income >US $50,000 was negatively associated (OR 0.75, 95% CI 0.57-0.98; OR 66, 95% CI 0.48-0.91; Table 5). Users aged between 35 and 44 years (OR 1.51, 95% CI 1.13-2.03) and 45 and 64 years (OR 1.57, 95% CI 1.18-2.10) had better intentions of discussing recommendations than younger users (aged 18-34 years). Compared with White users, Latino (OR 1.96, 95% CI 1.22-3.25) and Black (OR 2.37, 95% CI 1.57-3.66) users had stronger intentions to discuss recommendations with a provider, and Black users were twice as likely to intend to do so. Confidence in Buoy (OR 1.54, 95% CI 1.34-1.76), perceived utility (OR 1.32, 95% CI 1.10-1.58), and anxiety reduction because of using Buoy (OR 1.43, 95% CI 1.24-1.63) were associated with higher intention to discuss Buoy's recommendations.

Overall, users had strong intentions to follow Buoy's recommendations, and users who self-reported very good or excellent health had the strongest intention to wait or watch or self-treat (OR 1.92, 95% CI 1.04-3.65; Table 5). Those who reported Buoy as easy to read and understand were 2.2 times (95% CI 1.21-4.14) more likely to intend to wait or watch or self-treat than those who reported lower comprehensibility for Buoy. Users with health insurance (OR 2.21, 95% CI 1.36-3.57) and a regular provider (OR 1.59, 95% CI 1.11-2.28) had the strongest intentions to seek care. Confidence in Buoy (OR 1.87, 95% CI 1.56-2.25) and anxiety reduction because of Buoy (OR 1.54, 95% CI 1.29-1.83) were also associated with a higher intention to seek care.

Table 4. Intentions to follow and discuss Buoy recommendations (N=2437).

Item	Overall	Male (n=368)	Female (n=2069)	White (n=1227)	Latino (n=139)	Black (n=189)	Asian or Pacific Islander (n=86)
Intentions to follow Buoy's recommendations (n=1886), n (%)	1428 (75.71)	187 (9.91)	1241 (65.8)	908 (48.14)	116 (6.15)	149 (7.9)	62 (3.29)
Wait and watch (n=283), n (%)	249 (87.9)	24 (9.6)	225 (79.5)	146 (51.6)	23 (8.1)	29 (10.2)	22 (7.8)
Self-treatment (n=385), n (%)	339 (88.1)	50 (14.7)	289 (75.1)	226 (58.7)	34 (8.8)	32 (8.3)	8 (2.1)
Phone call or in-person visit in the next 3 days (n=107), n (%)	81 (75.7)	14 (13.1)	67 (62.6)	49 (45.8)	7 (6.5)	8 (7.5)	3 (2.8)
Primary care physician in 2 weeks (n=688), n (%)	487 (70.7)	60 (12.3)	427 (62.1)	317 (46.1)	35 (5.1)	48 (7.0)	17 (2.5)
Primary care physician in 1 to 2 days (n=336), n (%)	205 (61.0)	29 (14.1)	176 (52.4)	137 (40.8)	9 (2.7)	22 (6.5)	10 (3.0)
In-person visit that day or as soon as possible (n=87), n (%)	67 (77.0)	10 (11.5)	57 (65.5)	33 (37.9)	8 (9.2)	10 (11.5)	2 (2.3)
Intentions to discuss Buoy's recommendations (n=1830), n (%)	1198 (65.46)	156 (8.52)	1042 (56.94)	758 (41.42)	109 (5.96)	150 (8.19)	51 (2.79)

Table 5. Intentions to follow Buoy's recommendations.

Predictors	Discuss Buoy's recommendations		No medical intention		Medical intention	
	OR[a]	*P* value	OR	*P* value	OR	*P* value
Intercept	0.02 (0.01-0.06)	<.001[b]	0.04 (0.00-0.75)	.03[b]	0.02 (0.00-0.11)	<.001[b]
Age 35 to 44 years	1. 51 (1.13-2.03)	.006[b]	0.66 (0.32-1.40)	.28	1.09 (0.74-1.60)	.67
Age 45 to 64 years	1.57 (1.18-2.10)	.002[b]	0.57 (0.26-1.26)	.16	1.07 (0.74-1.55)	.70
Age ≥65 years	1.31 (0.79-2.21)	.30	0.97 (0.28-4.08)	.96	1.12 (0.58-2.27)	.74
Female	0.86 (0.62-1.20)	.39	0.79 (0.31-1.80)	.59	1.00 (0.65-1.54)	.99
Black	2.37 (1.57-3.66)	<.001[b]	0.62 (0.27-1.57)	.23	1.49 (0.89-2.54)	.14
Latino	1.96 (1.22-3.25)	.007[b]	1.56 (0.48-7.12)	.50	1.38 (0.74-2.68)	.33
Asian or Pacific Islander	1.04 (0.62-1.74)	.99	0.79 (0.24-3.23)	.72	0.82 (0.43-1.64)	.57
Other ethnicities	1.56 (0.80-3.18)	.20	0.67 (0.18-3.39)	.58	0.94 (0.41-2.28)	.89
Have insurance	0.79 (0.52-1.18)	.25	0.74 (0.24-2.01)	.57	2.21 (1.36-3.57)	.001[b]
Have regular provider	1.37 (1.04-1.82)	.03	0.51 (0.21-1.14)	.12	1.59 (1.11-2.28)	.01[b]
General health status: very good or excellent	1.09 (0.86-1.38)	.50	1.92 (1.04-3.65)	.04[b]	0.95 (0.70-1.29)	.73
Some college	0.95 (0.68-1.38)	.77	1.03 (0.41-2.46)	.95	0.89 (0.57-1.38)	.61
College degree	0.73 (0.52-1.04)	.08	0.54 (0.21-1.29)	.18	0.69 (0.43-1.08)	.11
US $50,000-99,999	0.75 (0.57-0.98)	.03[b]	1.55 (0.76-3.18)	.22	1.20 (0.85-1.70)	.31
≥US $100,000	0.66 (0.48-0.91)	.01[b]	1.74 (0.76-4.17)	.20	0.92 (0.61-1.38)	.68
Comprehensibility of Buoy	1.19 (0.93-1.53)	.17	2.24 (1.21-4.14)	.01[b]	0.90 (0.65-1.22)	.49
Confidence in Buoy	1.54 (1.34-1.76)	<.001[b]	2.23 (1.61-3.14)	<.001[b]	1.87 (1.56-2.25)	<.001[b]
Perceived utility of Buoy	1.32 (1.10-1.58)	.002[b]	1.02 (0.63-1.62)	.93	1.12 (0.90-1.39)	.32
Emotional consequences of using Buoy	1.43 (1.24-1.63)	<.001[b]	1.02 (0.66-1.53)	.93	1.54 (1.29-1.83)	<.001[b]

[a] OR: odds ratio.

[b] Significant association.

Discussion

Principal Findings

This study sought to understand who uses web-based AI-powered symptom checkers and for what purposes. The demographic profile of Buoy users was similar to that described in other studies of web-based health information seekers, suggesting that older, marginalized groups continue to be digitally excluded. Consistent with data on internet-based health-seeking behaviors more generally [49], most Buoy users were middle-aged (or younger), female, and highly educated. More research is needed to better understand older adults' web-based health information–seeking behaviors and support their medical and health decisions [50]. Although a scoping review of articles examining AI-driven symptom checkers from various perspectives found that those who do not have access to health care services are more likely to use symptom checkers [51], Buoy users overwhelmingly reported having health insurance. This finding does not negate the possibility that users were motivated by financial considerations, as most contemporary health plans require an out-of-pocket copayment. Nevertheless, this suggests that other considerations such as convenience were also salient.

Along these lines, prior research has identified an association between stigmatizing conditions and the use of symptom checkers [51]. In this study, gynecological problems were among the top 3 symptom groups. Furthermore, across presenting symptoms or diagnoses, approximately 11.04% (269/2437) of the respondents were "too embarrassed" to seek in-person care. Taken together, these findings suggest that symptom checkers might be particularly useful for users affected by conditions considered personal, embarrassing, stigmatizing, not warranting the physician's attention, or requiring potentially uncomfortable or psychologically stressful physical examinations (such as pelvic examinations).

In examining the reasons for using the tool, approximately one-third of the respondents had persistent symptoms that failed to resolve spontaneously, another one-third had new symptoms, and the rest either thought they did not need professional attention or (as mentioned previously) were too embarrassed to seek care. Thus, some patients used the symptom checker because they had significant health-related concerns; some because they lacked sufficient concern to warrant in-person care; and some because they had issues with perceived quality, cost, or convenience of available care or simply wanted a second opinion. Symptoms that persist longer than expected have been identified as strong drivers of health-related anxiety and, thus, health care use [52]. At the same time, valuing convenience and lack of trust in the health care system (factors that may be particularly prominent among young people and racially and ethnically minoritized groups, respectively) have been associated with a lower propensity to use formal health care services [53].

Regarding the user experience, users had high levels of confidence in Buoy and found it useful. Moreover, users trusted Buoy more than any other nonmedical source. Perceived confidence, utility, and trust were associated with a stronger intention to discuss Buoy's recommendations with a physician. This finding is in line with a study examining patient perspectives on the usefulness of a symptom checker [11]. Most Buoy users found the tool useful for diagnosis, and most reported that they would use it again. Although the experiences of users who discussed recommendations with their physicians varied, most felt that physicians were open to discussing the results of the tool. This is an important finding, as users may not follow recommendations to seek care if they believe that acting on the advice of a symptom checker will be questioned or even belittled by their physician, regardless of their confidence in the tool.

This study ultimately advances the understanding of web-based health information–seeking behaviors and outcomes by linking objective measures of information seeking from a web-based AI-powered system with clinical and psychosocial outcomes. The results demonstrate the potential utility of an artificially powered web-based health information tool to improve outcomes for users. Symptom checkers have been described as a means of addressing the lack of access to physicians and reducing unnecessary office visits [4].

There is a lack of research on whether the use of symptom checkers translates into medical care–seeking behaviors [4,33]. Future research should examine the effects of such tools on medical care seeking, specifically how users interpret recommendations, whether the recommendations are followed, and how user responses vary among sociodemographic groups. For example, one might surmise that individuals with limited access to care or with prior negative health care experiences might be more likely to attend to, appreciate, and follow such recommendations than their more privileged counterparts. Although symptom checkers may empower users to make more informed decisions, they might paradoxically worsen health disparities if their use were less accessible to some groups. Currently, web-based symptom checkers are mostly available for free. As web-based symptom checker companies establish partnerships with employers and health insurance companies to ensure profits, not all users may be equally ready or able to pay for symptom checking.

Limitations and Strengths

We partnered with the Buoy technical and medical staff to sample the users. Owing to our partnership approach, we were able to obtain the specific symptoms reported by the participants as the primary reason for using Buoy, as well as the possible diagnoses identified by Buoy and Buoy's triage recommendation. This allowed for the comparison and validation of self-reported data. We also obtained from Buoy the symptoms, diagnoses, triage, and sex of eligible users who opted not to participate in our study. This allowed us to compare our sample to the population of users and assess potential bias. In addition, a benefit of a collaborative approach is the potential to overcome the self-report limitations of prior studies. Thus, we obtained from Buoy the paths that individuals took and Buoy's final recommendation. We were also able to match the initial reason for the consultation to the reason reported in the survey and assess the extent to which respondents understood the recommendation and intended to act upon it. By leveraging a public or private partnership, we were able to explore the use and effects of a web-based symptom checker, which has important implications for health equity and the health care system during and after the COVID-19 crisis.

The limitations of this study include the use of cross-sectional data, which limited the ability to make any causal inferences, and the potential lack of applicability to other web-based symptom checkers. In addition, we did not assess the actual search terms entered by users. Finally, our study used a limited definition of web-based health information. Searches for symptoms using a web-based symptom checker differ from other forms of health-related information communicated through the internet. For example, web-based health communities can also be a source of social support [54] and peer-to-peer medical advice [7].

Conclusions

The results of this study demonstrate the potential utility of a web-based health information tool to empower people to seek appropriate care and reduce health-related anxiety. An interactive symptom checker might provide more personalized and potentially reliable medical information than other forms of web-based health information–seeking. Despite encouraging results suggesting that the web-based tool may fulfill unmet health information needs among women and Black and Latino adults, analyses of the user base illustrate persistent second-level digital divide effects.

For web-based symptom checkers to make a meaningful contribution, they must not only be trusted by users but also meet their diverse needs, especially those concerning usability and comprehensibility. The inability to access web-based symptom checkers may also be associated with increased disparities in access to care, particularly among groups that have lagged historically in terms of digital access and literacy. Moreover, web-based symptom checker business models may further exacerbate these disparities. In contrast, AI technologies such as Buoy have the potential to alleviate disparities by allowing users to access accurate, actionable, and personalized advice within an evolving but often confusing web-based health information environment. Finally, there is a lack of evidence on whether web-based symptom checkers influence care-seeking behaviors. To address this gap, future research will use Buoy users' follow-up data to assess the extent to which users discuss their web-based findings with physicians, as well as barriers to the same and patient satisfaction.

Acknowledgments

This project was supported by a gift from Buoy Health to the University of California, Merced. Buoy Health staff provided access to participants and internal data used to validate self-reported data but had no role in data analysis or interpretation. The authors gratefully acknowledge Jovo Velasco, Stephanie Reyes, and Michael Vang for their data collection and coding support in this research. The contents are solely the responsibility of the authors and do not necessarily represent the official views of Buoy Health or the University of California.

Conflicts of Interest

The authors received funding and some technical assistance in recruiting from Buoy Health, although they were not involved in the research design, data analysis, or interpretation.

Multimedia Appendix 1
Buoy triage levels and recommendations.
[DOC File , 38 KB - jmir_v24i8e36322_app1.doc]

Multimedia Appendix 2
Symptom and diagnosis codes and frequencies.
[DOCX File , 24 KB - jmir_v24i8e36322_app2.docx]

References

1. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020 Nov 30;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]
2. McCarthy M. Sixty seconds on . . . symptom checkers. BMJ 2016 Oct 20;355:i5596. [doi: 10.1136/bmj.i5596] [Medline: 27765782]
3. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web use for symptom appraisal of physical health conditions: a systematic review. J Med Internet Res 2017 Jun 13;19(6):e202 [FREE Full text] [doi: 10.2196/jmir.6755] [Medline: 28611017]
4. Semigran HL, Linder JA, 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]
5. Morita T, Rahman A, Hasegawa T, Ozaki A, Tanimoto T. The potential possibility of symptom checker. Int J Health Policy Manag 2017 Oct 01;6(10):615-616 [FREE Full text] [doi: 10.15171/ijhpm.2017.41] [Medline: 28949479]
6. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 8;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
7. Fox S. After Dr Google: peer-to-peer health care. Pediatrics 2013 Jun;131 Suppl 4:S224-S225. [doi: 10.1542/peds.2012-3786K] [Medline: 23729765]
8. Hartzband P, Groopman J. Untangling the Web--patients, doctors, and the Internet. N Engl J Med 2010 Mar 25;362(12):1063-1066. [doi: 10.1056/NEJMp0911938] [Medline: 20335581]
9. Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of self-diagnosis health chatbots in real-world settings: case study. J Med Internet Res 2021 Jan 06;23(1):e19928 [FREE Full text] [doi: 10.2196/19928] [Medline: 33404508]
10. Tsai CH, You Y, Gui X, Kou Y, Carroll JM. Exploring and promoting diagnostic transparency and explainability in online symptom checkers. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 1-17. [doi: 10.1145/3411764.3445101]
11. Meyer AN, Giardina TD, Spitzmueller C, Shahid U, Scott TM, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
12. Katz E, Blumler JG, Gurevitch M. Uses and gratifications research. Public Opin Q 1973;37(4):509-523. [doi: 10.1086/268109]
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. Massey PM. Where Do U.S. adults who do not use the Internet get health information? Examining digital health information disparities from 2008 to 2013. J Health Commun 2016;21(1):118-124. [doi: 10.1080/10810730.2015.1058444] [Medline: 26166484]

15. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
16. Ramirez AS, Leyva B, Graff K, Nelson DE, Huerta E. Seeking information on behalf of others: an analysis of calls to a Spanish-language radio health program. Health Promot Pract 2015 Jul;16(4):501-509. [doi: 10.1177/1524839915574246] [Medline: 25716191]
17. Abrahamson JA, Fisher KE, Turner AG, Durrance JC, Turner TC. Lay information mediary behavior uncovered: exploring how nonprofessionals seek health information for themselves and others online. J Med Libr Assoc 2008 Oct;96(4):310-323 [FREE Full text] [doi: 10.3163/1536-5050.96.4.006] [Medline: 18974809]
18. Kelly B, Hornik R, Romantan A, Schwartz JS, Armstrong K, DeMichele A, et al. Cancer information scanning and seeking in the general population. J Health Commun 2010 Oct;15(7):734-753 [FREE Full text] [doi: 10.1080/10810730.2010.514029] [Medline: 21104503]
19. Niederdeppe J, Hornik RC, Kelly BJ, Frosch DL, Romantan A, Stevens RS, et al. Examining the dimensions of cancer-related information seeking and scanning behavior. Health Commun 2007;22(2):153-167. [doi: 10.1080/10410230701454189] [Medline: 17668995]
20. Rooks RN, Wiltshire JC, Elder K, BeLue R, Gary LC. Health information seeking and use outside of the medical encounter: is it associated with race and ethnicity? Soc Sci Med 2012 Jan;74(2):176-184. [doi: 10.1016/j.socscimed.2011.09.040] [Medline: 22154611]
21. Grasso KL, Bell RA. Understanding health information seeking: a test of the risk perception attitude framework. J Health Commun 2015;20(12):1406-1414. [doi: 10.1080/10810730.2015.1018634] [Medline: 26161622]
22. Shim M, Kelly B, Hornik R. Cancer information scanning and seeking behavior is associated with knowledge, lifestyle choices, and screening. J Health Commun 2006;11 Suppl 1:157-172. [doi: 10.1080/10810730600637475] [Medline: 16641081]
23. Fareed N, Swoboda CM, Jonnalagadda P, Walker DM, Huerta TR. Differences between races in health information seeking and trust over time: evidence from a cross-sectional, pooled analyses of HINTS data. Am J Health Promot 2021 Jan;35(1):84-92. [doi: 10.1177/0890117120934609] [Medline: 32588638]
24. Jiang S, Street RL. Pathway linking Internet health information seeking to better health: a moderated mediation study. Health Commun 2017 Aug;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
25. Ramírez AS, Freres D, Martinez LS, Lewis N, Bourgoin A, Kelly BJ, et al. Information seeking from media and family/friends increases the likelihood of engaging in healthy lifestyle behaviors. J Health Commun 2013;18(5):527-542 [FREE Full text] [doi: 10.1080/10810730.2012.743632] [Medline: 23472825]
26. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
27. Moldovan-Johnson M, Tan AS, Hornik RC. Navigating the cancer information environment: the reciprocal relationship between patient-clinician information engagement and information seeking from nonmedical sources. Health Commun 2014;29(10):974-983 [FREE Full text] [doi: 10.1080/10410236.2013.822770] [Medline: 24359259]
28. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the Internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
29. Cocco AM, Zordan R, Taylor DM, Weiland TJ, Dilley SJ, Kant J, et al. Dr Google in the ED: searching for online health information by adult emergency department patients. Med J Aust 2018 Oct 15;209(8):342-347. [doi: 10.5694/mja17.00889] [Medline: 30107763]
30. Buoy Health: Check Symptoms & Find the Right Care. URL: https://www.buoyhealth.com/ [accessed 2022-08-10]
31. Mack H. Digital health startup Buoy launches AI-powered, symptom-checking chatbot. MobiHealthNews. 2017 Mar 8. URL: https://www.mobihealthnews.com/content/digital-health-startup-buoy-launches-ai-powered-symptom-checking-chatbot [accessed 2022-08-10]
32. Buoy Health - A chatbot that helps diagnose your symptoms. Product Hunt. 2017. URL: https://www.producthunt.com/posts/buoy-health [accessed 2022-08-10]
33. Winn AN, Somai M, Fergestrom N, Crotty BH. Association of use of online symptom checkers with patients' plans for seeking care. JAMA Netw Open 2019 Dec 02;2(12):e1918561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18561] [Medline: 31880791]
34. Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. Med Care 2005 Jun;43(6):607-615. [doi: 10.1097/01.mlr.0000163658.65008.ec] [Medline: 15908856]
35. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach. New York, NY, USA: Psychology Press; 2010.
36. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-484. [doi: 10.1080/10810730490504233] [Medline: 15513791]

37. Sbaffi L, Rowley J. Trust and credibility in Web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
38. Sivakumar G, Mares ML. The doctor versus the Internet: effects of low-, medium-, and high-quality websites on intentions to follow the doctor's advice. Health Commun 2017 Dec;32(12):1454-1462. [doi: 10.1080/10410236.2016.1228030] [Medline: 27767346]
39. van Weert JC, van Noort G, Bol N, van Dijk L, Tates K, Jansen J. Tailored information for cancer patients on the Internet: effects of visual cues and language complexity on information recall and satisfaction. Patient Educ Couns 2011 Sep;84(3):368-378. [doi: 10.1016/j.pec.2011.04.006] [Medline: 21550757]
40. 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]
41. White RW, Horvitz E. Experiences with Web search on medical concerns and self diagnosis. AMIA Annu Symp Proc 2009 Nov 14;2009:696-700 [FREE Full text] [Medline: 20351943]
42. Schneider D, Appleton L, McLemore T. A reason for visit classification for ambulatory care. Vital Health Stat 2 1979 Feb(78):i-63 [FREE Full text] [Medline: 433160]
43. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
44. Shapiro MF, Ware Jr JE, Sherbourne CD. Effects of cost sharing on seeking care for serious and minor symptoms. Results of a randomized controlled trial. Ann Intern Med 1986 Feb;104(2):246-251. [doi: 10.7326/0003-4819-104-2-246] [Medline: 3946953]
45. Li J, Theng YL, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2016 Dec;22(4):804-814 [FREE Full text] [doi: 10.1177/1460458215595851] [Medline: 26261218]
46. Kwon JH, Kye SY, Park EY, Oh KH, Park K. What predicts the trust of online health information? Epidemiol Health 2015 Jun 28;37:e2015030 [FREE Full text] [doi: 10.4178/epih/e2015030] [Medline: 26212505]
47. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, 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 Dec;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
48. Thai CL, Gaysynsky A, Falisi A, Chou WY, Blake K, Hesse BW. Trust in health information sources and channels, then and now: evidence from the health information national trends survey (2005–2013). In: Hale TM, Chou WY, Cotten SR, Khilani A, editors. eHealth: Current Evidence, Promises, Perils and Future Directions (Studies in Media and Communications, Vol. 15). Bingley, UK: Emerald Publishing Limited; 2018:43-67.
49. 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]
50. 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]
51. Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The use of artificially intelligent self-diagnosing digital platforms by the general public: scoping review. JMIR Med Inform 2019 May 01;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]
52. Kravitz RL, Callahan EJ, Paterniti D, Antonius D, Dunham M, Lewis CE. Prevalence and sources of patients' unmet expectations for care. Ann Intern Med 1996 Nov 01;125(9):730-737. [doi: 10.7326/0003-4819-125-9-199611010-00004] [Medline: 8929006]
53. Musa D, Schulz R, Harris R, Silverman M, Thomas SB. Trust in the health care system and the use of preventive health services by older black and white adults. Am J Public Health 2009 Jul;99(7):1293-1299. [doi: 10.2105/AJPH.2007.123927] [Medline: 18923129]
54. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 01;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]

Abbreviations

AI: artificial intelligence
NAMCS: National Ambulatory Medical Care Survey
OR: odds ratio
RQ: research question

Edited by T Leung; submitted 17.01.22; peer-reviewed by N Fareed, S Song, Y Qian; comments to author 16.03.22; revised version received 21.04.22; accepted 24.05.22; published 19.08.22.

Please cite as:
Arellano Carmona K, Chittamuru D, Kravitz RL, Ramondt S, Ramírez AS
Health Information Seeking From an Intelligent Web-Based Symptom Checker: Cross-sectional Questionnaire Study
J Med Internet Res 2022;24(8):e36322
URL: https://www.jmir.org/2022/8/e36322
doi:10.2196/36322
PMID:35984690

Review

Misinformation About COVID-19 Vaccines on Social Media: Rapid Review

Ingjerd Skafle[1,2], MA; Anders Nordahl-Hansen[3], PhD; Daniel S Quintana[4,5,6,7], PhD; Rolf Wynn[8,9], PhD; Elia Gabarron[3,10], PhD

[1]Faculty of Health, Welfare, and Organisation, Østfold University College, Halden, Norway
[2]Faculty of Medicine, University of Oslo, Oslo, Norway
[3]Department of Education, ICT, and Learning, Østfold University College, Halden, Norway
[4]Department of Psychology, University of Oslo, Oslo, Norway
[5]KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway
[6]Norwegian Centre for Mental Disorders Research (NORMENT), University of Oslo, Oslo, Norway
[7]NevSom, Department of Rare Disorders & Disabilities, Oslo University Hospital, Oslo, Norway
[8]Department of Clinical Medicine, The Artic University of Norway, Tromsø, Norway
[9]Division of Mental Health and Substance Use, University Hospital of North Norway, Tromsø, Norway
[10]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

Corresponding Author:
Ingjerd Skafle, MA
Faculty of Health, Welfare, and Organisation
Østfold University College
B R A Veien 4
Halden, 1757
Norway
Phone: 47 69 60 80 00
Email: ingjerd.skafle@hiof.no

Abstract

Background: The development of COVID-19 vaccines has been crucial in fighting the pandemic. However, misinformation about the COVID-19 pandemic and vaccines is spread on social media platforms at a rate that has made the World Health Organization coin the phrase *infodemic*. False claims about adverse vaccine side effects, such as vaccines being the cause of autism, were already considered a threat to global health before the outbreak of COVID-19.

Objective: We aimed to synthesize the existing research on misinformation about COVID-19 vaccines spread on social media platforms and its effects. The secondary aim was to gain insight and gather knowledge about whether misinformation about autism and COVID-19 vaccines is being spread on social media platforms.

Methods: We performed a literature search on September 9, 2021, and searched PubMed, PsycINFO, ERIC, EMBASE, Cochrane Library, and the Cochrane COVID-19 Study Register. We included publications in peer-reviewed journals that fulfilled the following criteria: original empirical studies, studies that assessed social media and misinformation, and studies about COVID-19 vaccines. Thematic analysis was used to identify the patterns (themes) of misinformation. Narrative qualitative synthesis was undertaken with the guidance of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 Statement and the Synthesis Without Meta-analysis reporting guideline. The risk of bias was assessed using the Joanna Briggs Institute Critical Appraisal tool. Ratings of the certainty of evidence were based on recommendations from the Grading of Recommendations Assessment, Development and Evaluation Working Group.

Results: The search yielded 757 records, with 45 articles selected for this review. We identified 3 main themes of misinformation: medical misinformation, vaccine development, and conspiracies. Twitter was the most studied social media platform, followed by Facebook, YouTube, and Instagram. A vast majority of studies were from industrialized Western countries. We identified 19 studies in which the effect of social media misinformation on vaccine hesitancy was measured or discussed. These studies implied that the misinformation spread on social media had a negative effect on vaccine hesitancy and uptake. Only 1 study contained misinformation about autism as a side effect of COVID-19 vaccines.

Conclusions: To prevent these misconceptions from taking hold, health authorities should openly address and discuss these false claims with both cultural and religious awareness in mind. Our review showed that there is a need to examine the effect of social media misinformation on vaccine hesitancy with a more robust experimental design. Furthermore, this review also demonstrated that more studies are needed from the Global South and on social media platforms other than the major platforms such as Twitter and Facebook.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021277524; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021277524

International Registered Report Identifier (IRRID): RR2-10.31219/osf.io/tyevj

(J Med Internet Res 2022;24(8):e37367) doi:10.2196/37367

KEYWORDS

social media; misinformation; COVID-19 vaccines; vaccination hesitancy; autism spectrum disorder

Introduction

Background

An unprecedented global effort has been undertaken to develop vaccines that protect against COVID-19. However, there is a grave concern that vaccine hesitancy will be a major obstacle to reaching herd immunity. In 2019, the World Health Organization (WHO) had already named vaccine hesitancy as 1 of 10 threats to global health [1]. Global vaccine distribution equity is also a major challenge. Figures from February 2022 show that 61.9% of the world's population has received at least one dose of a COVID-19 vaccine, but only 10.6% of people in the Global South have received a dose [2]. Furthermore, the rate of people receiving a COVID-19 vaccine in some high-income countries where vaccines are available and free has dropped [3]. The WHO reiterates that COVID-19 vaccines remain critical and are considered effective against severe disease and death [4].

The reasons behind COVID-19 vaccine hesitancy are complex. Fear of side effects and concerns about the pace at which the vaccines were developed have been cited as primary reasons behind this hesitancy [5]. In addition, misinformation about COVID-19 and vaccines has spread on social media platforms at a rate that has made the WHO coin the phrase infodemic [6]. An infodemic is "too much information including false or misleading information in digital and physical environments during a disease outbreak" [7,8].

A well-known false claim is that the measles, mumps, and rubella (MMR) vaccine can cause autism [9]. The claim has since been empirically refuted many times but is still stated as a major concern for some parents [10]. Motta and Steccula [11] examined American public opinion data on MMR safety collected before and after a retracted 1998 study linking autism to MMR. The researchers detected a statistically significant increase in public concern about MMR safety following the retracted study and the media attention it received. This suggests that misleading vaccine information can impact public confidence in vaccines and cause skepticism about vaccines in general. Since the retracted 1998 study, groups of vaccine deniers or *antivaxxers* have grown, and claims that vaccines are harmful have spread to almost all vaccines [12]. Pullan and Dey [13] analyzed search patterns in Google Trends during the early stage of the pandemic in 2020 and found that search interest in COVID-19 vaccines had understandably increased, but also found that well-known antivaccine searches such as "autism" and "mercury" also had a growing presence and similar spikes as search patterns for COVID-19 vaccines. These results confirm that the false claim of associations between autism and MMR vaccines has become an argument for all types of vaccines and also possibly a concern when it comes to COVID-19 vaccines. Therefore, we examined whether misinformation on social media in recent times linked autism to COVID-19 vaccines.

Furthermore, vaccine hesitancy based on misinformation seems to be a worldwide phenomenon regardless of the uneven distribution of COVID-19 vaccines [4]. Social media plays a crucial role in disseminating both correct information and misinformation about infectious diseases and vaccines [14]. Wilson and Wiysonge [15] showed, in a global cross-national analysis of geographically coded tweets and vaccination rates from 166 countries, that there was a significant relationship between social media use and vaccine hesitancy. However, there has been a joint effort by several of the largest social media platforms and technology companies to combat the spread of misinformation about COVID-19 [16].

Objective

We aimed to synthesize the existing research on misinformation about COVID-19 vaccines spread on social media platforms and its effects. The secondary aim was to gain insight and gather knowledge about whether misinformation about autism and COVID-19 vaccines is being spread on social media platforms. The following questions guided our inquiry: What is known about misinformation regarding COVID-19 vaccines spread on social media platforms? What is known about the effects of misinformation about COVID-19 vaccines spread on social media platforms? What is known about social media misinformation on COVID-19 vaccines concerning autism spectrum disorder?

Methods

Design

We followed the guidance from Cochrane Rapid Reviews [17]. We chose a rapid review protocol in line with the recommendations by Cochrane; that is, the need "for timely evidence for decision-making purposes including to address urgent and emergent health issues and questions deemed to be of high priority” [17]. The need to address vaccine hesitancy toward COVID-19 vaccines is an emergent health issue. The narrative qualitative synthesis was undertaken with the guidance of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 Statement [18] and the Synthesis Without Meta-analysis reporting guideline [19].

Search Strategy and Selection Criteria

With the help and expertise of an information retrieval specialist, we used the following search string in this rapid review: (“misinformation” OR “disinformation” OR “information”) AND (“social media” OR “Facebook” OR “Twitter” OR “Instagram” OR “WhatsApp” OR “Telegram” OR “Tumblr” OR “Pinterest” OR “YouTube” OR “VKontakte” OR “Snapchat” OR “TikTok” OR “Weibo” OR “WeChat” OR “Reddit”) AND (“covid*” OR “corona*” OR “pandemic” OR “Sars-CoV-2” AND “vaccine*” OR “vaccination*”).

No date or language limitations were used. The full search strategy of the information retrieval specialist is available in Multimedia Appendix 1.

Publications were excluded if the studies were not original empirical research, if studies examined vaccines in general and not COVID-19 vaccines, if studies did not examine social media misinformation, and if data were gathered before the COVID-19 vaccine Pfizer-BioNTech phase 3 clinical trial [20].

Data Collection Process and Extraction

This review was registered with the PROSPERO international register of systematic reviews (CRD42021277524). Systematic searches in the PubMed, PsycINFO, ERIC, EMBASE, Cochrane Library, and Cochrane COVID-19 Study Register databases were conducted by an information retrieval specialist on September 9, 2021. Duplicates were identified and removed by IS and EG. We used Rayyan [21] as the screening tool. Rayyan is a web application and mobile app for systematic reviews. It eases the process of the initial screening of abstracts and titles and helps researchers save time when they share and compare include-exclude decisions. All titles and abstracts were screened by IS and ANH independently. In the initial search, no date restriction was set. However, during the piloting of the title and abstract screening, IS and ANH discussed the fact that there were studies that explored misinformation about COVID-19 vaccines at a very early stage in the pandemic, before any COVID-19 vaccines were a reality. We decided that we needed a threshold date as to when we believed we found misinformation about the actual COVID-19 vaccines to be relevant, as misinformation at a very early stage would be about a potential vaccine. Therefore, we decided to include studies that were conducted during and after the Pfizer-BioNTech phase 3 clinical trial, because then the news about an actual vaccine was starting to spread around the world and thus starting to become a reality. We chose the Pfizer-BioNTech vaccine because it was the first COVID-19 vaccine to be approved by the WHO [22].

Of the 319 titles and abstracts screened, IS and ANH disagreed on 35. The disagreements were resolved through discussions between the 2 reviewers and if an extra opinion was needed, EG was consulted. Of these articles, 1 article was in German, 2 were in Spanish, and the rest were in English. IS can understand German and EG speaks Spanish. IS and EG performed a further assessment of the eligibility of the full-text records and conducted a pilot exercise using the same 10 full-text articles to calibrate and test the review form. After the screening, both reviewers assessed the articles that the other had excluded. ANH assisted with conflicts and discussed doubts surrounding the included or excluded articles. The data extraction from the included articles involved 2 reviewers (IS and EG), where IS extracted data using a piloted form and EG checked for the correctness and completeness of the extracted data. Data from the included articles were extracted based on design and study population, type of misinformation, effect of misinformation, misinformation about autism, ethical considerations, and social media channels. The agreed evidence was then synthesized narratively.

To synthesize the knowledge gathered about the types of misinformation, a thematic analysis was performed [23]. After the data extraction, IS gathered the data on the content of the misinformation. The data extracts on misinformation were then coded by ANH. IS and ANH searched for themes based on the codes and agreed upon 3 final themes of misinformation: conspiracies, medical misinformation, and vaccine development. EG approved the themes. Multimedia Appendix 2 provides an overview of the thematic analysis that was undertaken.

Assessment of Risk of Bias

The risk of bias was graded according to the Joanna Briggs Institute (JBI) Critical Appraisal tool “Checklist for Analytical Cross-sectional Studies” [24] by 1 experienced reviewer (DSQ). The evaluation was based on answers to 8 questions (yes, no, or not applicable). The studies were classified as having low (>70%), moderate (40%-70%), or high (<40%) risk of bias. A complete overview of the assessment can be found in Multimedia Appendix 3 [25-69].

Assessment of the Quality of the Evidence

One experienced reviewer (RW) assigned certainty of evidence ratings based on recommendations by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group [70]. The included studies that looked at associations were given a narrative GRADE score related to the outcome “Association between social media misinformation and vaccine hesitancy.” The level of quality of evidence was classified as very low, low, moderate, or high. A complete overview of the assessment can be found in Multimedia Appendix 4 [25-69].

Data Synthesis

Narrative synthesis was undertaken with the guidance of the PRISMA 2020 Statement [18] and Synthesis Without Meta-analysis reporting guideline [19]. In the synthesis, findings from our included studies were grouped according to study design, population, social media sample, types of social media, types of misinformation reported, misinformation about autism, the reported effect of the misinformation on vaccine hesitancy, and the assessments of risk of bias and quality of evidence. When synthesizing the findings narratively, studies with a low risk of bias or high quality of evidence will be highlighted on several occasions.

Results

Study Selection and Risk of Bias

We identified 45 relevant studies (Figure 1). The list of excluded articles during the full-text review and the reasons for exclusion are reported in Multimedia Appendix 5. The risk of bias in 53% (24/45) of the included studies was classified as low, according to the JBI Critical Appraisal tool [25-42,47-52]. In total, 18% (8/45) of the studies showed a moderate risk of bias [43-46,53-56]. Finally, 27% (12/45) of the included studies showed a high risk of bias [57-68]. Of the 45 studies, in 1 (2%) study [69], none of the questions in the JBI tool were applicable.

We grouped the studies into 2 major categories according to data sampling. One group gathered data through surveys, interviews, or focus groups (Table 1). The other group gathered data from social media platforms (Table 2). The largest total population sample in the first group of 22 studies (Table 1) came from Europe, with 27,975 respondents in total. All respondents were described as adults or >18 years, except for 2 studies in the United States where the participants were aged ≥65 years [28,34]. Another exception was 1 study from Slovenia, where participants aged ≥15 years were included [35]. Gender has not been a focal point in any of the 45 included studies.

Data were extracted from social media platforms in 23 of the included studies. These studies formed the second group (Table 2).

The 12 studies that were assessed to have a high risk of bias were found in the second group (Table 2), whereas the studies in Table 1 had a low or moderate risk of bias according to the JBI tool.

Many of the studies did not name social media platforms in the first group (Table 1) but rather discussed social media platforms in general. However, some studies did specify which social media platforms they were assessing. Figure 2 summarizes the types of social media platforms specified in the 45 included studies.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for new systematic reviews.

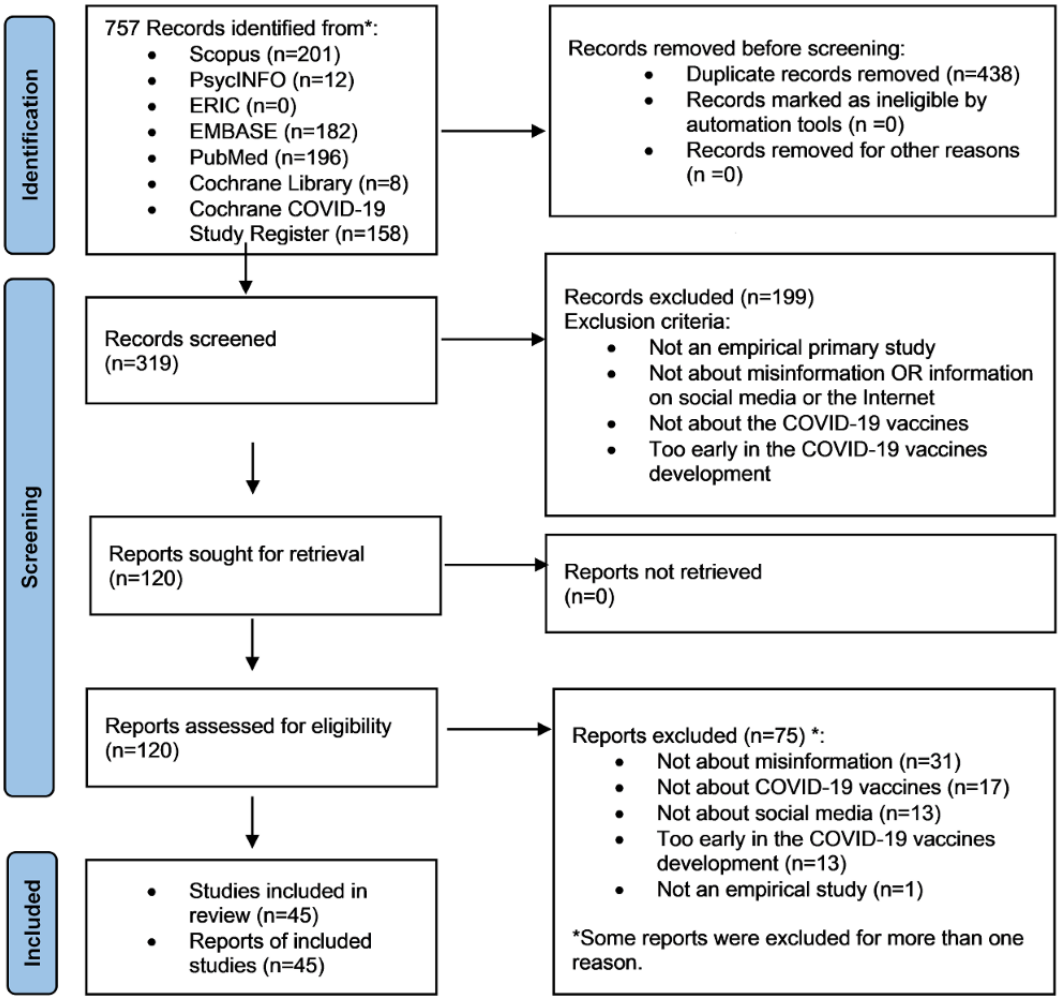


Table 1. Studies in which data were collected through surveys, observations, or interviews (n=22).

Study	Country	Study period	Study design	Type of social media	Social media or population sample	Type of misinformation reported	Risk of bias (JBI[a])
Alibrahim and Awad [25], 2021	Kuwait	March 26 to April 26, 2021	Cross-sectional study	Not specified	4147 adults, ≥18 years	COVID-19 is not a serious infection that requires vaccination	Low
Allington et al [26], 2021	United Kingdom	November 21 to December 21, 2020	Cross-sectional study	Not specified	4343 UK residents, aged 18-75 years	Conspiracy theories (not specified)	Low
Aloweidi et al [27], 2021	Jordan	January 22 to February 28, 2021	Cross-sectional study	Not specified	646 adults	The vaccines are unsafe; effect of the vaccines on a genetic level; causes chronic illnesses; may lead to infertility; can affect their offspring; contains toxic heavy metals and neurotoxic materials; it is a part of a secret research	Low
Bhagianadh and Arora [28], 2021	United States	October to November 2020	Longitudinal survey	Not specified	5784 Medicare enrollees, ≥65 years	Distrust of government narrative about vaccines; vaccine will cause COVID-19	Low
Brodziak et al [29], 2021	Poland	January 26 to February 28, 2021	Survey	Not specified	635 adult patients with cancer	The vaccine contains bodies of aborted children; COVID-19 does not exist	Low
Chadwick et al [30], 2021	United Kingdom	September 24 to October 17, 2020	Cross-sectional study	Not specified	5114 adults in the United Kingdom	Conspiracies (not specified)	Low
Ebrahimi et al [31], 2021	Norway	January 23 to February 2, 2021	Cross-sectional study	Not specified	4571 Norwegian adults	Not specified	Low
Kanyike et al [32], 2021	Uganda	Monday, March 15, and Sunday, March 21, 2021	Cross-sectional study	Not specified	600 medical students, ≥18 years	Negative information about COVID-19	Low
Karabela et al [33], 2021	Turkey	February 1, 2021, to February 28, 2021	Cross-sectional study	Social media, WhatsApp, and YouTube	1216 adults	Conspiracy theories (not specified)	Low
Park et al [34], 2021	United States	October to November 2020	Cross-sectional study	Not specified	6478 Medicare beneficiaries	The belief that COVID-19 is not that dangerous	Low
Petravić et al [35], 2021	Slovenia	December 17 to December 27, 2020	Cross-sectional study	Not specified	12,042 Slovenian residents, ≥15 years. Analysis of responses from the 2320 respondents (12%) who answered the open-ended question	The vaccines will cause a genocide; COVID-19 is the same as influenza	Low
Sallam et al [36], 2021	Jordan	January 19 to January 23, 2021	Cross-sectional study	Not specified	1106 university students	COVID-19 was man-made for enforcing vaccinations; COVID-19 vaccinations intends to implant microchips into people to control them; COVID-19 vaccination will lead to infertility	Low
Sallam et al [37], 2021	Jordan, Kuwait, and Saudi Arabia	December 4 to December 18, 2020	Cross-sectional study	Facebook, Instagram, Twitter, and WhatsApp	3414 respondents	An artificial origin of the virus; the disease was man-made to enforce vaccination; microchip implanting and infertility claims	Low

Study	Country	Study period	Study design	Type of social media	Social media or population sample	Type of misinformation reported	Risk of bias (JBI[a])
Sharevski and Gover [38], 2021	United States	January and February 2021	Cross-sectional quasi-experimental study	Twitter	304 respondents, ≥18 years	Exaggeration of rare side effects of COVID-19 vaccines	Low
Zhang et al [39], 2021	China	September 1 to September 7, 2020	Cross-sectional study	WeChat, WeChat moments, Weibo, TikTok	2053 Chinese factory workers (full-time employees) ≥18 years	Negative information about COVID-19 vaccines	Low
Zhang et al [40], 2021	China	September 1 to September 7, 2020	Cross-sectional study	WeChat, WeChat moments, Weibo, TikTok	2053 Chinese parents, ≥18 years	Negative information about COVID-19 vaccines	Low
Costantino et al [41], 2021	Italy	December 2020 to March 2021	Cross-sectional study	Not specified	363 adults	Unfavorable information about COVID-19 vaccines	Low
Jennings et al [42], 2021	United Kingdom	Survey: December 12 to December 18, 2020. Focus groups: November 30 to December 7, 2020	Cross-sectional qualitative and quantitative (mixed method) study	TikTok, Instagram, Snapchat, Twitter; Facebook, YouTube	1476 UK adults participated in the survey; 29 adults in the United Kingdom participated in the focus groups	Conspiracy theories (not specified)	Low
El-Far Cardo et al [43], 2021	Germany	August and November 2020	Cross-sectional study	Facebook, Twitter, Telegram	808 persons	COVID-19 is not a health threat	Moderate
Knights et al [44], 2021	United Kingdom	June 18 and November 30, 2020	Cross-sectional qualitative study	Not specified	64 primary care professionals and administrative staff and 17 recently arrived migrants	5G conspiracy theory	Moderate
Berry et al [45], 2021	United States	December 30, 2020, to January 15, 2021	Qualitative observational study	Not specified	193 skilled nursing facility workers	Vaccines cause COVID-19; microchip; the virus has been around for a long time and killed many people since 1918; fear of racist motives and the safety of the vaccines; the vaccines have fetal cells from abortions	Moderate
Choudhary et al [46], 2021	India	February 18 to February 28, 2021	Cross-sectional study	Not specified	272 Indian adults, ≥18 years	COVID-19 is a conspiracy	Moderate

[a] JBI: Joanna Briggs Institute.

Table 2. Studies in which data were collected from social media platforms (n=23).

Study	Country	Study period	Study design	Type of social media	Social Media or population sample	Type of misinformation reported	Risk of bias (JBI[a])
Chan et al [47], 2021	The United Kingdom	December 10, 2020	Cross-sectional study (social media data extraction)	YouTube	48 COVID-19 vaccine–related videos on YouTube	Misinformation about COVID-19 vaccines (not specified). Only 2 (4.2%) videos made nonfactual claims.	Low
Herrera-Peco et al [48], 2021	Spain	December 14 to December 28, 2020	Cross-sectional study (social media data extraction)	Twitter	5040 Twitter users participated, generating a total of 1,664,261 impressions	Messenger RNA vaccines will produce changes in human DNA; government and pharmaceutical industries are allies; adverse effects leading to genocide.	Low
Hughes et al [49], 2021	United States	October 2020 to November 2020	Cross-sectional study (social media data extraction and modeling)	Facebook, Twitter, YouTube, and Instagram	Using hashtag and keyword searchers, a team of subject matter experts identified 20 channels (ie, bounded sources of content, such as a social media account), which appeared to contain a high degree of antivaccine content or COVID denialism.	Corrupt elites; physical deformities; mental illness; microchips that violate your autonomy and privacy; the people who intentionally created the COVID vaccine are shadowy and suspicious.	Low
Larrondo-Ureta [50], 2021	Spanish-speaking countries	December 2020 and February 2021	Cross-sectional study (social media data extraction)	Twitter	62,045 tweets and 258,843 retweets	Antivaccine discourse (not specified)	Low
Liu and Liu [51], 2021	English-speaking countries	November 1 to November 22, 2020	Cross-sectional study (social media data extraction)	Twitter	5000 COVID-19 vaccine–related tweets, which were posted by 4796 unique users.	Microchips; alters DNA; women become sterile.	Low
Sobkowicz and Sobkowicz [52], 2021	United States and Poland	March 1, 2021	Cross-sectional study (social media data extraction)	Reddit and Interia	Reddit and Interia antivaccine groups	Antivaccination discussions about COVID-19 vaccines.	Low
Guntuku et al [53], 2021	United States	December 1, 2020, to February 28, 2021	Cross-sectional study (social media data extraction)	Twitter	78.1 million vaccine-related tweets	Evangelical hubs posted conspiracy theories about Bill Gates and China.	Moderate
Hernández-García et al [54], 2021	Spain	February 9 2021	Cross-sectional study (social media data extraction)	YouTube	118 YouTube videos	Hoaxes and conspiracy theories (not specified).	Moderate

Study	Country	Study period	Study design	Type of social media	Social Media or population sample	Type of misinformation reported	Risk of bias (JBI[a])
Islam et al [55], 2021	Australia	December 31, 2019, to November 30, 2020	Cross-sectional study (social media data extraction)	Facebook, YouTube, and Twitter	637 news articles, social media narratives, web-based reports, and blogs spread on social media	Daughter of the Russian president had died after receiving the second dose of COVID-19 vaccine; children and soldiers dying after receiving the vaccine in multiple countries; conspiracy theory about Bill Gates; COVID-19 vaccine can monitor the human population and take over the world; COVID-19 vaccines contain a microchip through which biometric data could be collected, and large businesses could send signals to the chips using 5G networks; crucial phases of the clinical trials were skipped; COVID-19 vaccine contains cells from aborted fetus or genes from pigs.	Moderate
Kwok et al [56], 2021	Australia	January 22 and October 20, 2020	Cross-sectional study (social media data extraction)	Twitter	31,100 COVID-19 vaccine–related tweets	Conspiracy theories such as the “mark of the beast” and microchips in vaccines.	Moderate
Alliheibi et al [5,8], 2021	Saudi Arabia	December 15, 2020, to May 25, 2021	Cross-sectional study (social media data extraction)	Twitter	37,467 Arabic tweets from 23,748 users	COVID-19 vaccination is a cover for a plan devised by Bill Gates to implant trackable microchips to control people.	High
Baines et al [58], 2021	United States	November 20, 2020, to January 6, 2021	Cross-sectional study (social media data extraction)	Parler	400 random parleys from a large sample of 7000 parleys	Sterilization possibilities for men and women; COVID-19 vaccine to control the population; Bill Gates and Anthony Fauci had instigated measures (ie, microchips and enzymes in the vaccine) to control the population through the administration of the COVID-19 vaccine; governments and certain powerful individuals “planned” this health crisis to vaccinate children without parental consent as part of the new world order to control future populations.	High
Basch et al [59], 2021	United States	December 2020	Cross-sectional study (social media data extraction)	TikTok	100 videos studied garnered 35,338,600 views	38 videos discouraged the vaccine; 3 videos claimed that the vaccine is a hoax.	High

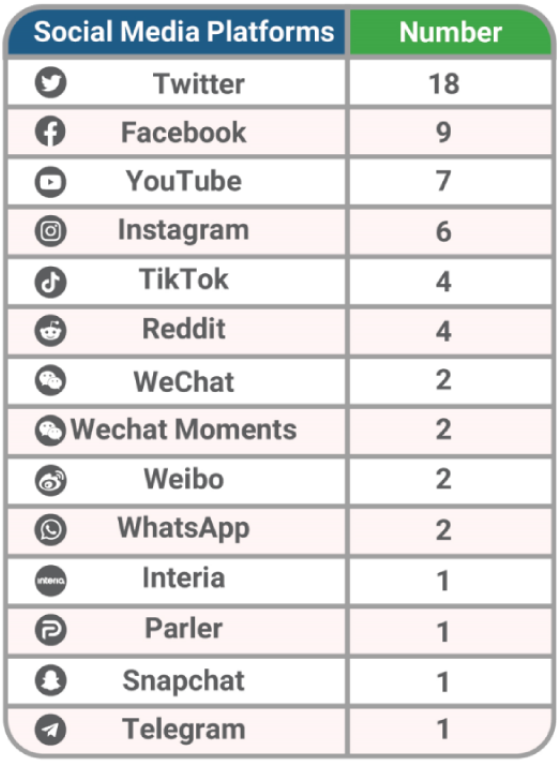
Study	Country	Study period	Study design	Type of social media	Social Media or population sample	Type of misinformation reported	Risk of bias (JBI[a])
Boucher et al [60], 2021	Canada	November 19 and November 26, 2020	Cross-sectional study (social media data extraction)	Twitter	636,516 English and French tweets	COVID-19 vaccines are poison and the messenger RNA technology has not been tested yet and is harmful.	High
Criss et al [61], 2021	United States	October 2020 to January 2021	Cross-sectional study (social media data extraction)	Twitter	1110 tweets	Misleading information that countered scientific research about the vaccines; the government using vaccines to insert microchips and control the population; the immune system is stronger than the vaccines; race extermination conspiracy that claims that the vaccine was created to "kill off [people of color] POC."	High
Herrera-Peco et al [62], 2021	Spain	December 8 to December 23, 2020	Cross-sectional study (social media data extraction)	Twitter	6080 Twitter interactions (n=499 of those are single tweets)	Deny the existence of the virus; the vaccine will modify the DNA of human beings; industry lobbies to kill older adults and leave young adults with Bells syndrome.	High
Melton et al [63], 2021	United States	December 1, 2020, to May 15, 2021	Cross-sectional study (social media data extraction)	Reddit	13 Reddit communities	Misinformation about side effects.	High
Pascual-Ferrá et al [64], 2021	United States	December 29, 2019, to January 2, 2021	Cross-sectional study (social media data extraction)	Facebook, Instagram, Reddit, and YouTube	Peaks and interactions	Viral video of a nurse fainting after vaccine uptake. Misinformation about COVID-19 vaccines (not specified).	High
Rotolo et al [65], 2021	United States	March 19, 2020, and June 16, 2021	Cross-sectional study (social media data extraction)	Facebook, Twitter, and Instagram	Aim: share 49 infographics to counter vaccine hesitancy.	COVID-19 myths.	High
Savolainen [66], 2021	Finland	February 2021	Cross-sectional study (social media data extraction)	Reddit, from the subreddit VaxxHappened	40 threads contained in total 1877 messages	Misinformation about COVID-19 vaccines (not specified).	High
Thelwall et al [67], 2021	United Kingdom	March 10 to December 5, 2020	Cross-sectional study (social media data extraction)	Twitter	446 COVID-19 vaccine–hesitant tweets in English	Deep state conspiracy; depopulation; microchips; Bill Gates; fearing that people of color are at risk for experimentation—motivated by the infamous US federal government Tuskegee Syphilis study ending in 1972 that secretly experimented on poor African American men.	High

Study	Country	Study period	Study design	Type of social media	Social Media or population sample	Type of misinformation reported	Risk of bias (JBI[a])
Wawrzuta et al [68], 2021	Poland	November 1, 2020, to May 1, 2021	Cross-sectional study (social media data extraction)	Facebook	3414 Facebook comments	The vaccine was created only for the profit of pharmaceutical companies; conspiracy theories, hidden vaccine effects (eg, chips); the vaccine will be dangerous to health; the vaccine has existed before the COVID-19 pandemic.	High
Doyno et al [69], 2021	Unites States	January to April 2021	Quasi-experimental study	YouTube, Twitter, Facebook, and Instagram	Information campaign with 79 COVID-19 vaccine–related videos in English, Cantonese, Spanish, Mandarin, and Polish	Misinformation (not specified).	N/A[b]

[a] JBI: Joanna Briggs Institute.

[b] N/A: not applicable.

Figure 2. Social media platforms.



Social Media Platforms	Number
Twitter	18
Facebook	9
YouTube	7
Instagram	6
TikTok	4
Reddit	4
WeChat	2
Wechat Moments	2
Weibo	2
WhatsApp	2
Interia	1
Parler	1
Snapchat	1
Telegram	1

Thematic Analysis

Figure 3 provides an overview of the 3 overarching themes of misinformation identified from the thematic analysis, and examples from the data extraction and codes that laid ground for the final 3 themes are seen in the inner circles of the figure.

Of the 45 included studies, 18 (40%) studies reported misinformation across all 3 categories [27-29,35-37,45,48,49,51,55,57, 58,60-62,67,68], 9 (20%) studies reported only on conspiracies [26,30,33,42,44,46,53,54,56], 6 (13%) studies were concerned specifically with medical misinformation [25,34,38,43,63,64], and 12 (27%) studies reported on COVID-19 vaccine misinformation or antivaccine discourse without going into further detail [31,35,39-41,47,52,59,65,66,69].

Figure 3. Types of misinformation about the COVID-19 vaccine on social media platforms.

Effects of Social Media Misinformation

We identified 19 studies that made assumptions regarding the effects of social media misinformation on vaccine hesitancy (Table 3). The evaluation of the certainty of evidence of these 19 studies that measured the “Association between social media misinformation and COVID-19 vaccine hesitancy” was classified as moderate or low to moderate according to GRADE in 2 cases [38,65]. For the rest of the studies, the certainty of evidence according to GRADE was considered low or very low.

The 2 studies with a higher certainty of evidence had an experimental design. Rotolo et al [65] aimed to develop and distribute infographics that addressed COVID-19 vaccine hesitancy and misinformation. Although their infographics reached thousands of people, they were unable to determine the impact on vaccine hesitancy. Sharevski and Gover [38] analyzed the perceived accuracy of COVID-19 vaccine–related tweets when they were moderated by smart device technology that Twitter applies to COVID-19 misinformation. The results from the 304 participants suggested that vaccine-hesitant users ignored warnings as long as the tweets aligned with their personal beliefs.

Table 3. Studies in which the effect of social media misinformation is measured or discussed (n=19).

Study	Reported effect of misinformation	Certainty of evidence (GRADE[a])
Sharevski and Gover [38], 2021	Amazon Alexa was not able to dispel any biases that were rooted in personal beliefs. One's hesitancy from COVID-19 vaccination sufficed for biased perception of the information from Alexa despite any labeling as long as the tweets echoed their skeptical outlook on the whole COVID-19 vaccination effort.	Moderate
Rotolo et al [65], 2021	Each infographic reached thousands to tens of thousands of people. We do not know whether those who viewed these infographics changed their perspective on vaccination, so we are unable to conclude their impact on vaccine hesitancy based on this study alone.	Low to moderate
Allington et al [26], 2021	Informational reliance on all social media platforms was positively correlated with vaccine hesitancy; this correlation was strongest concerning Facebook and YouTube (R_S[b]=0.15 and R_S=0.18, respectively). Coronavirus conspiracy suspicions and general vaccine attitudes appear uniquely predictive, jointly explaining 35% of variance.	Low
Bhagianadh and Arora [28], 2021	Those depending on social media as the main source of information on COVID-19 expressed higher negative vaccine intent (OR[c] 3.36, 95% CI 1.44-7.82). Among those who expressed a negative vaccine intent, 40% (n=298) expressed no trust in government, and 10% (n=74) said that the vaccines cause COVID-19.	Low
Boucher et al [60], 2021	The study showed 2 clusters opposite to these vaccine acceptant clusters exhibiting more vaccine-hesitant narratives. There were 23.4% (n=146,191) of conversations on Twitter during this period of observation that can be directly attributed to vaccine hesitancy.	Low
Chadwick et al [30], 2021	Combinations of news avoidance and high levels of the news-finds-me attitude and social media dependence and high levels of conspiracy mentality are most likely to be associated with web-based discouragement of vaccination.	Low
Jennings et al [42], 2021	Holding conspiracy beliefs is a significant predictor of vaccine hesitancy. In the bivariate analysis, there is some support for a relationship between social media use (Snapchat, TikTok, YouTube, and Instagram) and increased vaccine hesitancy. YouTube users were significantly less willing to be vaccinated, with a two-thirds likelihood of vaccine willingness compared with nonusers.	Low
Liu and Liu [51], 2021	279 tweets stated their behavioral intentions. A total of 97 tweets were labeled with positive behavioral intentions, while 182 tweets contained negative behavioral intentions.	Low
Park et al [34], 2021	The study found that social media dependence and high levels of conspiracy mentality were most likely to be associated with web-based discouragement of vaccination. The likelihood of COVID-19 vaccine uptake was significantly lower among those relying on social media (OR 0.40, 95% CI 0.25-0.65)	Low
Zhang et al [39], 2021	Regarding social media influence, higher frequency of exposure to positive information related to COVID-19 vaccination was associated with a higher intention to receive a COVID-19 vaccination at market rate (AOR[d] 1.53, 95% CI 1.39-1.70) or a free vaccination (AOR 1.52, 95% CI 1.35-1.71).	Low
Zhang et al [40], 2021	Higher exposure to positive information related to COVID-19 vaccination on social media was associated with higher parental acceptability of COVID-19 vaccination (AOR 1.35, 95% CI 1.17-1.56). Higher exposure to negative information related to COVID-19 vaccination was negatively associated with the dependent variable (AOR 0.85, 95% CI 0.74-0.99).	Low
Aloweidi et al [27], 2021	The effect of social media (OR 1.21, 95% CI 1.04-1.41; *P*=.01) was significantly associated with the willingness to take COVID-19 vaccine. Circulated information about COVID-19 vaccines on social media platforms that they believed in: it is unsafe (n=283, 43.8%); effect of the vaccines on a genetic level (n=87, 13.5%); causes chronic illnesses (n=60, 9.3%); may lead to infertility (n=43, 6.7%); can affect their offspring (n=56, 8.7%); toxic heavy metals and neurotoxic materials (n=47, 7.3%); it is a part of a secret research (n=101, 15.6%)	Very low to low
Brodziak et al [29], 2021	A total of 432 (68%) used social media every day. Unwilling to vaccinate against COVID-19: social media as a source of information about vaccinations (OR 1.42, 95% CI 0.72-2.80). Not a significant predictor; attitudes toward COVID-19 vaccines: afraid of the vaccine's side effects (n=284, 44.7%); afraid of the composition of the vaccine (n=239, 37.6%); contains bodies of aborted children (n=49, 7.7%); COVID-19 does not exist (n=42, 6.6%)	Very low to low
Ebrahimi et al [31], 2021	Individuals with a preference for social media platforms as compared with those preferring source-verified media platforms had a near 2-fold (ie, 1.64) odds of being hesitant toward vaccination. Belief in superiority of natural immunity: OR 2.663, 95% CI 2.350-3.028; *P*<.001	Very low to low
El-Far Cardo et al [43], 2020	Factors that were negatively associated to get vaccinated were using social media in general as an information source about COVID-19 (*P*=.01) and the use of Facebook (*P*=.05) or Telegram (*P*=.05). However, using Twitter was not significantly associated with adverse vaccination intentions (*P*=.56). Believing that COVID-19 is not dangerous was associated with unwillingness to get vaccinated.	Very low to low

Study	Reported effect of misinformation	Certainty of evidence (GRADE[a])
Petravić et al [35], 2021	Those who trusted alternative media sources (alternative explanations on social media) and had a distrust of the government were more vaccine hesitant.	Very low to low
Sallam et al [36], 2021	The lowest rate of intention to get the vaccine was among those who depended on social media platforms (19.8%) compared with dependence on medical doctors, scientists, and scientific journals (47.2%, $P<.001$). Conspiracy beliefs were evaluated using the validated VCBS[e], with higher scores implying embrace of conspiracies. A significantly higher VCBS score was correlated with reluctance to get the vaccine ($P<.001$).	Very low to low
Costantino et al [41], 2014	A total of 71.4% (n=60) responded that unfavorable information about COVID-19 vaccines obtained from the internet, social media, or media was associated with the decision to not take the vaccine.	Very low
Karabela et al [33], 2021	Although the correlation was not significant, of the participants, those who considered having vaccination mostly trusted YouTube as their source of information. In contrast, the participants who stated that they would have the COVID-19 vaccine did not trust social media sites such as Facebook, Twitter, and Instagram ($P<.005$). There was a positive and low-level relationship between attitudes toward COVID-19 vaccines and conspiracy theories ($r=0.214$).	Very low

[a] GRADE: Grading of Recommendations Assessment, Development and Evaluation.

[b] R_S: Spearman Rank Correlation Coefficient.

[c] OR: odds ratio.

[d] AOR: adjusted odds ratio.

[e] VCBS: Vaccine Conspiracy Belief Scale.

Allington et al [26] analyzed findings from a web-based survey conducted with a sample of 4343 adults in the United Kingdom. They found a positive correlation between trust in social media and vaccine hesitancy and the strongest link was found for YouTube and Facebook. Conspiracy suspicions about COVID-19 and general vaccine attitudes appeared to be uniquely predictive, jointly explaining 35% of the variance. Boucher et al [60] analyzed 636,516 English and French tweets. A total of 23.4% (n=146,191) of the conversations on Twitter during the study period could be directly attributed to vaccine hesitancy. A British study by Liu and Liu [30] of 5114 adults found that social media dependence and high levels of conspiracy mentality were most likely to be associated with web-based discouragement of vaccination. In a study of 4571 Norwegian adults, individuals who preferred social media platforms had nearly 2-fold (ie, 1.64) odds of being hesitant toward COVID-19 vaccination compared with those preferring source-verified media platforms [31]. In addition, those who held the belief of the superiority of natural immunity over vaccination were more vaccine hesitant (odds ratio 2.663, 95% CI 2.350-3.028; $P<.001$). Petravić et al [35] asked 12,042 Slovenian residents about their attitudes toward COVID-19 vaccines. Those who trusted alternative media sources and alternative explanations on social media were more vaccine hesitant. A total of 11 studies [27-29,33,34,36,39-43] discussed social media misinformation, vaccine uptake, and vaccine intentions.

Discussion

Principal Findings

The 45 included studies about misinformation on social media platforms about COVID-19 vaccines suggest that there should be great concern about the volume of misinformation being spread, and the association between COVID-19 vaccine misinformation and vaccine hesitancy. To our knowledge, this is the first review to analyze social media misinformation about COVID-19 vaccines. We identified 3 overall categories of misinformation, namely, medical misinformation, conspiracies, and distrust in vaccine development; however, the 3 categories are connected and sometimes overlapping, as distrust in vaccine development might be founded in conspiratorial beliefs about hidden power structures and corrupt elites. The included studies were predominantly from Europe and the United States, and therefore, there is a lack of information, especially from African and South American countries. Twitter was the most studied platform, with Facebook and YouTube being in the second and third place, respectively.

Fear of side effects is a major concern when it comes to vaccine hesitancy, and as this review shows, this concern can easily turn into medical misinformation and exaggerations of side effects. To synthesize what is known about social media misinformation about COVID-19 vaccines from the included studies, a thematic analysis was undertaken. The coded extract of data that made up the theme medical misinformation contained misinformation about side effects such as infertility, chronic illness, changes in DNA, physical deformities, and mental illness. Only one study mentioned autism as an adverse side effect of COVID-19 vaccines [63]. Knowing that the side effects of the vaccines are a major concern [5], medical misinformation has the potential to do a lot of harm.

When we examined the types of reported misinformation, we also found that a lot of misinformation is grounded in conspiracy theories. Some of these conspiracy theories have become infamous, such as the belief that there are secret societies and hidden power structures run by corrupt elites. These elites are believed to be networking with big pharmaceutical companies to make money or to depopulate the world. There are also conspiracy theories about racially motivated depopulation. For example, we found 3 studies from the United States that mentioned the fear of racist motives by official health authorities as a reason for vaccine hesitancy [45,61,67]. Some of this fear has historical roots in the United States, as one of these studies [67], for instance, brought up the Tuskegee Syphilis study. This was a clinical study (1932-1972) in which the United States Public Health Service used African Americans to observe untreated syphilis and therefore denied them treatment [71]. This exemplifies that a lack of trust in public health institutions might have deep historical roots in some countries and cultures. Other issues to be aware of are religious concerns and vaccine hesitancy. We found several studies that reported on misinformation about the content and development of vaccines and in some studies [29,45,55], we found very explicit language (eg, “pigs” and “cells from aborted children”). Such wording can cause worry in some religious communities.

The second objective of this review was to examine the effects of social media misinformation about COVID-19 vaccines. The 19 studies identified in Table 3 interpreted the results as associations among social media use, misinformation, and vaccine hesitancy. According to the JBI and GRADE evaluations, there is a need for more robust designs to become more certain regarding the actual effect of social media misinformation on vaccine hesitancy. Only 1 study, an intervention study regarding the impact of addressing misinformation on Twitter users, was assessed to have a low risk of bias and moderate quality of evidence [38]. In addition, 4 studies reported significance levels of associations, but the effect size was not reported [31,32,35,50]. Other studies in this review showed that social media platforms did not necessarily spread misinformation to a great extent, perhaps reflecting that the effort made by some social media platforms to halt misinformation has worked. Chan et al [47] examined 48 COVID-19 vaccine–related videos on YouTube in December 2020 and found only 2 videos (4.2%) that made nonfactual claims. Hernández-García et al [54] also examined YouTube videos during February 2021 and found that only 2 out of 110 videos contained COVID-19 vaccine hoaxes or conspiracy theories. Pascual-Ferrá et al [64] examined social media data from Facebook, Instagram, Reddit, and YouTube and did not find evidence of the dominance of misinformation. However, what is being spread and discussed in closed groups is another question that needs to be examined further. Another valid approach would also be to examine comment sections. Although antivaccine content has been prevented from surfacing in searches, this does not prevent people from commenting about their beliefs or posting other types of information in the comment section. It is controversial to deplatform people [72] and might even do harm, as these people might be seen as someone speaking against the establishment, which are, in essence, some of the core beliefs of some conspiracy theorists.

Surprisingly, there was a dearth of studies examining misinformation about autism and COVID-19 vaccines. Considering the history of misinformation about vaccines and autism over the past 2 decades, more research should focus on this topic. One could also speculate whether this would have played out differently if COVID-19 vaccines were more targeted toward younger children. Future research should also aim to examine social media platforms such as TikTok, which is a very popular platform worldwide, and is often used by people who are younger than, for instance, the average Twitter user [73]. The low inclusion of some social media platforms such as TikTok or Telegram is a limitation, as certain parts of the population and particular communities are not included.

When addressing vaccine hesitancy, one should be careful before labeling all vaccine-hesitant people as antivaxxers or misinformed people. The primary concerns from people who say that they are vaccine hesitant are the safety of the vaccines and the rapid pace of their development [7]. However, being hesitant and skeptical does not mean that these people are unwilling to take the vaccines but rather that they have some concerns that should be adequately addressed to convince them of the safety and efficacy of the vaccines. To understand a complex issue such as vaccine hesitancy, knowledge about sociodemographic conditions and cultural awareness is key. In addition, countries with a more undemocratic regime will suffer from a lack of trust in official authorities, which may damage an official vaccine campaign. People’s trust in the government varies between countries and cultures. Although some countries have a tradition for mandatory vaccination, this is less acceptable in other countries.

The issue of trust is also an important issue to be considered. “Fake news” became a buzzword in the last decade and the term was used not only to actually coin false news but also to spread distrust to news agencies and official actors, accusing them of spreading falsehoods. Vosoughi et al [74] aimed to understand how false news spread and examined a set of rumors (n=126,000) spread by 3 million people on Twitter from 2006 to 2017. The results showed that false news spread much faster and reached a larger audience than real news. Social media has contributed to a far more complex information landscape than before and has created new challenges when it comes to building trust in official actors. These are issues that need to be addressed and analyzed in future studies of misinformation about vaccines.

Limitations

We did not include gray literature or preprints in this review. The rapid pace at which the pandemic is moving makes preprint research particularly relevant. However, although peer review is not a guarantee of quality, we decided not to include gray literature or preprints and limited eligible articles to peer-reviewed manuscripts. We did not contact researchers with potential projects on this subject matter. Furthermore, the searched databases were selected based on the topic at hand. There will always be a chance that other, more specified or general databases would capture other studies.

A limitation of the evidence included in this review is that, in our assessment, all but one received a low score on the assessment of quality of evidence. However, it is a challenge in the process of assessing quality of evidence and risk of bias, that the included studies have a range of different designs, each with its strengths and weaknesses. We applied 2 tools in this regard: 1 from the JBI [24] and 1 from the GRADE Working Group [70]. Neither tool provides a complete picture of the included studies, but they may help the reader in obtaining a broader view of the included studies.

Furthermore, there is a poor correlation between self-reported social media use and actual use [75]. A high proportion of these studies extracted data from Twitter because Twitter has opened up access for researchers to extract data from its platform, making it more accessible compared with other social media platforms. The Twitter sample may not be representative of a random sample of the population, as its users tend to range in age from 25 to 34 years and are predominantly from the United States [76,77]. Furthermore, we did not assess the potential presence of social media bots (automated accounts) spreading incorrect information in these studies. We also did not discuss how social media algorithms partake in creating echo chambers [78]. These are well-known challenges in researching data gathered from social media [79]. Only 1 study included in this review was from an African country [32]. The study was from Uganda and included 600 participants. There were no studies from Middle or South American countries or Pacific Island countries and Australia. The studies included in this review focused mostly on high-income countries, thus making the conclusions and generalizations weaker in terms of applying them to Global South nations.

Conclusions

This review suggests that there should be great concern about the volume of misinformation being spread and the association between COVID-19 vaccine misinformation and vaccine hesitancy. Many studies have shown that there is a link between misinformation on social media and COVID-19 vaccine hesitancy. However, there is a need to examine this effect using a more robust experimental design to assess this effect. It is possible to conduct more experimental studies in an ethical manner in a laboratory setting; for instance, a study to see whether people are able to distinguish between false and true information and how they do so. Such a study would, of course, have to be based on informed consent and be approved by an ethics committee. It is also possible to improve observational studies that extract data from social media by gathering more representative data (eg, including data from several social media platforms, different audiences, several languages, and covering longer periods). There are many types of misinformation that are spread on social media platforms, and to prevent these myths from taking hold, health authorities should openly address and discuss these false claims with both cultural and religious awareness in mind. This review showed that a greater variation in studies is needed when it comes to both social media platforms and geographic location. We only found one study that mentioned misinformation about autism and COVID-19 vaccines, but taking the history of autism and the antivaxx community into account, we believe that this an issue that should be given attention in future research.

Although some major tech companies have taken steps to prevent misinformation, more action is needed to stop this infodemic. One valid approach proposed for infodemic management is first information monitoring (infoveillance); second, to enhance and build eHealth literacy and science literacy capacity; third, to encourage quality improvement processes such as fact-checking and peer review; and finally, to encourage accurate and timely knowledge translation [80].

Misinformation about COVID-19 vaccines is still thriving on social media platforms. However, this undertaking represents a balance between people’s right to speak their minds and strategies to counter the spread of misinformation.

Acknowledgments

The authors would like to thank senior librarian Kjell Erik Johnsen for his help and input in conducting the systematic searches.

Data Availability

All relevant data are available in this paper and its supplementary files. Correspondence and requests for additional material should be addressed to the corresponding author (IS).

Authors' Contributions

IS, EG, and ANH conceived the idea for this study. IS and ANH screened the titles and abstracts. IS and EG full-text screened the articles. IS performed data extraction with verification from EG. DSQ assessed the risk of bias, and RW evaluated the quality of evidence. IS took the lead in writing the manuscript. All authors discussed, revised, and contributed to the final manuscript. All authors have read and agreed to the final published version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Documentation of systematic literature search.
[PDF File (Adobe PDF File), 100 KB - jmir_v24i8e37367_app1.pdf]

Multimedia Appendix 2
Thematic analysis.
[PDF File (Adobe PDF File), 108 KB - jmir_v24i8e37367_app2.pdf]

Multimedia Appendix 3
Assessment of risk of bias.
[PDF File (Adobe PDF File), 215 KB - jmir_v24i8e37367_app3.pdf]

Multimedia Appendix 4
Grading of Recommendations, Development, and Evaluation (GRADE) scores from the GRADE handbook for quality of evidence.
[PDF File (Adobe PDF File), 166 KB - jmir_v24i8e37367_app4.pdf]

Multimedia Appendix 5
List of excluded full-text articles.
[PDF File (Adobe PDF File), 164 KB - jmir_v24i8e37367_app5.pdf]

References

1. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2021-10-24]
2. Coronavirus (COVID-19) Vaccinations. Our World In Data. URL: https://ourworldindata.org/covid-vaccinations [accessed 2022-02-18]
3. Tracking Coronavirus around the world. The New York Times. URL: https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html [accessed 2021-12-03]
4. Update on Omicron. World Health Organization. URL: https://www.who.int/news/item/28-11-2021-update-on-omicron [accessed 2021-12-03]
5. Machingaidze S, Wiysonge CS. Understanding COVID-19 vaccine hesitancy. Nat Med 2021 Aug;27(8):1338-1339. [doi: 10.1038/s41591-021-01459-7] [Medline: 34272500]
6. Gabarron E, Oyeyemi SO, Wynn R. COVID-19-related misinformation on social media: a systematic review. Bull World Health Organ 2021 Jun 01;99(6):455-63A [FREE Full text] [doi: 10.2471/BLT.20.276782] [Medline: 34108756]
7. Infodemic. World Health Organization. URL: https://www.who.int/health-topics/infodemic#tab=tab_1 [accessed 2021-10-21]
8. When the Buzz Bites Back. Washington Post. URL: https://www.washingtonpost.com/archive/opinions/2003/05/11/when-the-buzz-bites-back/bc8cd84f-cab6-4648-bf58-0277261af6cd/ [accessed 2022-04-07]
9. Lord C, Brugha TS, Charman T, Cusack J, Dumas G, Frazier T, et al. Autism spectrum disorder. Nat Rev Dis Primers 2020 Jan 16;6(1):5 [FREE Full text] [doi: 10.1038/s41572-019-0138-4] [Medline: 31949163]
10. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger J. Vaccine hesitancy: an overview. Hum Vaccin Immunother 2013 Aug;9(8):1763-1773 [FREE Full text] [doi: 10.4161/hv.24657] [Medline: 23584253]
11. Motta M, Stecula D. Quantifying the effect of Wakefield et al. (1998) on skepticism about MMR vaccine safety in the U.S. PLoS One 2021;16(8):e0256395 [FREE Full text] [doi: 10.1371/journal.pone.0256395] [Medline: 34411172]
12. Benoit SL, Mauldin RF. The "anti-vax" movement: a quantitative report on vaccine beliefs and knowledge across social media. BMC Public Health 2021 Nov 17;21(1):2106 [FREE Full text] [doi: 10.1186/s12889-021-12114-8] [Medline: 34789206]
13. Pullan S, Dey M. Vaccine hesitancy and anti-vaccination in the time of COVID-19: a Google Trends analysis. Vaccine 2021 Apr 01;39(14):1877-1881 [FREE Full text] [doi: 10.1016/j.vaccine.2021.03.019] [Medline: 33715904]
14. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? BMJ 2014 Oct 14;349:g6178. [doi: 10.1136/bmj.g6178] [Medline: 25315514]
15. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
16. Immunizing the public against misinformation. World Health Organization. URL: https://www.who.int/news-room/feature-stories/detail/immunizing-the-public-against-misinformation [accessed 2021-10-21]
17. Cochrane Rapid Reviews. Interim Guidance from the Cochrane Rapid Reviews Methods Group. Cochrane Rapid Reviews Methods Group. URL: http://methods.cochrane.org/sites/methods.cochrane.org.rapidreviews/files/uploads/cochrane_rr_-_guidance-23mar2020-final.pdf [accessed 2022-07-26]
18. 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]
19. 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]
20. Pfizer and BioNTech conclude phase 3 study of COVID-19 vaccine candidate, meeting all primary efficacy endpoints. Pfizer. URL: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-conclude-phase-3-study-covid-19-vaccine [accessed 2022-01-18]

21. 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]
22. 11 Vaccines Granted Emergency Use Listing (EUL) by WHO. COVID 19 Vaccine Tracker. URL: https://covid19.trackvaccines.org/agency/who/ [accessed 2022-04-25]
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
24. JBI critical appraisal checklist for analytical cross sectional studies. Joanna Briggs Institute. URL: https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Analytical_Cross_Sectional_Studies2017_0.pdf [accessed 2022-01-04]
25. Alibrahim J, Awad A. COVID-19 vaccine hesitancy among the public in Kuwait: a cross-sectional survey. Int J Environ Res Public Health 2021 Aug 22;18(16):8836 [FREE Full text] [doi: 10.3390/ijerph18168836] [Medline: 34444585]
26. Allington D, McAndrew S, Moxham-Hall V, Duffy B. Coronavirus conspiracy suspicions, general vaccine attitudes, trust and coronavirus information source as predictors of vaccine hesitancy among UK residents during the COVID-19 pandemic. Psychol Med 2021 Apr 12:1-12 [FREE Full text] [doi: 10.1017/S0033291721001434] [Medline: 33843509]
27. Aloweidi A, Bsisu I, Suleiman A, Abu-Halaweh S, Almustafa M, Aqel M, et al. Hesitancy towards COVID-19 Vaccines: an analytical cross-sectional study. Int J Environ Res Public Health 2021 May 12;18(10):5111 [FREE Full text] [doi: 10.3390/ijerph18105111] [Medline: 34065888]
28. Bhagianadh D, Arora K. COVID-19 vaccine hesitancy among community-dwelling older adults: the role of information sources. J Appl Gerontol 2022 Jan;41(1):4-11. [doi: 10.1177/07334648211037507] [Medline: 34365856]
29. Brodziak A, Sigorski D, Osmola M, Wilk M, Gawlik-Urban A, Kiszka J, et al. Attitudes of patients with cancer towards vaccinations-results of online survey with special focus on the vaccination against COVID-19. Vaccines (Basel) 2021 Apr 21;9(5):411 [FREE Full text] [doi: 10.3390/vaccines9050411] [Medline: 33919048]
30. Chadwick A, Kaiser J, Vaccari C, Freeman D, Lambe S, Loe BS, et al. Online social endorsement and COVID-19 vaccine hesitancy in the united kingdom. Social Media Soc 2021 Apr 05;7(2):205630512110088. [doi: 10.1177/20563051211008817]
31. Ebrahimi OV, Johnson MS, Ebling S, Amundsen OM, Halsøy Ø, Hoffart A, et al. Risk, trust, and flawed assumptions: vaccine hesitancy during the COVID-19 pandemic. Front Public Health 2021;9:700213 [FREE Full text] [doi: 10.3389/fpubh.2021.700213] [Medline: 34277557]
32. Kanyike AM, Olum R, Kajjimu J, Ojilong D, Akech GM, Nassozi DR, et al. Acceptance of the coronavirus disease-2019 vaccine among medical students in Uganda. Trop Med Health 2021 May 13;49(1):37 [FREE Full text] [doi: 10.1186/s41182-021-00331-1] [Medline: 33985592]
33. Karabela SN, Coşkun F, Hoşgör H. Investigation of the relationships between perceived causes of COVID-19, attitudes towards vaccine and level of trust in information sources from the perspective of Infodemic: the case of Turkey. BMC Public Health 2021 Jun 23;21(1):1195 [FREE Full text] [doi: 10.1186/s12889-021-11262-1] [Medline: 34158015]
34. Park S, Massey PM, Stimpson JP. Primary source of information about COVID-19 as a determinant of perception of COVID-19 severity and vaccine uptake : source of information and COVID-19. J Gen Intern Med 2021 Oct;36(10):3088-3095 [FREE Full text] [doi: 10.1007/s11606-021-07080-1] [Medline: 34378115]
35. Petravić L, Arh R, Gabrovec T, Jazbec L, Rupčić N, Starešinič N, et al. Factors affecting attitudes towards COVID-19 vaccination: an online survey in Slovenia. Vaccines (Basel) 2021 Mar 12;9(3):247 [FREE Full text] [doi: 10.3390/vaccines9030247] [Medline: 33808958]
36. Sallam M, Dababseh D, Eid H, Hasan H, Taim D, Al-Mahzoum K, et al. Low COVID-19 vaccine acceptance is correlated with conspiracy beliefs among university students in Jordan. Int J Environ Res Public Health 2021 Mar 01;18(5):2407 [FREE Full text] [doi: 10.3390/ijerph18052407] [Medline: 33804558]
37. Sallam M, Dababseh D, Eid H, Al-Mahzoum K, Al-Haidar A, Taim D, et al. High rates of COVID-19 vaccine hesitancy and its association with conspiracy beliefs: a study in Jordan and Kuwait among other Arab countries. Vaccines (Basel) 2021 Jan 12;9(1):42 [FREE Full text] [doi: 10.3390/vaccines9010042] [Medline: 33445581]
38. Sharevski F, Gover D. Two truths and a lie: exploring soft moderation of COVID-19 misinformation with Amazon Alexa. In: Proceedings of the 16th International Conference on Availability, Reliability and Security. 2021 Presented at: ARES 2021: The 16th International Conference on Availability, Reliability and Security; Aug 17 - 20, 2021; Vienna Austria URL: https://doi.org/10.1145/3465481.3470017 [doi: 10.1145/3465481.3470017]
39. Zhang KC, Fang Y, Cao H, Chen H, Hu T, Chen Y, et al. Behavioral intention to receive a COVID-19 vaccination among Chinese factory workers: cross-sectional online survey. J Med Internet Res 2021 Mar 09;23(3):e24673 [FREE Full text] [doi: 10.2196/24673] [Medline: 33646966]
40. Zhang KC, Fang Y, Cao H, Chen H, Hu T, Chen YQ, et al. Parental acceptability of COVID-19 vaccination for children under the age of 18 years: cross-sectional online survey. JMIR Pediatr Parent 2020 Dec 30;3(2):e24827 [FREE Full text] [doi: 10.2196/24827] [Medline: 33326406]
41. Costantino C, Gori D, Alba D, Cimino L, Conforto A, Tomasello F, et al. Willingness to receive covid-19 vaccination in costumers accessing community pharmacies in the province of Palermo, Italy. EuroMediterranean Biomedical J 2021;16(19):80-84 [FREE Full text]

42. 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]
43. El-Far Cardo A, Kraus T, Kaifie A. Factors that shape people's attitudes towards the COVID-19 pandemic in Germany-the influence of MEDIA, politics and personal characteristics. Int J Environ Res Public Health 2021 Jul 22;18(15):7772 [FREE Full text] [doi: 10.3390/ijerph18157772] [Medline: 34360063]
44. Knights F, Carter J, Deal A, Crawshaw AF, Hayward SE, Jones L, et al. Impact of COVID-19 on migrants' access to primary care and implications for vaccine roll-out: a national qualitative study. Br J Gen Pract 2021 Aug;71(709):e583-e595 [FREE Full text] [doi: 10.3399/BJGP.2021.0028] [Medline: 33875420]
45. Berry SD, Johnson KS, Myles L, Herndon L, Montoya A, Fashaw S, et al. Lessons learned from frontline skilled nursing facility staff regarding COVID-19 vaccine hesitancy. J Am Geriatr Soc 2021 May;69(5):1140-1146 [FREE Full text] [doi: 10.1111/jgs.17136] [Medline: 33764497]
46. Choudhary K, Garapati P, Prasad SR, Murti K, Singh S, Kumar N. COVID-19 vaccine hesitancy assessment among the educated Indian population: an online survey. Int j pharmaceutical res 2021 Aug 02;13(03). [doi: 10.31838/ijpr/2021.13.03.196]
47. Chan C, Sounderajah V, Daniels E, Acharya A, Clarke J, Yalamanchili S, et al. The reliability and quality of YouTube videos as a source of public health information regarding COVID-19 vaccination: cross-sectional study. JMIR Public Health Surveill 2021 Jul 08;7(7):e29942 [FREE Full text] [doi: 10.2196/29942] [Medline: 34081599]
48. Herrera-Peco I, Ruiz Núñez C, Jiménez-Gómez B, Romero-Magdalena CS, Benítez De Gracia E. [COVID-19 and vaccination: analysis of public institution's role in information spread through Twitter.]. Rev Esp Salud Publica 2021 Jun 16;95:e202106084 [FREE Full text] [Medline: 34132241]
49. Hughes B, Miller-Idriss C, Piltch-Loeb R, Goldberg B, White K, Criezis M, et al. Development of a codebook of online anti-vaccination rhetoric to manage COVID-19 vaccine misinformation. Int J Environ Res Public Health 2021 Jul 15;18(14):7556 [FREE Full text] [doi: 10.3390/ijerph18147556] [Medline: 34300005]
50. Larrondo-Ureta A, Fernández S, Morales-i-Gras J. Desinformación, vacunas y Covid-19. Análisis de la infodemia y la conversación digital en Twitter. Revista Latina de Comunicacion Social 2021 Jun 07(79):1-18. [doi: 10.4185/rlcs-2021-1504]
51. Liu S, Liu J. Understanding behavioral intentions toward COVID-19 vaccines: theory-based content analysis of tweets. J Med Internet Res 2021 May 12;23(5):e28118 [FREE Full text] [doi: 10.2196/28118] [Medline: 33939625]
52. Sobkowicz P, Sobkowicz A. Agent based model of anti-vaccination movements: simulations and comparison with empirical data. Vaccines (Basel) 2021 Jul 21;9(8):809 [FREE Full text] [doi: 10.3390/vaccines9080809] [Medline: 34451934]
53. Guntuku SC, Buttenheim AM, Sherman G, Merchant RM. Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States. Vaccine 2021 Jul 05;39(30):4034-4038 [FREE Full text] [doi: 10.1016/j.vaccine.2021.06.014] [Medline: 34140171]
54. Hernández-García I, Gascón-Giménez I, Gascón-Giménez A, Giménez-Júlvez T. Information in Spanish on YouTube about Covid-19 vaccines. Hum Vaccin Immunother 2021 Nov 02;17(11):3916-3921. [doi: 10.1080/21645515.2021.1957416] [Medline: 34375570]
55. Islam MS, Kamal AM, Kabir A, Southern DL, Khan SH, Hasan SM, et al. COVID-19 vaccine rumors and conspiracy theories: the need for cognitive inoculation against misinformation to improve vaccine adherence. PLoS One 2021;16(5):e0251605 [FREE Full text] [doi: 10.1371/journal.pone.0251605] [Medline: 33979412]
56. Kwok SW, 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]
57. Alliheibi FM, Omar A, Al-Horais N. Opinion mining of Saudi responses to COVID-19 vaccines on Twitter. Int J Adv Comput Sci Appl 2021;12(6). [doi: 10.14569/ijacsa.2021.0120610]
58. Baines A, Ittefaq M, Abwao M. #Scamdemic, #Plandemic, or #Scaredemic: what Parler social media platform tells us about COVID-19 vaccine. Vaccines (Basel) 2021 Apr 22;9(5):421 [FREE Full text] [doi: 10.3390/vaccines9050421] [Medline: 33922343]
59. Basch CH, Meleo-Erwin Z, Fera J, Jaime C, Basch CE. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. Hum Vaccin Immunother 2021 Aug 03;17(8):2373-2377 [FREE Full text] [doi: 10.1080/21645515.2021.1894896] [Medline: 33764283]
60. Boucher J, Cornelson K, Benham JL, Fullerton MM, Tang T, Constantinescu C, et al. Analyzing social media to explore the attitudes and behaviors following the announcement of successful COVID-19 vaccine trials: infodemiology study. JMIR Infodemiology 2021;1(1):e28800 [FREE Full text] [doi: 10.2196/28800] [Medline: 34447924]
61. Criss S, Nguyen TT, Norton S, Virani I, Titherington E, Tillmanns EL, et al. Advocacy, hesitancy, and equity: exploring U.S. Race-related discussions of the COVID-19 vaccine on Twitter. Int J Environ Res Public Health 2021 May 26;18(11):5693 [FREE Full text] [doi: 10.3390/ijerph18115693] [Medline: 34073291]
62. Herrera-Peco I, Jiménez-Gómez B, Romero Magdalena CS, Deudero JJ, García-Puente M, Benítez De Gracia E, et al. Antivaccine movement and COVID-19 negationism: a content analysis of Spanish-written messages on Twitter. Vaccines (Basel) 2021 Jun 15;9(6):656 [FREE Full text] [doi: 10.3390/vaccines9060656] [Medline: 34203946]

63. 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]
64. Pascual-Ferrá P, Alperstein N, Barnett DJ. A multi-platform approach to monitoring negative dominance for COVID-19 vaccine-related information online. Disaster Med Public Health Prep 2021 May 03:1-24 [FREE Full text] [doi: 10.1017/dmp.2021.136] [Medline: 33938423]
65. 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]
66. Savolainen R. Assessing the credibility of COVID-19 vaccine mis/disinformation in online discussion. J Inform Sci 2021 Aug 19:016555152110406 [FREE Full text] [doi: 10.1177/01655515211040653]
67. Thelwall M, Kousha K, Thelwall S. Covid-19 vaccine hesitancy on English-language Twitter. El profesional de la información 2021 Mar 15:e300212 [FREE Full text] [doi: 10.3145/epi.2021.mar.12]
68. Wawrzuta D, Jaworski M, Gotlib J, Panczyk M. What arguments against COVID-19 vaccines run on Facebook in Poland: content analysis of comments. Vaccines (Basel) 2021 May 10;9(5):481 [FREE Full text] [doi: 10.3390/vaccines9050481] [Medline: 34068500]
69. Doyno CR, Fitzgerald JM, White CM, Sobieraj DM, Zacchera M. Innovative partnership in Connecticut to expand health professional eligibility to administer COVID-19 vaccines. J Am Pharm Assoc (2003) 2022;62(1):247-252 [FREE Full text] [doi: 10.1016/j.japh.2021.08.013] [Medline: 34462216]
70. GRADE Handbook. GRADE. URL: https://gdt.gradepro.org/app/handbook/handbook.html [accessed 2021-11-05]
71. The U.S. Public Health Service Syphilis Study at Tuskegee. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tuskegee/timeline.htm [accessed 2022-05-26]
72. Burki T. The online anti-vaccine movement in the age of COVID-19. Lancet Digital Health 2020 Oct;2(10):e504-e505 [FREE Full text] [doi: 10.1016/S2589-7500(20)30227-2] [Medline: 32984795]
73. Most popular social networks worldwide as of January 2022, ranked by number of monthly active users. Statista. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2022-01-03]
74. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Sci 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
75. Parry DA, Davidson BI, Sewall CJ, Fisher JT, Mieczkowski H, Quintana DS. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. Nat Hum Behav 2021 Nov;5(11):1535-1547. [doi: 10.1038/s41562-021-01117-5] [Medline: 34002052]
76. Distribution of Twitter users worldwide as of April 2021, by age group. Statista. URL: https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/ [accessed 2022-01-03]
77. Leading countries based on number of Twitter users as of January 2022. Statista. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 2022-01-04]
78. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2023301118 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
79. Orabi M, Mouheb D, Al Aghbari Z, Kamel I. Detection of bots in social media: a systematic review. Inform Process Manag 2020 Jul;57(4):102250 [FREE Full text] [doi: 10.1016/j.ipm.2020.102250]
80. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]

Abbreviations

GRADE: Grading of Recommendations Assessment, Development and Evaluation
JBI: Joanna Briggs Institute
MMR: measles, mumps, and rubella
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

Edited by M Gisondi, L Westafer; submitted 18.02.22; peer-reviewed by SF Tsao, T Vukušić Rukavina, M MacKay, S Kwok; comments to author 04.04.22; revised version received 25.04.22; accepted 24.05.22; published 04.08.22.

Please cite as:
Skafle I, Nordahl-Hansen A, Quintana DS, Wynn R, Gabarron E
Misinformation About COVID-19 Vaccines on Social Media: Rapid Review
J Med Internet Res 2022;24(8):e37367
URL: https://www.jmir.org/2022/8/e37367
doi:10.2196/37367
PMID:35816685

Original Paper

Assessing the Role of Social Bots During the COVID-19 Pandemic: Infodemic, Disagreement, and Criticism

Victor Suarez-Lledo[1,2], BSc, MSc; Javier Alvarez-Galvez[1,2], BSc, MSc, PhD

[1]Department of Biomedicine, Biotechnology and Public Health, University of Cadiz, Cádiz, Spain

[2]Computational Social Science DataLab, University Research Institute on Social Sciences, University of Cadiz, Jerez de la Frontera, Spain

Corresponding Author:
Victor Suarez-Lledo, BSc, MSc
Department of Biomedicine, Biotechnology and Public Health
University of Cadiz
Av Ana de Viya
Cádiz, 11009
Spain
Phone: 34 956019080
Email: victor.sanz@uca.es

Abstract

Background: Social media has changed the way we live and communicate, as well as offering unprecedented opportunities to improve many aspects of our lives, including health promotion and disease prevention. However, there is also a darker side to social media that is not always as evident as its possible benefits. In fact, social media has also opened the door to new social and health risks that are linked to health misinformation.

Objective: This study aimed to study the role of social media bots during the COVID-19 outbreak.

Methods: The Twitter streaming API was used to collect tweets regarding COVID-19 during the early stages of the outbreak. The Botometer tool was then used to obtain the likelihood of whether each account is a bot or not. Bot classification and topic-modeling techniques were used to interpret the Twitter conversation. Finally, the sentiment associated with the tweets was compared depending on the source of the tweet.

Results: Regarding the conversation topics, there were notable differences between the different accounts. The content of nonbot accounts was associated with the evolution of the pandemic, support, and advice. On the other hand, in the case of self-declared bots, the content consisted mainly of news, such as the existence of diagnostic tests, the evolution of the pandemic, and scientific findings. Finally, in the case of bots, the content was mostly political. Above all, there was a general overriding tone of criticism and disagreement. In relation to the sentiment analysis, the main differences were associated with the tone of the conversation. In the case of self-declared bots, this tended to be neutral, whereas the conversation of normal users scored positively. In contrast, bots tended to score negatively.

Conclusions: By classifying the accounts according to their likelihood of being bots and performing topic modeling, we were able to segment the Twitter conversation regarding COVID-19. Bot accounts tended to criticize the measures imposed to curb the pandemic, express disagreement with politicians, or question the veracity of the information shared on social media.

(J Med Internet Res 2022;24(8):e36085) doi:10.2196/36085

KEYWORDS

infodemics; social media; misinformation; epidemics; outbreaks; COVID-19; infodemiology; health promotion; pandemic; chatbot; social media bot; Twitter stream; Botometer; peer support

Introduction

Social media has radically changed the way we live and communicate. These new communication platforms offer unprecedented opportunities to improve many aspects of our lives, including public health [1,2]. They are useful to improving our access to evidence-based health information that can be fundamental to promoting healthy habits and fostering risk prevention [2]. In addition, the progressive growth of web-based, health-related knowledge and content has been found to be useful for patients who need to acquire medical skills and enhance their self-efficacy for adherence to treatments or therapies as well as for disease prevention [3].

Nevertheless, social media has also opened the door to new social and health risks [4,5]. Policies to mitigate misinformation and false health rumors are becoming increasingly common. In fact, some of the most widespread social media platforms such as Facebook, Instagram, and Twitter have implemented policies to combat the spread of misinformation regarding the COVID-19 pandemic. However, the web-based ecosystem is still overrun with health myths, hoaxes, and fake news stories that—either consciously or unconsciously—are propagated by social media users for different purposes. These messages can lead to attitude and behavior changes that may result in inadequate health decisions [6,7]. The effect of health misinformation has also been found to be determinant in health decision-making during risky situations and outbreaks such as the H5N1, Ebola, and Zika [5] viruses and the more recent COVID-19 pandemic [8,9]. Misleading messages have even hampered public health actions taken to tackle outbreaks [10-12]. For instance, in the context of the COVID-19 pandemic, misleading information has been detected regarding the origin of the virus, the potential treatments and protective measures available, and the real impact of the disease [13]. In one sample of tweets relating to COVID-19, 24.8% of the tweets included misinformation and 17.4% included unverifiable information [13]. Recently, much of the misinformation during the pandemic has focused on the debate regarding the vaccination process and the subsequent doubts the new vaccines have raised among the population [14].

Therefore, the role of social media during the COVID-19 pandemic has been critical. Although these new platforms have been useful to keep the public informed during the most critical moments of the pandemic, the responses by health authorities to combat the outbreak have been followed by a massive "infodemic," recently defined as "an overabundance of information—some accurate and some not—that makes it hard for people to find trustworthy sources and reliable guidance when they need it" [15]. Information consumption, opinion formation, and social contagion processes relating to COVID-19 across the social media ecosystem have become a major challenge for researchers [16], since these processes can strongly affect people's behavior and reduce the effectiveness of the countermeasures implemented by governments and health organizations [17].

Recently, misinformation dynamics have increased their complexity due to the emergence of so-called "social bots" (ie, automated web-based accounts). The role of social bots in the spread of misinformation on social media platforms has been widely recognized during political campaigns and election periods [18] and in relation to health debates, especially during health crises [19]. Regarding health communication on social media platforms, some studies have found that social bots are used to promote certain products to increase company profits and favor certain ideological positions [20] or contradict health evidence [21,22]. Bots have certain behavioral characteristics that make them potential super-spreaders of misinformation (eg, excessive posting and frequent retweeting of emerging news and tagging and mentions of influential topics or relevant figures) [20,23,24]. These accounts often use amplification as a strategy for the dissemination of content that misinforms based on the interests of the creators of these automatic accounts [25], although they are also often used as a tool to generate disagreement and social polarization [22].

The activity of social bots has dramatically increased in the context of the COVID-19 infodemic [25] due to their participation in the debate on the health measures to control the pandemic and the vaccines that have emerged during this period [26]. To date, it has been established that the progressive proliferation of social bots (and particularly unverified accounts) in the complex social media ecosystem may contribute to the increased spread of COVID-19 misinformation and the subsequent evolution of the pandemic, either by amplifying messages of dubious quality or generating polarization in relation to controversial issues [25]. However, a better understanding is needed on the role of these bots in the COVID-19 infodemic [27]. In an attempt to fill this knowledge gap, this study aimed to explore the role of social bots during the early stages of the COVID-19 pandemic. Our objective was to answer 3 basic questions: (1) What were the main conversation topics during the outbreak of COVID-19 on Twitter? (2) How do these topics vary depending on the information source (nonbots, bots, or self-declared bots)? and (3) How does the general tone of the conversation vary depending on the source?

Methods

Data Collection

Data collection started on March 16 and ended on June 15, 2020, using the Twitter streaming API with the following hashtags: *covid_19*, *covid19*, *covid*, and *coronavirus*. These hashtags were used during this period to capture the conversation during the first wave of COVID-19. To simplify the subsequent analysis, only tweets written in the English language were selected. The resulting data sample contained approximately 14 million tweets from about 285,000 different Twitter accounts.

Bot Classification

We used *Botometer* (formerly *BotOrNot*; OSoMe project) [28] to obtain the likelihood of whether each account is a bot or not. *Botometer* is a publicly available service that leverages more than 1000 features to evaluate the extent to which a Twitter account exhibits similarity to the known characteristics of social bots. As in other studies [29,30], 0.8 is the score used to classify an account as a bot. In addition, the percentage of bot accounts in benchmark studies is between 9% to 15% of the total number of accounts on Twitter [31]. In our case, this score classified approximately 14% of the accounts as bots.

In addition to the overall likelihood of being a bot, *Botometer* also gives specific scores for 6 different bot types: echo chamber, fake follower, financial, self-declared, spammer, and other. Given the differing nature of social bots, it was considered necessary to draw a distinction between self-declared bots and other types of bots. Self-declared bots are extracted from Botwiki [28,32].

Topic Modeling

Finally, together with the bot classification, we also applied topic-modeling techniques. This unsupervised classification approach allows the classification of texts, using techniques such as clustering to find groups of texts with similar content. In this case, we used latent dirichlet allocation (LDA), a popular topic-modeling technique which considers each document as a random mixture of various topics and each topic as a mixture of words [33].

To correctly interpret the results, we considered the distribution of the topics within the corpus, the keywords of each of the topics, and the intertopic distance [34]. Based on this, the most common topics of the different documents in the corpus were extracted. For each topic, we obtained the most relevant words and the 50 most characteristic tweets according to the model. We then carried out an inductive qualitative process to characterize each topic, followed by a descriptive process to codify the information [35]. Discrepancies were shared and resolved by mutual agreement. We also analyzed the distribution of the different types of accounts in the topics. This approach allowed us to determine the main conversation topics [36] and the most common ones for each type of account.

In addition, we plotted an intertopic distance map [34] to visualize the topics in a 2D space. The area of the topic circles is proportional to the number of tokens (ie, single words) that belong to each topic across the dictionary. The circles were plotted using a multidimensional scaling algorithm based on the words they comprise, with the topics that are closer together having more words in common.

Sentiment Analysis

For each of the groups, we used sentiment analysis to examine the tone or sentiment associated with the content. Sentiment analysis is an area of knowledge in the field of natural language processing, text analysis, and computational linguistics used to identify and extract subjective information from resources. In the case of text mining, sentiment analysis involves automatically mass-classifying documents based on the positive or negative connotation of the language in the document [37].

For the sentiment extraction, we used Valence Aware Dictionary and Sentiment Reasoner (VADER), a rule-based tool specifically attuned to sentiments expressed on social media platforms [38]. VADER uses a combination of sentiments associated with lexicons that are generally labeled according to their semantic orientation as positive or negative. Unlike other text analysis tools, VADER works well on texts extracted from social media platforms, because it does not need as much text as other tools [39-41].

Another feature of this method is the output value. Most sentiment analyses classify texts as positive, negative, and neutral; for example, texts with a predominance of words, expressions, or ways of writing perceived as positive are classified as positive. However, the method used here returns a sentiment score between –1 and 1, allowing a higher level of comparison between the different types of accounts.

Results

Bot Classification

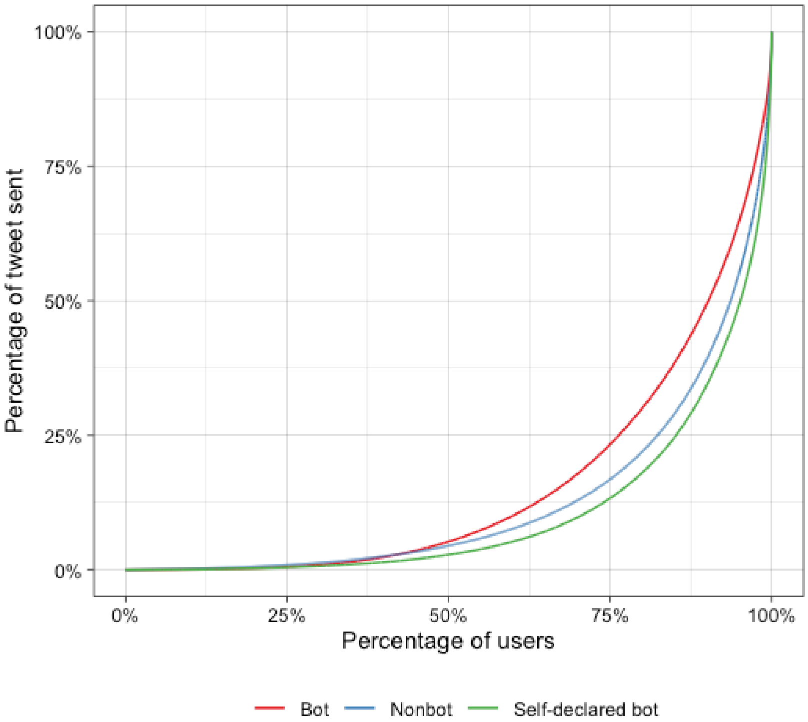
Table 1 shows the resulting classification. If the probability of an account being a bot is lower than 0.8, we considered it as a normal user (ie, nonbot). If the probability of an account being a self-declared bot is higher of 0.8, we classified it as a self-declared bot. Accounts with the probability of being a bot higher than 0.8 and the probability of being a self-declared bot lower than 0.8 were classified as bots. Of the 205,298 accounts, most (n=187,992, 91.6%) were normal users; 4.2% (n=8616) were classified with a high likelihood of being bot accounts; and 4.2% (n=8690) were classified as self-declared bots. Bot accounts posted an average of 123.3 tweets per user. During the 3-month time window, accounts classified as self-declared bots posted a slightly lower average of 121.1 tweets per user. However, accounts classified as having a low likelihood of being bots posted 42.5 tweets per user. These differences between the mean values were statistically significant ($F_{2,284,814}$=1056; P<.001). As also noted in Broniatowski et al [22], the most active accounts on average were those classified as bots.

Not all groups contributed to the same extent. Likewise, the contribution of the participants in the global discussion was highly unequal. The Gini index was used to measure this inequality. This index is a measure of the distribution, with a higher Gini index indicating greater inequality. Figure 1 shows these distributions, with a Gini index of 0.786 for self-declared bots, 0.744 for nonbots, and 0.686 for bots. Self-declared bots had the most unequal distribution: 75% (6517/8690) of self-declared bots posted 12.5% (131,559/1,052,471) of the total number of tweets. In contrast, 75% (6462/8616) of the bot accounts posted 25% (265,499/1,061,997) of the total tweets.

Table 1. Distribution of bot classification.

Source	Account (N=205,298), n (%)	Tweet (N=10,098,455)		
		n (%)	Mean	Median
Nonbot	187,992 (91.6)	7,983,987 (79.1)	42.5	9.0
Bot	8616 (4.2)	1,061,997 (10.5)	123.3	35.5
Self-declared bot	8690 (4.2)	1,052,471 (10.4)	121.1	15.0

Figure 1. Lorenz curve showing inequality in the number of tweets.



In the case of self-declared bots, the most active accounts spread official data (the number of COVID-19 cases and mortality, etc). Second, several of these accounts were digital magazines or independent news agencies. The descriptions of these accounts mentioned that they created messages to provide periodic reports on the situation and communicate the global evolution of COVID-19 or substantial changes in the evolution of the pandemic. Many of these accounts indicated that their purpose was informative. Given this situation, these profiles were separated from those classified as regular bots in the analysis.

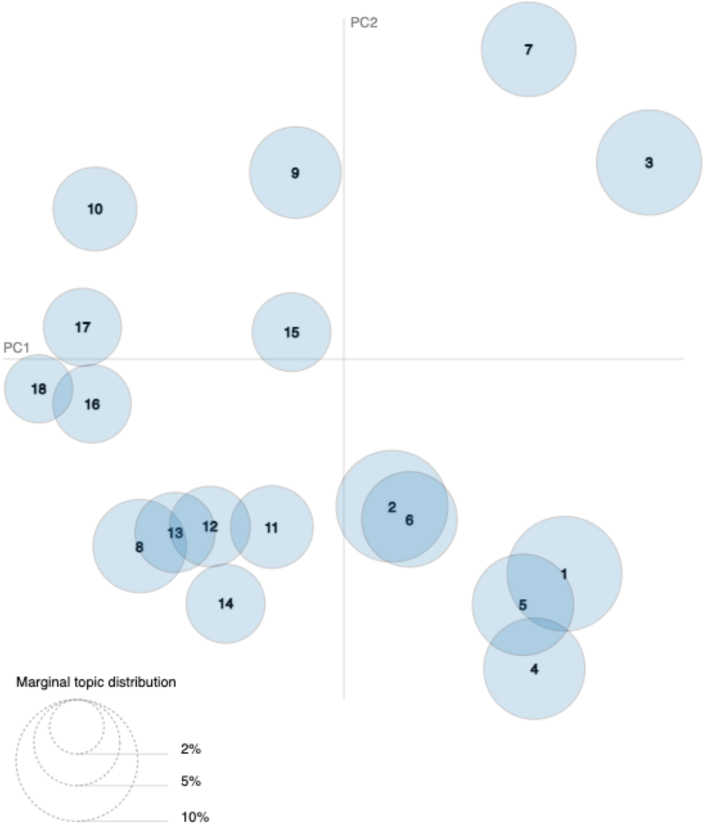
The descriptions of the accounts classified as bots were very different from each other. Many accounts identified themselves with technology companies. Others identified themselves as activists, either political, environmental, or even military. These accounts tweeted about the pandemic, the political measures taken, or complaints about the situation resulting from inaction.

Topic Modeling

After classifying the accounts, the topics were extracted using LDA. To select the correct number of topics, we relied on the coefficient of variation, which measures the coherence between the topics inferred by a model. In other words, the coefficient indicates which combination of topics is the most coherent. Higher values indicate that the topics are semantically interpretable. Topic coherence measures score a single topic by measuring the degree of semantic similarity between high-scoring words in the topic. This concept brings together several measures to assess the coherence between the topics. To choose the number of topics, the LDA model was reapplied with different outputs, and those with the highest coefficient of variation were selected (Multimedia Appendix 1). In total, 18 topics were extracted and plotted using the intertopic distance map.

In the intertopic distance map below (Figure 2), each bubble represents a topic. Each topic was assigned a number depending on the number of tweets inside it. Accordingly, Topic 1 had a higher percentage of tokens than Topic 2 and so on. The larger the bubble, the higher the number of tokens classified in this topic. The further the topics are away from each other, the more different they are. Therefore, there are not many differences between 2 nearby topics. On the contrary, there are greater differences if they are further apart.

Figure 2. Intertopic distance map. PC: principal component.



We also plotted the most common terms in a bar chart (Figure 3). The terms were sorted according to the number of times they appear. The colored bars show the estimated number of times a term is in each topic. The grey bars represent the overall frequency of each term in the corpus. When interpreting the results, it is not only necessary to consider the most common terms but also the most salient terms. Saliency is the product of weighting the probability of a word, *P*(w), by its distinctiveness, a measure of how informative the specific term is to determine the generative topic. Saliency is therefore a measure of the degree to which the word appears a small number of times or not at all in other topics [36].

We then carried out the qualitative phase, with both authors agreeing in 89% (16/18) of the cases. Table 2 shows the results obtained in the classification.

In Figure 4, each line is a topic, and each dot represents the percentage of accounts in each topic. Topic 1 contained tweets with information on the outbreak. The messages were focused on providing information about the advance of the pandemic and what actions need to be taken to stop it. The most common words were *stay*, *home*, and *family*. Other tweets shared this kind of information but for specific regions. For example, Topic 9 was focused on regions in Africa, and Topic 13 was focused on the lockdown in India.

These 2 topics have the most substantial differences between self-declared bots and the rest of the accounts. Topic 9 accumulated the highest percentage (1581/8690, 18.2%) of self-declared bot accounts, compared to bots (896/8616, 10.4%) and nonbots (20,115/187,992, 10.7%). Likewise, the percentage of self-declared bots in Topic 13 is 5.7% (495/8690), whereas the percentage for bots is 3.4% (293/8616) and 3.8% (7144/187,992) for nonbots.

Topic 2 contained information about the evolution of the pandemic. This topic was focused on the second wave and information on the number of deaths. The most common keywords were *case*, *death*, *report*, and *total*. In the following topics, the model groups' contents were related to specific measures to curb the pandemic. Topic 3 mentioned the lack of testing. Some topics reminded people to stay at home (Topic 4), of the importance of wearing a face mask (Topic 11), or of washing one's hands (Topic 12).

Other messages were related to US politics or President Trump. Most of the tweets in Topic 17 were about decisions by the US Congress. Topic 18 mentioned certain national political scandals. Topic 8 was focused on criticizing President Trump's policies. These tweets cast President Trump as a liar and irresponsible. Some of the most common keywords were *president*, *Trump*, *China*, *virus*, *year*, and *world*. This topic had the biggest difference between the percentage of bots and the rest of the accounts. In Topic 16, most of the tweets mentioned the lack of honesty of the US President. There were also complaints about the need to share true information and disregard rumors (Topic 15). In these last 2 topics, the percentage of bot accounts was slightly higher than the rest of the accounts.

Figure 3. Word distribution along with topics.

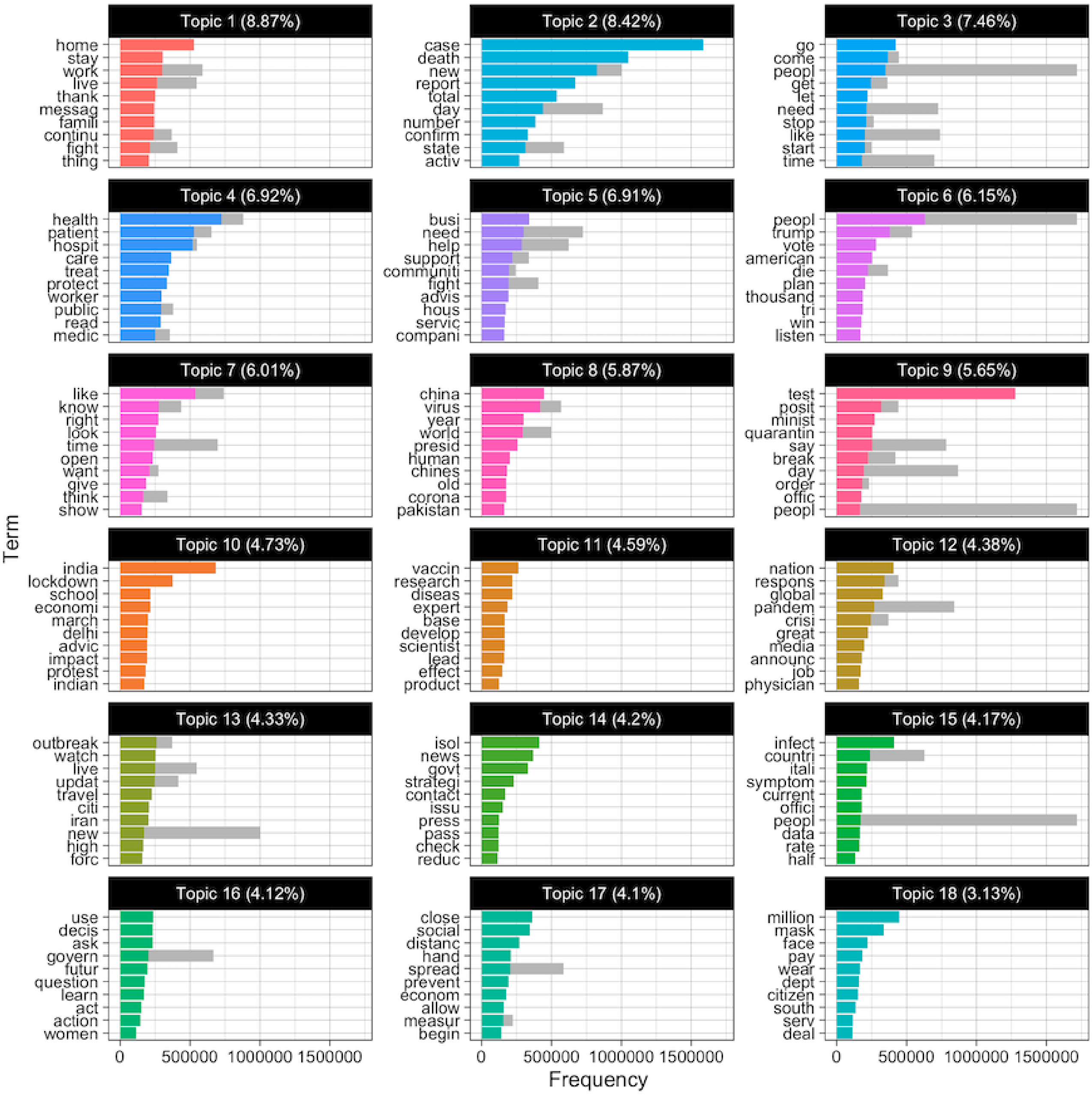
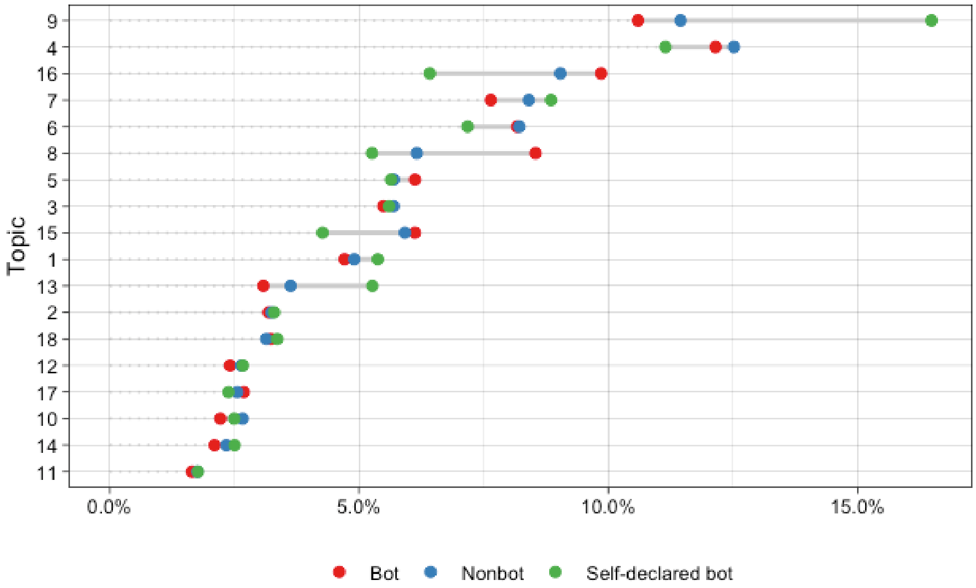


Table 2. Main idea for each topic.

ID	Topic
1	News about coronavirus
2	Second wave and vaccines
3	Complaints about lack of COVID-19 testing
4	Stay home
5	China and its relationship with the virus
6	Respect health care workers
7	Financial aid and charity during the pandemic
8	Trump and the pandemic
9	Reporting positive cases in Maharashtra and Africa
10	Pointing out that COVID-19 is different from influenza
11	Wearing face masks
12	Tips to prevent spreading COVID-19
13	Lockdown in India
14	Death of a famous person
15	Calls for real leadership
16	A call for honesty
17	Decisions in the US Congress
18	A national scandal

Figure 4. Account distribution within topics.

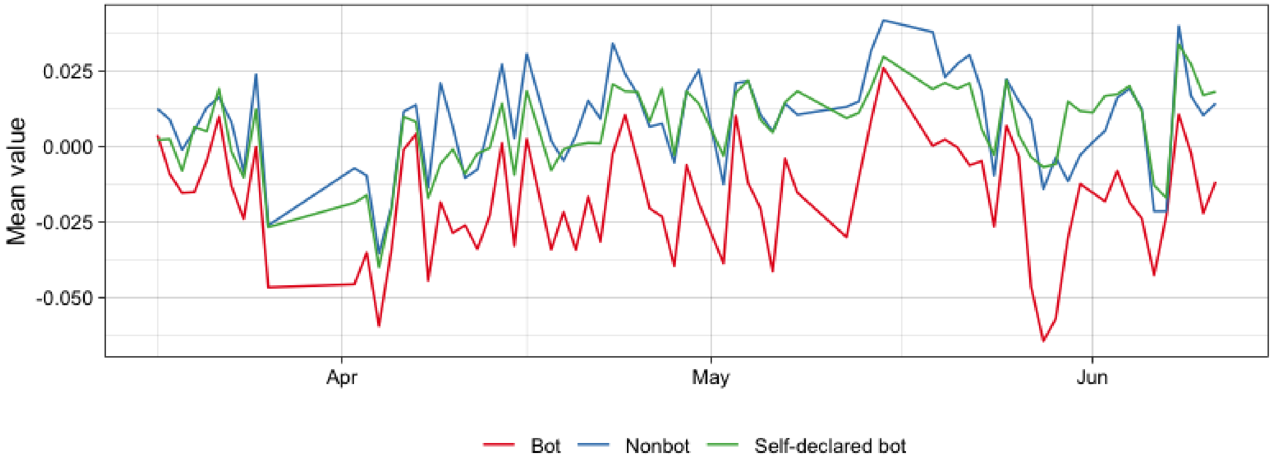


Sentiment Analysis

The mean value of the VADER score for each group was 0.0109 (SD 0.414) for nonbots, 0.00784 (SD 0.383) for self-declared bots, and –0.0155 (SD 0.427) for bots. An ANOVA test was used to check for statistically significant differences in the mean values of the groups ($F_{2,284,814}$=5216; P<.001). Figure 5 shows the evolution of the average scores over the period. The mean value was almost always lower in the case of bots, indicating a greater presence of words associated with negative feelings in this group. Accounts classified as self-declared bots were closer to values of 0. On the other hand, accounts classified as bots scored negatively.

These differences in sentiment between nonbots, self-declared bots, and undeclared bots are better understood if we consider the different topics that made up these conversations. Although most of the tweets posted by nonbots were focused on sharing the situation people were experiencing due to the outbreak, self-declared bots tended to inform and post news on the outbreak all over the world, and undeclared bots were generally focused on criticizing political measures, interpersonal blame between senators or governors, and criticism directed at governments or political leaders in relation to the mismanagement of the health crisis. At this point in the analysis, it seemed more likely that undeclared bots spread messages of disagreement, criticism, and complaints regarding the political and health authorities in view of the difficulties to adequately control the pandemic.

Figure 5. Mean value for Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analysis.



Discussion

Principal Findings

This study has allowed an assessment of the role of social bots on Twitter during the early stages of the COVID-19 pandemic. There were consistent differences between the different account types identified (self-declared bots, undeclared bots, and nonbots). Although the percentage of undeclared bots on Twitter is relatively low compared to the large number of human users, it has been established that bots are generally linked to web-based conversations characterized by controversy and polarization. In this sense, the role of these automatic agents is far from negligible, considering the role they play in the amplification of ideas and opinions that generate conflict in our societies [42,43].

The classification adopted has allowed the comparison of the different topics arising in the conversations of 3 different profiles of Twitter users during the initial months of the pandemic. Furthermore, to the best of our knowledge, this study has several advantages compared to other works that analyze sentiment in a general manner and regardless of the information source and type [44-46]. First, our study provides additional information on the information sources (nonbots, self-declared bots, and bots), particularly concerning the credibility of the different Twitter users. Second, it allows a deeper analysis of the Twitter conversation based on topics and the associated sentiments during the outbreak of the COVID-19 pandemic. Third, the comparison of the topics according to source shows there is internal consistency between the different types of accounts. Therefore, the differentiation of topics and sentiments linked to different Twitter user accounts (and particularly those relating to bots) is relevant for the identification, characterization, and monitoring of possible sources of disinformation that could emerge in the event of an infodemic [47].

On the other hand, the sentiment analysis also gives an idea of the strategy of undeclared bots or automated accounts in the context of the first months of the COVID-19 pandemic. Our study shows that social bots were used to criticize and harass political opponents rather than to provide useful information on health measures and self-protection behavior in a context where quality information was sorely needed in the face of widespread misinformation [47]. In-line with our results, a recent study indicates that right-wing self-media accounts and conspiracy theorists may give rise to this opinion polarization, whereas malicious bots may foster the diffusion of noncredible information [42]. We have not found large amounts of misinformation on health issues but rather major divisions regarding political decision-making processes and the measures to address the COVID-19 pandemic (eg, vaccines and protective measures, etc). In this sense, the conversation on automated accounts is directed more toward generating conflict and disagreement [43].

Despite these findings, additional evidence is needed to determine the social and health impacts of the misuse of social bots during the early months of the pandemic. Likewise, it is necessary to determine to what extent these agents have hindered the prevention and control of the health crisis by the different governments. In any case, this is a new working hypothesis that remains open and should be analyzed in detail in future studies.

Limitations and Strengths

This study is subject to several limitations. First, the data collected from Twitter is limited by the technical characteristics of the Twitter streaming API. Although the streaming API is more accurate than the REST API, it never returns the total number of tweets about the conversation [48]. Moreover, due to technical limitations, it is impossible to analyze the entire conversation. In addition, by selecting only tweets written in the English language, the content of the conversations is strongly focused on topics in the United States and United Kingdom. Second, the period analyzed is in the early stages of the outbreak, and the conversations tended to evolve just as the pandemic did. Third, when observing self-expression over the internet, only the thoughts and feelings the users chose to express at the time can be captured, which may be strategically composed to project a public persona [49]. Still, many mental health studies have shown that social media is a valuable outlet and source of support for its users [50]. Topic modeling is a good technique to obtain a general idea of the different topics within a conversation. However, the downside of this technique is that the number of topics must be preselected. In our case, we used the coefficient of variation to identify the optimal number for each group.

On the other hand, this study also has several strengths. First, it takes into account the credibility of the information source. This aspect is rarely addressed in studies of social media platforms [2]. Second, this study analyzes conversations regarding the outbreak of a pandemic, and social media sites are hot spots in such situations [5], with users increasing their information searches on these platforms.

Conclusions

By classifying the accounts according to the likelihood of being bots and applying topic modeling, we were able to segment the Twitter conversations regarding the COVID-19 pandemic. Nonbot accounts, for example, tended to share information or give advice on how to deal with the pandemic. The accounts declared as bots mostly shared information and statistics on the pandemic. Finally, accounts not declared as bots tended to criticize the measures imposed to curb the pandemic, express disagreement with politicians, or question the veracity of the information shared on social media platforms. We also used sentiment analysis to compare the tone of the conversations in these different groups. Self-declared bots had conversations with a neutral tone. The tone of messages written by nonbot accounts tended to be more positive than the former. On the contrary, the tone of undeclared bots was always more negative than the tone of self-declared bots. Therefore, it is necessary to work on the identification and monitoring of these agents in times of infodemics.

Acknowledgments

We would like to acknowledge the support of the University Research Institute for Sustainable Social Development, the University of Cádiz, and the Ramon y Cajal program run by the Spanish Ministry of Science and Innovation.

This work was supported by the Ramon and Cajal grant awarded to JAG (RYC-2016-19353) and the DCODES project (PID2020-118589RB-I00), financed by MCIN/ AEI/10.13039/501100011033.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Score of the coefficient of variation by the number of topics.
[PDF File (Adobe PDF File), 119 KB - jmir_v24i8e36085_app1.pdf]

References

1. McGloin AF, Eslami S. Digital and social media opportunities for dietary behaviour change. Proc Nutr Soc 2015 May;74(2):139-148. [doi: 10.1017/S0029665114001505] [Medline: 25319345]
2. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
3. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
4. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
5. Alvarez-Galvez J, Suarez-Lledo V, Rojas-Garcia A. Determinants of infodemics during disease outbreaks: a systematic review. Front Public Health 2021;9:603603 [FREE Full text] [doi: 10.3389/fpubh.2021.603603] [Medline: 33855006]
6. Cavallo DN, Chou WS, McQueen A, Ramirez A, Riley WT. Cancer prevention and control interventions using social media: user-generated approaches. Cancer Epidemiol Biomarkers Prev 2014 Sep;23(9):1953-1956 [FREE Full text] [doi: 10.1158/1055-9965.EPI-14-0593] [Medline: 25103820]
7. Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube. PLoS One 2014;9(10):e110171 [FREE Full text] [doi: 10.1371/journal.pone.0110171] [Medline: 25333470]
8. Funk S, Gilad E, Watkins C, Jansen VAA. The spread of awareness and its impact on epidemic outbreaks. Proc Natl Acad Sci U S A 2009 Apr 21;106(16):6872-6877 [FREE Full text] [doi: 10.1073/pnas.0810762106] [Medline: 19332788]
9. Betsch C. The role of the Internet in eliminating infectious diseases. managing perceptions and misperceptions of vaccination. Article in German. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2013 Sep 24;56(9):1279-1286. [doi: 10.1007/s00103-013-1793-3] [Medline: 23990090]

10. Househ M. Communicating Ebola through social media and electronic news media outlets: a cross-sectional study. Health Informatics J 2016 Sep 26;22(3):470-478 [FREE Full text] [doi: 10.1177/1460458214568037] [Medline: 25656678]
11. Catalan-Matamoros D, Peñafiel-Saiz C. How is communication of vaccines in traditional media: a systematic review. Perspect Public Health 2019 Jan 07;139(1):34-43. [doi: 10.1177/1757913918780142] [Medline: 29877122]
12. Levy JA, Strombeck R. Health benefits and risks of the Internet. J Med Syst 2002 Dec;26(6):495-510. [doi: 10.1023/a:1020288508362] [Medline: 12385532]
13. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
14. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
15. EPI-WIN: WHO Information Network for Epidemics. World Health Organization. 2020. URL: https://www.who.int/teams/epi-win [accessed 2020-03-09]
16. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
17. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
18. Caldarelli G, de Nicola R, del Vigna F, Petrocchi M, Saracco F. The role of bot squads in the political propaganda on Twitter. Commun Phys 2020 May 11;3(81):1-15. [doi: 10.1038/s42005-020-0340-4]
19. Allem J, Ferrara E. Could social bots pose a threat to public health? Am J Public Health 2018 Aug;108(8):1005-1006. [doi: 10.2105/AJPH.2018.304512] [Medline: 29995482]
20. Bessi A, Ferrara E. Social bots distort the 2016 U.S. Presidential election online discussion. First Monday 2016 Nov 03;21(11). [doi: 10.5210/fm.v21i11.7090]
21. Betsch C, Brewer NT, Brocard P, Davies P, Gaissmaier W, Haase N, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. Vaccine 2012 May 28;30(25):3727-3733. [doi: 10.1016/j.vaccine.2012.02.025] [Medline: 22365840]
22. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
23. Shao C, Ciampaglia GL, Varol O, Yang K, Flammini A, Menczer F. The spread of low-credibility content by social bots. Nat Commun 2018 Nov 20;9(1):4787 [FREE Full text] [doi: 10.1038/s41467-018-06930-7] [Medline: 30459415]
24. Yuan X, Schuchard RJ, Crooks AT. Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. Soc Media Soc 2019 Sep 04;5(3):205630511986546. [doi: 10.1177/2056305119865465]
25. Himelein-Wachowiak M, Giorgi S, Devoto A, Rahman M, Ungar L, Schwartz H, et al. Bots and misinformation spread on social media: implications for COVID-19. J Med Internet Res 2021 May 20;23(5):e26933 [FREE Full text] [doi: 10.2196/26933] [Medline: 33882014]
26. Zhang M, Qi X, Chen Z, Liu J. Social bots' involvement in the COVID-19 vaccine discussions on Twitter. Int J Environ Res Public Health 2022 Jan 31;19(3):1651 [FREE Full text] [doi: 10.3390/ijerph19031651] [Medline: 35162673]
27. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nat Hum Behav 2020 Dec 29;4(12):1285-1293. [doi: 10.1038/s41562-020-00994-6] [Medline: 33122812]
28. Sayyadiharikandeh M, Varol O, Yang K, Flammini A, Menczer F. Detection of novel social bots by ensembles of specialized classifiers. 2020 Oct 19 Presented at: CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management; October 19-23, 2020; Virtual event Ireland p. 2725-2732. [doi: 10.1145/3340531.3412698]
29. Luceri L, Badawy A, Deb A, Ferrara E. Red bots do it better: comparative analysis of social bot partisan behavior. 2019 May 13 Presented at: WWW '19: Companion Proceedings of The 2019 World Wide Web Conference; May 13-17, 2019; San Francisco, CA p. 1007-1012. [doi: 10.1145/3308560.3316735]
30. Yang K, Varol O, Davis CA, Ferrara E, Flammini A, Menczer F. Arming the public with artificial intelligence to counter social bots. Human Behav and Emerg Tech 2019 Feb 06;1(1):48-61. [doi: 10.1002/hbe2.115]
31. Varol O, Ferrara E, Davis C, Menczer F, Flammini A. Online human-bot interactions: detection, estimation, and characterization. 2017 May 03 Presented at: Proceedings of the Eleventh International AAAI Conference on Web and Social Media; May 15-18, 2017; Montreal, Quebec, Canada p. 280-289 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14871
32. Botwiki. URL: https://botwiki.org/ [accessed 2022-07-18]
33. Rehurek R, Sojka P. Gensim--python framework for vector space modelling. 2011 Presented at: EuroScipy 2011; August 25-28, 2011; Paris, France URL: https://www.fi.muni.cz/usr/sojka/posters/rehurek-sojka-scipy2011.pdf
34. Sievert C, Shirley K. LDAvis: a method for visualizing and interpreting topics. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces Baltimore, Maryland, USA: Association for Computational Linguistics; 2014 Jun Presented at: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces; June 27, 2014; Baltimore, MD p. 63-70. [doi: 10.3115/v1/w14-3110]

35. Colorafi KJ, Evans B. Qualitative descriptive methods in health science research. HERD 2016 Jul;9(4):16-25 [FREE Full text] [doi: 10.1177/1937586715614171] [Medline: 26791375]
36. Chuang J, Manning C, Heer J. Termite: visualization techniques for assessing textual topic models. New York, NY: Association for Computing Machinery; 2012 May 21 Presented at: AVI '12: Proceedings of the International Working Conference on Advanced Visual Interfaces; May 21-25, 2012; Capri Island, Italy p. 74-77. [doi: 10.1145/2254556.2254572]
37. Liu B. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies 2012 May 23;5(1):1-167. [doi: 10.2200/s00416ed1v01y201204hlt016]
38. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. 2014 May 16 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 216-225 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14550
39. Elbagir S, Yang J. Twitter sentiment analysis using natural language toolkit and VADER sentiment analyzer. In: Lecture Notes in Engineering and Computer Science. 2019 Presented at: Proceedings of The International MultiConference of Engineers and Computer Scientists 2019; March 13-15, 2019; Hong Kong, China p. 12-16. [doi: 10.1142/9789811215094_0005]
40. Pano T, Kashef R. A complete VADER-based sentiment analysis of Bitcoin (BTC) tweets during the era of COVID-19. Big Data Cogn Comput 2020 Nov 09;4(4):33. [doi: 10.3390/bdcc4040033]
41. Adarsh R, Patil A, Rayar S, Veena K. Comparison of VADER and LSTM for sentiment analysis. Int J Recent Technology Engineering 2019 Mar 01;7(6):540-543.
42. Xu W, Sasahara K. Characterizing the roles of bots on Twitter during the COVID-19 infodemic. J Comput Soc Sci 2022;5(1):591-609 [FREE Full text] [doi: 10.1007/s42001-021-00139-3] [Medline: 34485752]
43. Uyheng J, Carley KM. Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines. J Comput Soc Sci 2020;3(2):445-468 [FREE Full text] [doi: 10.1007/s42001-020-00087-4] [Medline: 33102925]
44. Pokharel BP. Twitter sentiment analysis during COVID-19 outbreak in Nepal. SSRN Journal 2020 Jun 11:1-26. [doi: 10.2139/ssrn.3624719]
45. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Jul 06;22(4):418-421 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]
46. Tsai MH, Wang Y. Analyzing Twitter data to evaluate people's attitudes towards public health policies and events in the era of COVID-19. Int J Environ Res Public Health 2021 Jun 10;18(12):6272 [FREE Full text] [doi: 10.3390/ijerph18126272] [Medline: 34200576]
47. Stella M, Ferrara E, de Domenico M. Bots increase exposure to negative and inflammatory content in online social systems. Proc Natl Acad Sci U S A 2018 Dec 04;115(49):12435-12440 [FREE Full text] [doi: 10.1073/pnas.1803470115] [Medline: 30459270]
48. Wang Y, Callan J, Zheng B. Should we use the sample? analyzing datasets sampled from Twitter’s stream API. ACM Trans Web 2015 Jun 20;9(3):1-23. [doi: 10.1145/2746366]
49. Marshall PD. Intercommunication and persona: the intercommunicative public self. Int J Interdisciplinary Stud Commun 2015;10(1):23-31. [doi: 10.18848/2324-7320/cgp/v10i01/53601]
50. McCloud RF, Kohler RE, Viswanath K. Cancer risk-promoting information: the communication environment of young adults. Am J Prev Med 2017 Sep;53(3S1):S63-S72 [FREE Full text] [doi: 10.1016/j.amepre.2017.03.025] [Medline: 28818248]

Abbreviations

LDA: latent dirichlet allocation
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by M Gisondi, L Westafer; submitted 31.12.21; peer-reviewed by M Anderson, J Uyheng, I Skafle; comments to author 13.02.22; revised version received 04.05.22; accepted 24.05.22; published 25.08.22.

Please cite as:
Suarez-Lledo V, Alvarez-Galvez J
Assessing the Role of Social Bots During the COVID-19 Pandemic: Infodemic, Disagreement, and Criticism
J Med Internet Res 2022;24(8):e36085
URL: https://www.jmir.org/2022/8/e36085
doi:10.2196/36085
PMID:35839385

Research Letter

Recognition of Gait Patterns in Older Adults Using Wearable Smartwatch Devices: Observational Study

Hyeon-Joo Kim[1*], BSc; Hyejoo Kim[1*], BSc; Jinyoon Park[1], MSc; Bumjo Oh[2], MD, MPH; Seung-Chan Kim[1], PhD

[1]Machine Learning Systems Lab, College of Sports Science, Sungkyunkwan University, Suwon, Republic of Korea

[2]Department of Family Medicine, Seoul Metropolitan Government - Seoul National University Boramae Medical Center, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Seung-Chan Kim, PhD
Machine Learning Systems Lab
College of Sports Science
Sungkyunkwan University
2066 Seoburo, Jangan-gu
Suseong Bldg #05111a
Suwon, 16419
Republic of Korea
Phone: 82 10 2533 1915
Email: seungk@g.skku.edu

(J Med Internet Res 2022;24(8):e39190) doi:10.2196/39190

KEYWORDS

activity recognition; machine learning; health monitoring; gait analysis; wearable; sequence classification; mobile health; mHealth; neural network

Introduction

It is challenging to routinely assess gait unless dedicated measuring devices are available. Inspired by a recent study that reported high classification performance of activity recognition tasks using smartwatches [1], we hypothesized that the recognition of gait-related activities in older adults can be formulated as a supervised learning problem. To quantify the complex gait motion, we focused on hand motion because disturbed hand motions are frequently reported as typical symptoms of neurodegenerative diseases [2].

Methods

Data Acquisition

We recruited 39 older adult participants (age: 80.4, SD 6.5 years; n=38, 73.7% women) from a local community. The number of participants for each class was as follows: cane-assisted gait (C0) (n=7), walker-assisted gait (C1) (n=5), gait with disturbances (C2) (n=21), gait without disturbances (C3) (n=6), and gait without disturbances in young controls (C4) (n=12). During the experiment, participants were asked to wear a smartwatch (DW9F1; Fossil Group, Inc) on each wrist and walk at a normal speed similar to their usual walk. Figure 1 shows example photographs taken during the experiment.

Figure 1. Five different gait styles: cane-assisted gait (C0), walker-assisted gait (C1), gait with disturbances (C2), gait without disturbances (C3), and gait without disturbances in young controls (C4).

Classification

The multivariate time-series (MTS) signals captured at a sampling rate of 50 Hz were segmented into . Here, represents the inertial motion at a specific moment, *t*. In this study, *D* was 12 (=6×2), since each smartwatch measures the 6-DOF (6 degrees of freedom) motion separately, and *T* was 100 (approximately 2s) so that each **x** could contain at least a full gait cycle. The task in our study was to infer the type of gait activity, , where *C* was 5. Our neural network systems, tailored to learn gait features from MTS data, were trained in an end-to-end fashion using state-of-the-art deep learning architectures, including Conv1D [3], long short-term memory (LSTM) [4], and an LSTM with an attention mechanism [5].

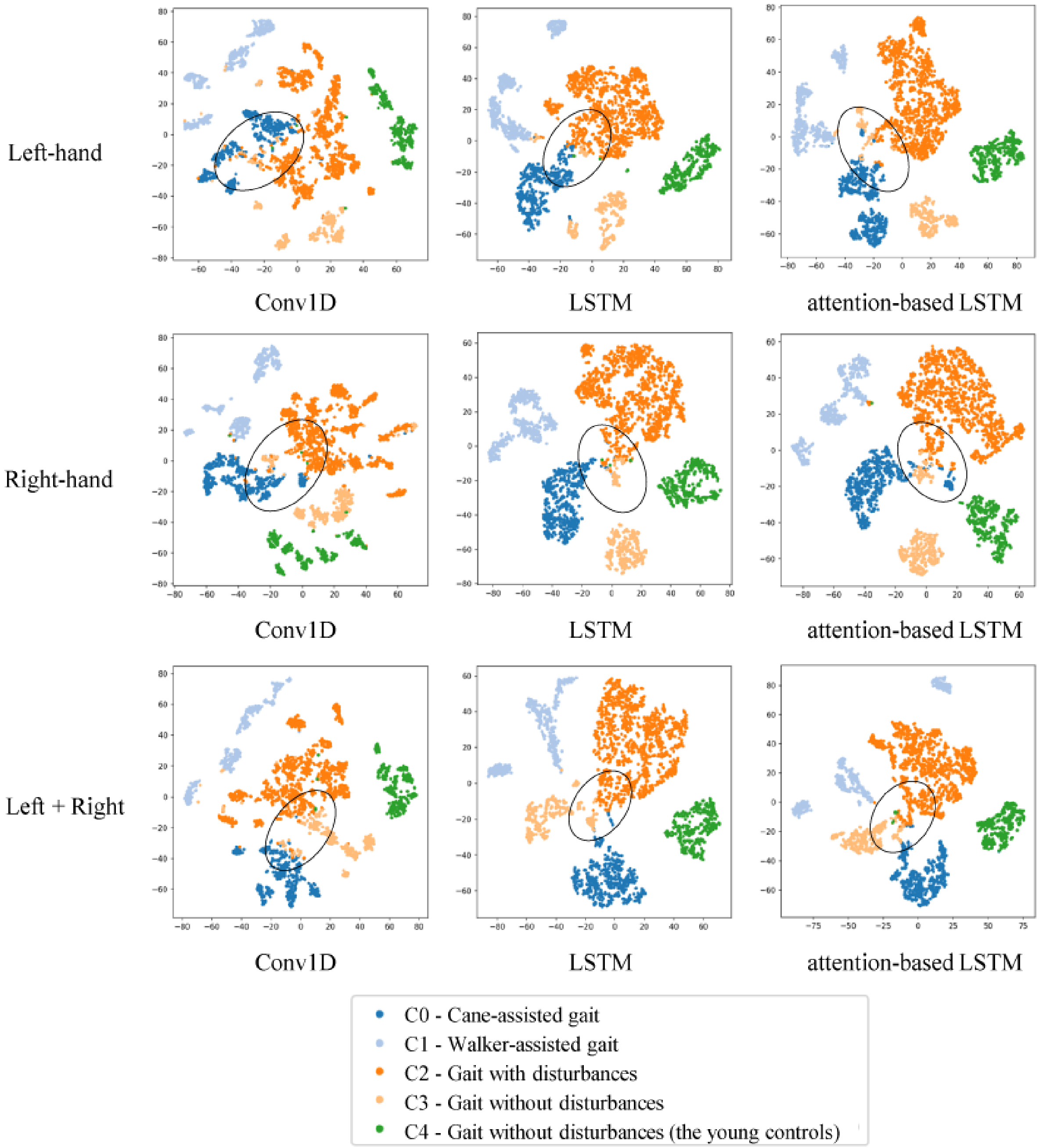
Ethics Approval

All participants were enrolled after institutional review board (IRB) approval (Sungkyunkwan University IRB approval number: SKKU 2021-12-014).

Results

We employed the accuracy and macro average of the F_1-score, F_m, as a measure of performance. For the both-hands condition, the accuracy (F_m) was 0.9757 (0.9728), 0.9736 (0.9699), and 0.9771 (0.9738) when Conv1D, LSTM, and attention-based LSTM were employed, respectively. In the case of the left-hand and right-hand conditions, the accuracies (F_m) obtained in the left-hand condition were 0.9652 (0.9623), 0.9611 (0.9583), and 0.9630 (0.9592), respectively. In the right-hand condition, the accuracies (F_m) were 0.9724 (0.9706), 0.9673 (0.9643), and 0.9673 (0.9635) for the same employed models, respectively. We also examined the learned representations as shown in Figure 2 using t-distributed stochastic neighbor embedding (t-SNE) [6], which visualizes the high-dimensional vectors by projecting them into a 2D space in such a way that similar points cluster together.

Figure 2. Feature visualization using t-distributed stochastic neighbor embedding. Each point is colored according to the predicted class. LSTM: long short-term memory.



Discussion

The experimental results demonstrated an acceptable classification performance (ie, both accuracy and the F_m score were higher than 0.95). However, there is systematic confusion, such as recognizing C3 as C2 (0.03-0.04 for the left hand, 0.05-0.07 for the right hand, and 0.05-0.06 for both hands, respectively) as shown in Figure 2 (see the region highlighted in black). It is noteworthy that the classification performance of the single-hand condition was similar to that of the both-hands condition, suggesting that wearing a single smartwatch is sufficient for the proposed gait assessment task. From the t-SNE plot, it was found that points from the LSTM and attention-based LSTM exhibit a more clustered distribution than those from the Conv1D model. We expect that the proposed approach can be applied to various health care applications for older adults (eg, wearable detection of gait disturbances).

Acknowledgments

This work was supported by a grant from the National Research Foundation of Korea (#NRF-2020R1C1C1010666). This work was also supported by Sungkyunkwan University and the BK21 FOUR (Graduate School Innovation) funded by the Ministry of Education (Korea) and the National Research Foundation of Korea.

Authors' Contributions

SCK and BO were responsible for the study concept and design; SCK and HK were involved in development; SCK, HJK, and JP conducted the analysis and interpreted the data; HK provided the visualizations; and all authors helped write the manuscript.

Conflicts of Interest

None declared.

References

1. Kim H, Kim HJ, Park J, Ryu JK, Kim SC. Recognition of fine-grained walking patterns using a smartwatch with deep attentive neural networks. Sensors (Basel) 2021 Sep 24;21(19):6393 [FREE Full text] [doi: 10.3390/s21196393] [Medline: 34640712]
2. Snijders AH, van de Warrenburg BP, Giladi N, Bloem BR. Neurological gait disorders in elderly people: clinical approach and classification. Lancet Neurol 2007 Jan;6(1):63-74. [doi: 10.1016/S1474-4422(06)70678-0]
3. Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP; October 25-29; Doha, Qatar p. 1746-1751 URL: https://aclanthology.org/D14-1181.pdf
4. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
5. Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. ACL Anthology 2015:1412-1421 [FREE Full text] [doi: 10.18653/v1/D15-1166]
6. Van der Maaten L, Hinton G. Visualizing data using t-SNE. JMLR 2008;9:2579-2605 [FREE Full text]

Abbreviations

6-DOF: 6 degrees of freedom
LSTM: long short-term memory
MTS: multivariate time-series
t-SNE: t-distributed stochastic neighbor embedding

Edited by R Kukafka; submitted 03.05.22; peer-reviewed by M Kraus, K Alexander; comments to author 15.07.22; revised version received 23.07.22; accepted 11.08.22; published 24.08.22.

Please cite as:
Kim HJ, Kim H, Park J, Oh B, Kim SC
Recognition of Gait Patterns in Older Adults Using Wearable Smartwatch Devices: Observational Study
J Med Internet Res 2022;24(8):e39190
URL: https://www.jmir.org/2022/8/e39190
doi:10.2196/39190
PMID:36001374

Retraction

Retraction: “Assessing the Dissemination of COVID-19 Articles Across Social Media With Altmetric and PlumX Metrics: Correlational Study”

Haley N Tornberg[1*], BS; Carine Moezinia[1*], MD; Chapman Wei[2], BS; Simone A Bernstein[3], MD; Chaplin Wei[4], BS; Refka Al-Beyati[5], MD; Theodore Quan[2], BS; David J Diemert[2], MD, FRCP

[1]Department of Medicine, Hospital for Special Surgery, New York, NY, United States
[2]Department of Medicine, The George Washington School of Medicine and Health Sciences, Washington, DC, United States
[3]Department of Psychiatry, Washington University School of Medicine and Barnes-Jewish Hospital, St. Louis, MO, United States
[4]Department of Medicine, American University of Antigua, Coolidge, Antigua and Barbuda
[5]Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, CA, United States
[*]these authors contributed equally

Corresponding Author:
Refka Al-Beyati, MD
Department of Medicine
David Geffen School of Medicine at University of California
10833 Le Conte Ave
Los Angeles, CA
United States
Phone: 1 3133102931
Email: Ralbeyat@gmail.com

Related Article:

Retraction of: https://www.jmir.org/2021/1/e21408/

(J Med Internet Res 2022;24(8):e41544) doi:10.2196/41544

The authors are retracting “Assessing the Dissemination of COVID-19 Articles Across Social Media With Altmetric and PlumX Metrics: Correlational Study” (J Med Internet Res 2021;23(1):e21408) in alignment with COPE guidelines and based on an honest error that invalidates the results.

Authors’ statement:

> *We sincerely regret our honest misunderstanding of the Altmetrics variables as it was discussed in our manuscript. As a team that values honest and ethical research, we are grateful to JMIR for bringing this grave error to our attention and appreciate our fellow peers for holding us to the highest standard of research. We strive to uphold integrity in our research and agree with the decision of the editors of JMIR to retract our original manuscript. We look forward to the opportunity to edit our work so our research accurately reflects the intention behind the Altmetrics variables.*

Submitted 29.07.22; this is a non–peer-reviewed article;accepted 29.07.22; published 10.08.22.

Please cite as:
Tornberg HN, Moezinia C, Wei C, Bernstein SA, Wei C, Al-Beyati R, Quan T, Diemert DJ
Retraction: “Assessing the Dissemination of COVID-19 Articles Across Social Media With Altmetric and PlumX Metrics: Correlational Study”
J Med Internet Res 2022;24(8):e41544
URL: https://www.jmir.org/2022/8/e41544
doi:10.2196/41544
PMID:35947853

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org