Journal of Medical Internet Research

Journal Impact Factorâ ¢ (JIF) (2023): 5.8
Volume 27 (2025) ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Viewpoint

Short Paper

Review

Apps and Digital Resources for Child Neurodevelopment, Mental Health, and Well-Being: Review, Evaluation, and Reflection on Current Resources

Kelsie Ann Boulton[1,2,3], PhD; Makana Hilton[1,2,3], MD; Emilia Sutton[1,2,3], MBrainMindSci; Adam John Guastella[1,2,3], PhD

[1]Children's Hospital Westmead Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]Child Neurodevelopment and Mental Health Team, Brain and Mind Centre, University of Sydney, Sydney, Australia

[3]Clinic for Autism and Neurodevelopmental (CAN) Research, Brain and Mind Centre, University of Sydney, Sydney, Australia

Corresponding Author:
Kelsie Ann Boulton, PhD
Children's Hospital Westmead Clinical School
Faculty of Medicine and Health
University of Sydney
94 Mallett Street
Sydney, 2050
Australia
Phone: 61 0293510984
Email: kelsie.boulton@sydney.edu.au

Abstract

Background: An increase in the prevalence of neurodevelopmental conditions worldwide, alongside resource constraints within clinical services, has led to increased interest in health information technologies, such as apps and digital resources. Digital tools are often viewed as a solution to bridge this divide and to increase supports for families. There is, however, a paucity of research that has evaluated digital health tools, their potential benefits for child neurodevelopment and associated concerns (eg, mental health, well-being), and their benefit for families.

Objective: This study conducted the first review of existing mobile apps and digital resources targeted at supporting the needs of children with developmental concerns or neurodevelopmental conditions.

Methods: We identified 3435 separate resources, of which 112 (43 apps and 69 digital resources) met the criteria. These resources were categorized according to their purpose or target and were then reviewed based on their engagement, information quality, and evidence base using the Adapted Mobile App Rating Scale.

Results: The most common condition of concern targeted by apps and digital resources was autism (19/112, 17% resources), with retrieved resources focusing on supporting challenging behaviors, promoting speech, language, and social development, and providing options for alternative and assistive communication. Other common areas of concern targeted by apps and digital resources included language and communication (16/112, 14.3%) and attention-deficit/hyperactivity disorder (11/112, 9.8%). Results showed that reviewed resources were engaging, with high levels of accessibility and functionality. Resources had various functions, including developmental or behavioral tasks targeted at children, assistive communication support, scheduling support, journaling, and advice, activities, and strategies for parents. The information quality of resources, such as credibility of source and evidence base was, however, mostly low. Apps and digital resources with good credibility and an existing evidence base were largely developed in partnership with research, health, or government institutions, and were rated significantly higher on overall quality compared with apps and digital resources not developed in partnership with such institutions (apps; t_{41}=–4.35, P<.001; digital resources; t_{67}=–4.95, P<.001).

Conclusions: The lack of evidence base across resources means that it is extremely difficult to provide recommendations to families with respect to apps or digital resources that may support their needs. Frameworks for the development of new tools are discussed, highlighting the novel approaches required to demonstrate the efficacy of tools for improving outcomes for children and families. Such a framework requires collaboration with multiple stakeholders (software developers, researchers, regulatory bodies, clinicians, children, and families) and engagement across multiple levels of expertise (app development, implementation, and dissemination within services, policy, and clinical regulations), to harness the potential of digital health for improving outcomes and promoting support in child neurodevelopment, which at this juncture remains largely underdeveloped.

(J Med Internet Res 2025;27:e58693) doi:10.2196/58693

KEYWORDS

digital tools; neurodevelopmental conditions; mental health; digital health; implementation; digital interventions; child neurodevelopment; digital technology; mobile phone

Introduction

In recent years, there has been a reported increase in the prevalence of neurodevelopmental conditions (NDCs) [1,2] and a rising awareness of the importance of early diagnosis and intervention in children with signs of developmental delay [3,4]. This has been matched with a growing number of recommendations and guidelines for multidisciplinary supports provided as early as possible [5-7]. Early identification and evidence-based supports are argued to be crucial for enhancing outcomes in the 1 in 10 children diagnosed with a neurodevelopmental condition [4].

Despite this, there is consistent evidence across nations of the considerable delays and barriers children and families face when seeking assessment [8,9] and support [10,11]. In addition, the wide range of multidisciplinary support needs often experienced by children and caregivers are frequently reported to result in fragmented care, where co-occurring conditions (eg, mental health) are often left unaddressed [12-14]. Caregivers can often report feeling overwhelmed and confused, and comment on the lack of coordination of care and difficulty in obtaining trustworthy, accessible information about supports [15-17].

It is unsurprising, given these resource constraints, that many families turn to health information technologies, such as apps and digital tools, to obtain information to supplement their health care [18,19]. The potential utility of such technologies has been reinforced internationally, with global commissioned reports recommending the integration of digital technologies to facilitate and optimize health care for children and adolescents [20,21]. Digital tools may bridge the divide between existing limited resources and growing demand [22], providing more instant access to supports that may be delivered in a flexible manner. Such tools may also bridge geographical barriers, facilitate access to care for diverse communities, and provide additional opportunities for assessment and feedback to support child and family needs across time [23-25].

Despite enthusiasm for the use of innovative digital solutions [26,27], there is an urgent need to understand the digital tools that currently exist for children with NDCs and their families, the platforms that these tools can be accessed on, and whether these tools are reliable, credible and have an existing evidence base to support their implementation. There is also a need for such evaluations to map existing resources so that supports can be strategically developed where gaps exist. For example, it has previously been shown that, despite the promise of digital tools for mental health, only a minority of apps purported to monitor and manage mental health symptoms have clinically validated evidence of effectiveness [28,29]. More recently, a meta-analysis of the efficacy of smartphone apps for symptoms of depression and anxiety reported overall small effects for symptom amelioration [30]. In addition, while there can be many valuable resources available, there are also many poor sources of information [31-33]. As a result, individuals may instead be influenced by factors such as the star rating of an app, which have been shown to have little relationship with clinical utility [34].

In addition to evaluating digital tools in terms of their evidence base and credibility, it is critical to also evaluate these tools in terms of their engagement, aesthetics, and user quality. Worldwide, there is consistent evidence that poor quality of an app or digital resource is associated with abandonment [35]. In contrast, the design and visual features of apps including an engaging interface, and being intuitive and easy to navigate, all increase engagement and use [36]. To illustrate, digital mental health tools have been shown to have continued uptake as low as 0.1% when rolled out at scale and this low rate is believed to be associated with low appeal and poor app maintenance to promote engagement [37]. This means that the ongoing use of a digital tool, no matter the evidence behind it, is likely to fail without high appearance and usability ratings [23,37]. This may be particularly important for families with NDCs as research shows that personal characteristics moderate the type of apps and digital tools that are used [38]. People with NDCs are more likely to learn from well-structured, easy-to-understand, and clear information, further highlighting the need to evaluate the quality and aesthetics of apps and digital resources [39].

The objective of this study was to review available mobile apps and digital resources that are targeted at supporting the needs of children with developmental concerns or NDCs. The study first sought to conduct a systematic review of all existing mobile apps and digital resources and to categorize them according to their purpose or target. Second, the review then sought to evaluate the information quality, utility, and evidence base for each resource by using the Adapted Mobile App Rating Scale (A-MARS) [40] to demonstrate the quality and evidence base of existing tools.

Methods

Design

This review used a stepwise approach in a similar manner to a systematic review. Our approach included a search strategy, assessment against prespecified eligibility criteria, app, and digital resource selection through an initial screening of all identified apps and digital resources, a full review of included apps and digital resources, data extraction and analysis, and quality assessment of included apps and digital resources using a reliable quality assessment tool, the A-MARS [40,41].

Search Strategy

A search of the main online app stores, iTunes (Apple Inc, Australia) and Google Play (Google Inc, Australia), and a search of digital resources using the Google search engine was conducted from October 2023 to November 2023. To avoid a selection bias from personal Google algorithms, search histories were cleared before each search term as per previous recommendations [42]. Search terms were identified following consultations with researchers with expertise in NDCs, clinicians practicing in neurodevelopmental assessment services, and families with lived experience of NDCs. Search terms were selected based on keywords caregivers of children with developmental concerns or NDCs may use when attempting to access apps and digital resources. The final search terms also demonstrated the best performance in identifying apps and digital resources of interest for this review during preliminary searches. In total, 28 search terms were identified by an expert clinical team based on *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) terms for neurodevelopment and transdiagnostic terms that families might typically use. Search terms included neurodevelopment kids, intellectual disability kids, global developmental delay, communication disorder, language disorder kids, stuttering kids, autism spectrum disorder, autism kids, attention-deficit/hyperactivity disorder (ADHD) kids, ADHD kids, learning disorder kids, motor disorder kids, Tourette’s disorder kids, cerebral palsy kids, dyslexia kids, epilepsy kids, anxious kids, depressed kids, sleep problem kids, attention kids, social problems kids, restricted behaviors kids, repetitive behaviors kids, social skills kids, well-being kids, education kids, emotion kids, cognition kids.

Eligibility Criteria

To be considered for inclusion in this review, apps, and digital resources were required to meet the following inclusion criteria: (1) contain keywords of the search term in the title or description, (2) have a focus on the needs of children with developmental concerns or NDCs (eg, social skills, well-being, academic skills, or mental health), (3) be free of charge, and (4) be in the English language. To ensure the resources contained relevant and applicable information that was catered to families of children with developmental concerns or NDCs, a variety of exclusion criteria was applied. Apps and resources were excluded if they (1) had no actionable information for families to use (eg, purely game, social media site, news article, research paper, clinic website, or home assessment), (2) if they were selling products (eg, books, materials, or therapy sessions), (3) if the information was not relevant to our target population (eg, app or resource targeted only at adults or included information was overly general), or (4) if they were in unsuitable formats for review with the A-MARS (ie, PDF document, podcast, or video).

Selection Process

Three authors (MH, ES, and KB) carried out the app store and search engine searches. To balance feasibility and search comprehensiveness, the first 50 resources for each search term were selected from both the app stores and the search engine [43]. A predesigned Microsoft Excel spreadsheet developed for this review was used to enter information about the apps and digital resources. Information entered included the name of the app or digital resource, the name of the app developer or the URL for digital resources, the app store or stores in which the app was available, and the search terms or terms that identified each app or digital resource.

Following removal of duplicates between search terms, 1 rater (MH) conducted an initial screen for eligibility, by reviewing the title, description, and home page of the app or digital resource. Additional screening was conducted on the remaining apps and digital resources to confirm inclusion in the review. This screening was conducted by 2 raters (MH and ES) and involved the installation of the app and a more detailed review of the digital resource, including a review of written information, pictures, and videos. Any disagreements were discussed with a 3rd rater (KB) and resolved by consensus. Reasons for exclusion were recorded.

Data Extraction

Apps and digital resources identified as eligible in the screening phase were then assessed using the A-MARS [40]. The A-MARS consists of 28 items, developed to assess the quality of mobile apps and e-tools (eg, websites and digital resources). The A-MARS provides an evaluation of app and e-tool quality by grading each app or digital resource on several domains, as described below (Textbox 1). Each item was scored using a 5-point Likert scale (1: inadequate, 2: poor, 3: acceptable, 4: good, and 5: excellent). Supplementary questions within the A-MARS capture information pertaining to the health-related quality of the app or digital resource. Eight scores are calculated for the A-MARS, including the mean score for each domain (ie, engagement, functionality, aesthetics, information, subjective quality, and health-related quality), a mean quality score based on the engagement, functionality, aesthetics and information domains, and a mean total score.

Textbox 1. List of Adapted Mobile App Rating Scale subscales and subscale descriptions.

Subscale and description

Engagement (5 items)

- Is the app or digital resource engaging and interesting for the user? Does it have prompts (eg, send alerts, messages, or reminders)?

Functionality (4 items)

- How does the app or digital resource function, does it have a logical flow and design? Is it easy to navigate?

Aesthetics (3 items)

- How appealing is the app or digital resource in terms of its overall visual appeal, graphic design, and stylistic consistency?

Information (6 items)

- Does the app or digital resource contain high-quality information (eg, text, feedback, measures, or references) from a credible source?

Subjective quality (4 items)

- Subjective quality rating of the app or digital resource.

Health-related quality (6 items)

- Does the app or digital resource provide access to other resources, strategies linked to the target issue, or the option for real-time tracking?

All apps and digital resources were rated using the A-MARS instrument. Three expert raters conducted this review: (1) a senior research fellow with a PhD in Psychology and 10 years' experience working with pediatric NDCs and digital health; (2) a senior research assistant with a Master's degree in Brain and Mind Sciences and 3 years' experience in working with NDCs and digital tools to support children with NDCs and their families; and (3) a research affiliate with Bachelors' degrees in Psychology and Medical and Health Sciences and a Postgraduate Diploma in Psychology. Two of these raters were also people with a diagnosed NDC. All raters reviewed the A-MARS in depth before conducting initial pilot ratings. Two apps and 2 digital resources were initially reviewed independently for training purposes. After independently rating the apps and digital resources, the raters met to compare and review results and to resolve discrepancies in ratings. To reach a consensus, the raters reviewed the A-MARS in depth to improve the alignment of ratings. The remaining apps and digital resources were then rated independently by 2 raters. Based on the methodology described by others [40,41], each rater trialed the apps and digital resources for a minimum of 10 minutes and then independently rated their quality using the A-MARS.

Statistical Analyses

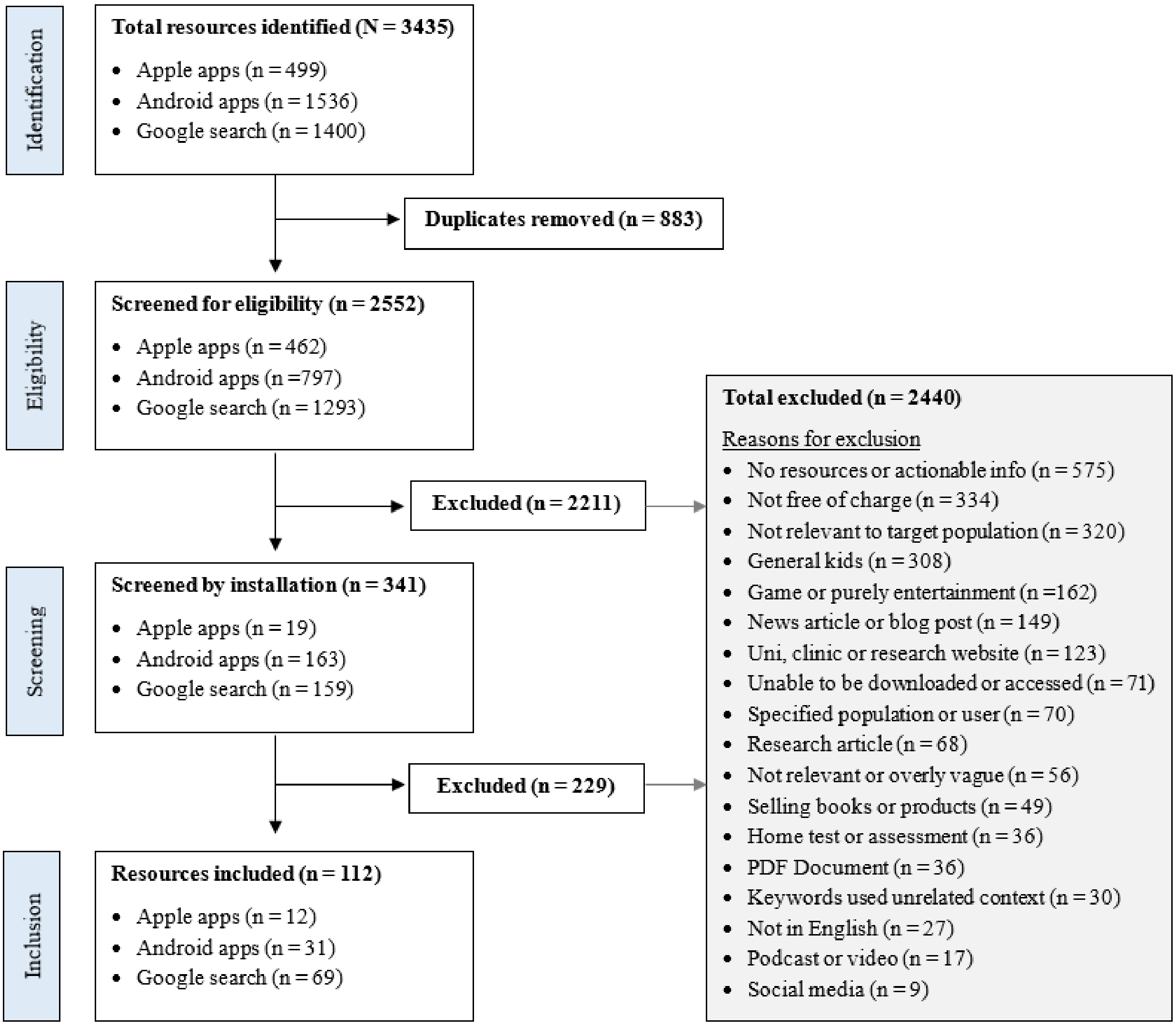
Statistical analyses were conducted using SPSS (version 28; IBM Corp). Descriptive statistics were used to summarize the A-MARS domain scores, the mean quality score, and the mean total score across apps and digital resources. Cronbach α and intraclass correlation coefficients were used to calculate the internal consistency and interrater reliability of the A-MARS domain scores, the quality score, and the total score. As this is the first study to our knowledge that has applied the A-MARS in apps and digital resources targeted at child development and NDCs, 2-tailed independent samples *t* tests were also used to compare mean domain scores, quality scores, and total scores between apps and digital resources. Bonferroni corrections were used for these comparisons and a *P* value of <.006 was considered statistically significant. To understand how the involvement of a reputable and credible organization impacted ratings for apps and digital resources, 2-tailed independent samples *t* tests were used to compare domain scores, quality scores, and total scores for apps and digital resources that had been developed in partnership with a university, health or government institution, relative to those that had not been developed in partnership with such institutions.

Results

Overview

The process of identification and inclusion of apps and digital resources is outlined in Figure 1. A total of 3435 apps and digital resources were initially identified. After the removal of duplicates, 2211 apps and digital resources were excluded following initial screening by title and description, and an additional 229 apps and digital resources were excluded following installation and more detailed screening. The most common reason for exclusion was an absence of actionable information that children and families could use (575/2440, 23.6%), followed by not being free of charge (334/2440, 13.6%), and not being relevant to the target population (320/2440, 13.1%). A total of 12 Apple apps, 31 Android apps, and 69 digital resources were included in the final review of this paper.

Figure 1. Flowchart of selection of included resources.



Features of Included Apps and Digital Resources

Table 1 displays the primary area of focus for retrieved apps and digital resources. The most common conditions that the resources targeted were autism, language, or communication difficulties, and ADHD. Both apps and digital resources were retrieved for most searched terms; however, resources retrieved for anxiety, attention, and developmental delay were mainly digital resources. Comparatively, the resources retrieved for epilepsy, cerebral palsy, and child well-being were mainly apps. Retrieved apps and digital resources had a range of functions, including developmental or behavioral tasks targeted at children, assistive communication support, scheduling support, journaling, and advice, activities, and strategies for parents.

Table 1. Primary focus areas included apps and digital resources.

Topic area	Apps, n (%)	Digital resources, n (%)	Apps and digital resources combined, n (%)
Autism	5 (11.6)	14 (20.3)	19 (17.0)
Language and communication	7 (16.3)	9 (13.0)	16 (14.3)
ADHD[a]	6 (14.0)	5 (7.2)	11 (9.8)
Anxiety	1 (2.3)	9 (13.0)	10 (8.9)
Dyslexia	3 (7.0)	6 (8.7)	9 (8.0)
Mood and depression	2 (4.7)	5 (7.2)	7 (6.3)
Social skills	2 (4.7)	4 (5.8)	6 (5.4)
Emotions	2 (4.7)	4 (5.8)	6 (5.4)
Intellectual disability	3 (7.0)	2 (2.9)	5 (4.5)
Epilepsy	5 (11.6)	0 (0)	5 (4.5)
Attention	1 (2.3)	4 (5.8)	5 (4.5)
Sleep	2 (4.7)	2 (2.9)	4 (3.6)
Cerebral palsy	2 (4.7)	1 (1.4)	3 (2.7)
Developmental delay	0 (0)	3 (4.3)	3 (2.7)
Child well-being	2 (4.7)	0 (0)	2 (1.8)
Tourette	0 (0)	1 (1.4)	1 (0.9)

[a] ADHD: attention deficit hyperactivity disorder.

Overall Assessment of Apps and Digital Resources

Across the 43 apps and 69 digital resources, excellent internal consistency and inter-rater reliability was found across the total and sub-domain A-MARS scales. Detailed item and domain statistics for the reviewed apps are displayed in Tables S1 and S2 in Multimedia Appendix 1. Table 2 displays the mean and median scores of the reviewed apps (n=43) and digital resources (n=69) across the 5 A-MARS domains, as well as for the quality and total scores. The apps, taken together, scored higher than the minimum acceptable score of 3.0 across most A-MARS domains, excluding the health-related quality domain (Table 2). Apps typically received low scores on this domain due to a lack of additional resources available (mean 1.81, SD=1.22), and limited access to further support or related information (mean 1.85, SD=1.24). Furthermore, the digital resources scored higher than 3.0 across all A-MARS domains except the subjective quality domain (Table 2).

Table 2. Adapted Mobile App Rating Scale ratings for apps and digital resources.

Apps				Digital resources			Values	
Domain	Mean (SD)	Median (range)	N (%) scoring ≥3.0	Mean (SD)	Median (range)	N (%) scoring ≥3.0	T-statistic (*df*)	*P* value
Engagement	3.67 (0.74)	3.60 (2.40-4.90)	36 (83.7%)	3.19 (0.54)	3.13 (2.00-4.50)	45 (65.2%)	3.97 (110)	<.001
Functionality	4.13 (0.68)	4.25 (2.25-5.00)	40 (93.0%)	4.09 (0.55)	4.13 (2.50-5.00)	66 (95.7%)	0.37 (110)	.72
Aesthetics	3.73 (0.72)	4.00 (1.84-5.00)	38 (88.4%)	3.66 (0.68)	3.67 (2.00-5.00)	62 (89.9%)	0.56 (110)	.57
Information	3.06 (0.95)	3.13 (1.00-4.75)	27 (62.8%)	3.57 (0.59)	3.67 (2.33-4.75)	61 (88.4%)	-3.16 (110)	.002
Subjective quality	3.21 (0.88)	3.50 (1.50-4.88)	26 (60.5%)	2.89 (0.68)	2.88 (1.38-4.25)	27 (39.1%)	2.04 (110)	.04
Health-related quality	2.58 (0.91)	2.33 (1.00-4.83)	13 (30.2%)	3.34 (0.78)	3.33 (2.00-5.00)	45 (65.2%)	-4.74 (110)	<.001
Quality score	3.40 (0.67)	3.40 (2.17-4.88)	32 (74.4%)	3.62 (0.53)	3.66 (2.38-4.69)	60 (87.0%)	–2.01 (110)	.04
Total score	3.40 (0.64)	3.29 (2.00-4.72)	31 (72.1%)	3.48 (0.55)	3.50 (2.32-4.67)	54 (78.3%)	–0.73 (110)	.47

Both apps and digital resources scored higher than 3.0 across quality and total scores. Apps and digital resources scored high on the functionality domain, with consistently high average scores across subcriteria. Of note, included apps scored relatively low in the information quality domain primarily due to questionable source credibility (mean 2.52, SD 0.79). Furthermore, a minority of apps (3/43, 7%) and no digital resources met the criteria for a verifiable evidence base as outlined by the A-MARS (has been trialed or tested and published in scientific literature). Independent sample *t* tests revealed statistically significant differences in mean scores between apps and digital resources on the engagement, information, and health-related quality domains (*P*<.006). While apps had higher ratings relative to digital resources for engagement, digital resources had higher ratings relative to apps on information and health-related quality domains.

Overall, apps and digital resources that had been developed in partnership with a health, university, or government institution, or where such an institution was listed as a partner, were rated more highly across domains of engagement, functionality, information, subjective quality, and health-related quality, as well as on overall quality and total scores. As shown in Tables 3 and 4, these effects were moderate to large for most comparisons. In total, 19% (8/43) of apps were developed in partnership with a health, university, or government institution, and 48% (33/69) of digital resources were developed in partnership with these institutions.

Table 3. Adapted Mobile App Rating Scale ratings for apps split by partnerships with health, university, or government institutions.

Domain	Apps				
	Institution partnership (n=8), mean (SD)	No institution partnership (n=35), mean (SD)	Effect size	T-statistic (*df*)	*P* value
Engagement	4.23 (0.75)	3.54 (0.68)	0.69	–2.53 (41)	.02
Functionality	4.64 (0.30)	4.01 (0.69)	0.64	–2.53 (41)	.02
Aesthetics	4.15 (0.62)	3.64 (0.73)	0.70	–1.85 (41)	.07
Information	4.30 (0.42)	2.78 (0.80)	0.75	–7.52 (41)	<.001
Subjective quality	4.05 (0.58)	3.02 (0.83)	0.80	–3.29 (41)	.002
Health-related quality	3.53 (1.33)	2.36 (0.63)	0.79	–2.42 (41)	.04
Quality score	4.13 (0.58)	3.23 (0.57)	0.57	–4.04 (41)	<.001
Total score	4.14 (0.51)	3.23 (0.54)	0.54	–4.35 (41)	<.001

Table 4. Adapted Mobile App Rating Scale ratings for digital resources split by partnerships with health, university, or government institutions.

Domain	Digital resources				
	Institution partnership (n=33), mean (SD)	No institution partnership (n=36), mean (SD)	Effect size	T-statistic (*df*)	*P* value
Engagement	3.47 (0.51)	2.93 (0.44)	0.47	–4.76 (67)	<.001
Functionality	4.36 (0.46)	3.83 (0.51)	0.48	–4.58 (67)	<.001
Aesthetics	3.98 (0.62)	3.36 (0.59)	0.6	–4.24 (67)	<.001
Information	3.87 (0.53)	3.30 (0.50)	0.51	–4.60 (67)	<.001
Subjective quality	3.13 (0.70)	2.67 (0.58)	0.64	–2.95 (67)	.004
Health-related quality	3.69 (0.76)	3.02 (0.65)	0.7	–3.95 (67)	<.001
Quality score	3.92 (0.44)	3.35 (0.44)	0.45	–5.18 (67)	<.001
Total score	3.77 (0.50)	3.21 (0.45)	0.47	–4.95 (67)	<.001

Individual Assessment of Apps and Digital Resources

Figures 2 and 3 display scores of the top 5 and bottom 5 apps in terms of their A-MARS ratings. These radar charts show how individual apps and digital resources scored on each A-MARS domain, as well as on the overall quality and total scores. Showing the difference between the top 5 and bottom 5 highlights the disparity between apps and digital resources on A-MARS ratings. In general, the top 5 apps and digital resources were characterized by high levels of factually correct information, relevant resources, good visuals, and engaging and user-friendly interfaces. Highly rated apps and digital resources were developed in collaboration with research, health, or government institutions. In contrast, the bottom 5 apps and digital resources were characterized by less involvement with a reputable institution, defective or inactive links, difficult or cumbersome interfaces, and a lack of information relevant to the purported concern or condition.

Figure 2. Radar chart of the top 5 and bottom 5 ranked apps. ADHD: attention-deficit/hyperactivity disorder.

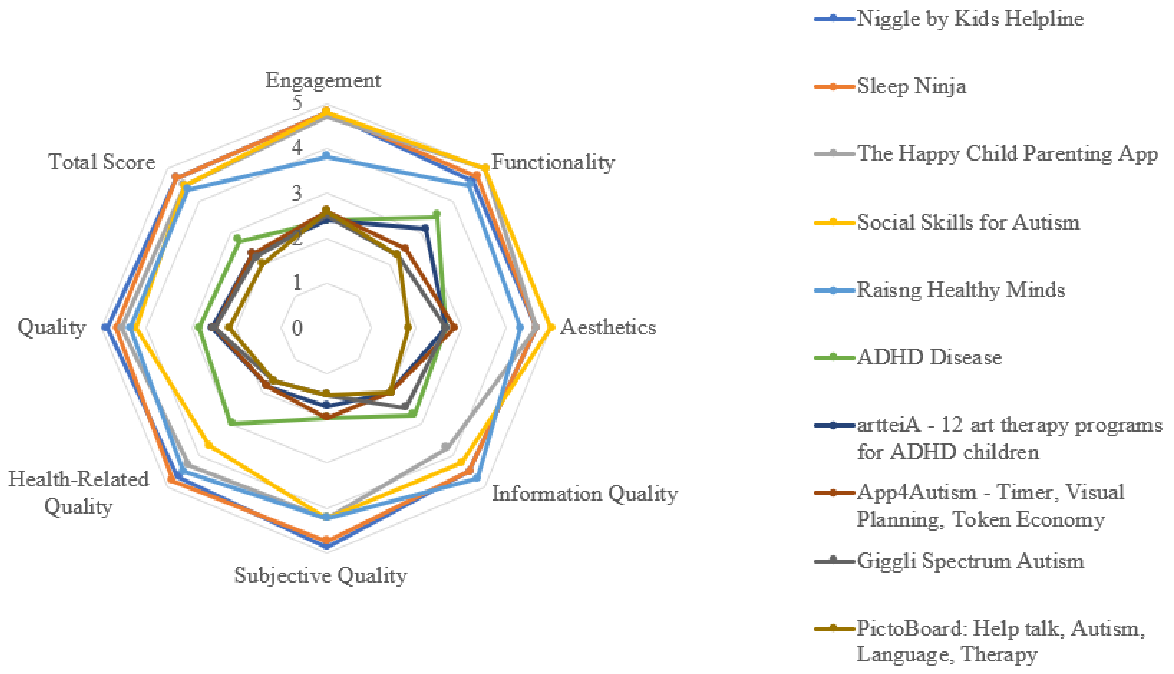
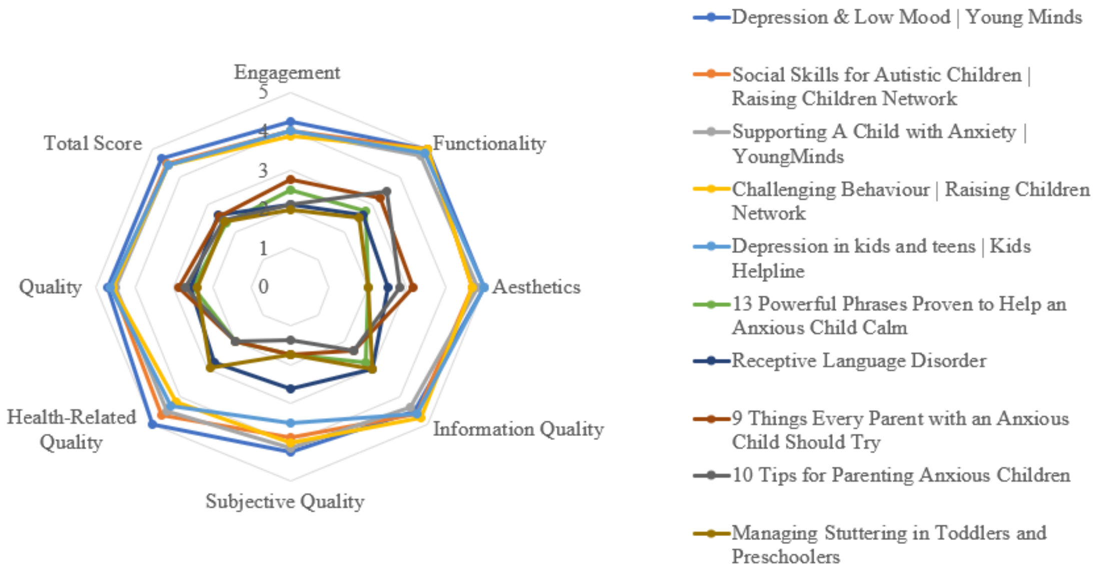


Figure 3. Radar chart of the top 5 and bottom 5 ranked digital resources.



Discussion

This study reviewed existing apps and digital resources that are targeted at the needs of children with developmental concerns or NDCs. To our knowledge, this is the first study that has evaluated apps and digital resources for child neurodevelopment and mental health broadly. From the initial 3435 resources identified, the study reviewed 112 resources in total (43 apps and 69 digital resources). Developed for children with NDCs and their families, the most common resources focus on supports for autism, language or communication difficulties, and ADHD. Most apps and digital resources exceeded the minimum acceptable score of 3.0, suggesting that retrieved resources were engaging, with high levels of accessibility and functionality. However, the results also showed items relating to information quality, such as credibility of source and evidence base, were generally low. Only 7% of apps or digital resources had an established evidence base. Interestingly, mobile apps and digital resources that were developed in partnership with a health, university, or government institution were rated significantly higher across most domains in comparison to apps and resources that were not developed in partnership with such an organization. Unfortunately, the review showed there was almost no evidence for their efficacy or effectiveness in promoting supportive behaviors or improving outcomes. As a result, there is a significant existing divide between the availability of digital tools that can be integrated into family systems to support needs and evidence that warrants recommendations from clinical teams. While the findings of the review provide an evaluation and collation of existing tools, this conclusion highlights the urgent need to develop new frameworks to test efficacy for improving outcomes and promoting supportive behaviors.

These results show that most apps and digital resources alike met the minimum acceptable criteria in terms of quality (eg, highly engaging and functional). However, we did observe some differences as a function of resource type (ie, apps or digital resources). We found that apps displayed higher ratings on the domain of engagement, suggesting that the reviewed apps were more interesting, customizable, and entertaining, compared to digital resources. In contrast, reviewed digital resources had higher ratings on domains of information and health-related quality compared with apps. This indicates that the quality and credibility of information, as well as any additional resources and strategies provided, was higher in the reviewed digital resources.

Apps and digital resources were rated highly in terms of their engagement and functionality, but somewhat lower on information quality, particularly with respect to credibility of source and existing evidence base. We observed that apps and digital resources that had been developed by or in collaboration with a research, health, or government institutions were rated significantly more highly on information quality and credibility. This aligns with studies that have used evaluation tools like the Mobile App Rating Scale and A-MARS in other fields [40,44]. However, we note that very few apps or digital resources had a scientific evidence base to support their use, which points to the dearth of evidence evaluating the effectiveness of apps and digital resources [40,45]. This lack of evidence base may relate to the iterative, fast-paced nature of technological development and the linear, incremental approach to clinical science and research-based interventions. With established lags of up to 17 years for research to be translated into clinical practice, software developers may opt to move to large-scale dissemination of their product before developing a rigorous evidence base [24]. However, the consequence of this is an abundance of apps and digital resources that have little to no evidence base and may not be beneficial for consumers [27]. To resolve this issue, a novel framework with investment prioritized for development is critical, whereby the effectiveness of apps and digital resources can be rigorously evaluated in a streamlined and timely manner. Such a framework needs to incorporate the swift evaluation of real-world implementation alongside both the evaluation of resource development and the evaluation of efficacy. It is unlikely to be achieved by applying standard academic approaches or randomized controlled trial methods and may need to incorporate adaptive research designs within implementation science methodologies. Integrated research enabling platforms [46-48] that permit rapid implementation of clinical trials for multiple apps and resources within existing ethics and governance frameworks are also required for swift and sustained translation.

Our results show that the most common condition of concern targeted by apps and digital resources was autism (19/112, 17.0% of all apps and digital resources), with retrieved resources focusing on supporting challenging behaviors, promoting speech, language, and social development, and providing options for alternative and assistive communication. Given the growing prevalence of conditions like autism [1], an increase in the number of health information technologies focused on these conditions is relevant and has the potential to provide children and families with a variety of support tools. However, the proportion of resources targeted at comorbid conditions, such as mental health, was lower (10/112, 8.9% targeted at anxiety and 7/112, 6.3% targeted at mood or depression). Mental ill health is one of the fastest-growing health problems in children and adolescents, particularly in the wake of the COVID-19 pandemic [49]. Moreover, children with existing developmental delays and NDCs are more likely to experience mental health conditions relative to their neurotypical peers [12]. Together these findings point to a gap in existing digital resources for children with co-occurring developmental and mental health concerns.

It is unlikely that a single resource will meet the varied needs of children and families. The application and evaluation of digital resources to uplift support in the community may require integrated approaches targeting different needs at different stages of child development and different stages of support seeking. Recent arguments have also been made for the integration of digital navigators that can further facilitate the engagement and implementation of digital health tools within multidisciplinary care teams and standard clinical care [40,50]. With the increasing frequency of digital navigators in other disciplines, such as mental health [51,52], health systems providing child development and assessment services can benefit from adopting a similar approach, to support children and families in accessing reliable, credible, and efficacious digital tools across their health care journey.

While this study provided a first review of apps and digital resources targeted at child neurodevelopment, mental health, and well-being, we note certain limitations. The evaluation of apps and digital resources can be made difficult due to the growing number of app evaluation frameworks [50]. While the evaluation tool we selected, the A-MARS, has sound reliability [53], has been used across disciplines [54,55], and has been adapted for use specifically with digital resources [40], we note that there are challenges common across evaluation tools [50]. While it was deemed the most suitable evaluation tool for this review, we note specific limitations of the A-MARS, such as potential subjectivity in ratings, a relative lack of focus on the clinical effectiveness of health information, and a lack of validation across cultures and languages, which may warrant further evaluation of apps and digital resources with additional evaluation tools. However, there is currently no consensus for the evaluation and regulation of apps and digital resources, making it difficult to conduct standardized evaluations that can provide clear guidance and recommendations to patients and clinicians. We also note that we restricted our search to apps and digital resources that were available free of charge. While this may have excluded potentially valuable paid resources, we argue that reviewing freely available resources increases the value of this study for use in the broader community and serves to increase equity in access to supports and resources. In a similar vein, we restricted our search to English language apps and digital resources, potentially limiting the applicability of our findings for non–English speaking populations. However, we have detailed a process that future researchers can use to identify and evaluate apps and digital resources in their region of interest, thereby facilitating evaluations in different countries and contexts. Further, we included a large number of search terms, which resulted in many retrieved resources. As a team we discussed restricting the search strategy; however, it was noted that children with developmental concerns or NDCs often experience multiple transdiagnostic concerns, with caregivers likely to search for support across many of these domains. As such, our more extensive search strategy was considered appropriate in order to capture the broad needs and concerns experienced by this patient population.

A crucial next step for this field is to understand how accessible, reliable, evidence-based health information technologies that meet the needs of children, families, and health care providers can be developed and disseminated across clinical settings. This requires collaboration with multiple stakeholders (app developers, researchers, regulatory bodies, clinicians, children, and families) as well as a reconceptualization of how to implement and recommend health information technologies across clinical services for maximum uptake and maintained engagement. At the development level, for example, partnerships between industry, research, and end-users are key to drive the timely co-production of evidence-based, accessible digital tools. Furthermore, given findings from digital mental health [56], a focus on developing tools that include gamification features may be key to optimizing uptake and engagement by children and families. At the implementation and dissemination level, existing models of care may need to be rethought to optimally engage recipients (ie, patients and clinicians) and encourage the integration of digital tools and solutions into clinical practice. Meanwhile, at the policy level, clinical regulation and recommendations require flexibility to enable innovative digital solutions to be adopted and embedded within health services and systems. While complex and multifaceted, such a framework would enable the development of evidence-based digital tools that are primed for accessibility and engagement with children and families and would facilitate the embedding of digital health into clinical service settings in a sustainable manner, uplifting the capacity of services to provide access to reliable, evidence-based digital tools that are likely to be used to support children and families.

In the current technology-driven world, apps and digital resources are being increasingly used and promoted as a source of information or an adjunct to support and therapy. In this review, we found that a large number of apps and digital resources targeted at child neurodevelopment, mental health, and well-being are engaging and functional, but lack an adequate evidence base. Those apps and digital resources that did show good credibility and an evidence base were largely developed in partnership with research, health, or government institutions. As a result, there is a pressing need to recognize, value, and promote this type of collaboration when developing new digital tools, and to also develop a framework where new digital tools can be rigorously evaluated in a timely manner, to promote swift translation from evidence-based research and development into practice. Clinicians and health care professionals can then be supported to recommend reliable, evidence-based apps and digital resources to children and families based on their individual needs, providing an ideal opportunity to evaluate the effectiveness of these tools for enhancing outcomes.

Acknowledgments

This work was supported by a Research Development Grant from the Brain and Mind Centre and an Emerging Researcher Grant from The Research Foundation of Cerebral Palsy Alliance (ERG01921).

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KAB and AJG were involved in the study conceptualization and supervision. KAB, MH, and ES were involved in data extraction, data curation, and formal analysis. KAB and AJG wrote the original draft of the manuscript. All authors provided review and feedback on the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interrater reliability and internal consistency of the A-MARS (Adapted Mobile App Rating Scale) items and domain scores.
[DOCX File , 21 KB - jmir_v27i1e58693_app1.docx]

Multimedia Appendix 2
PRISMA Checklist.
[PDF File (Adobe PDF File), 136 KB - jmir_v27i1e58693_app2.pdf]

References

1. Maenner MJ, Shaw KA, Bakian AV, Bilder DA, Durkin MS, Esler A, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2018. MMWR Surveill Summ 2021;70(11):1-16 [FREE Full text] [doi: 10.15585/mmwr.ss7011a1] [Medline: 34855725]
2. Zablotsky B, Black LI, Maenner MJ, Schieve LA, Danielson ML, Bitsko RH, et al. Prevalence and trends of developmental disabilities among children in the United States: 2009-2017. Pediatrics 2019;144(4):e20190811 [FREE Full text] [doi: 10.1542/peds.2019-0811] [Medline: 31558576]
3. Clark MLE, Vinen Z, Barbaro J, Dissanayake C. School age outcomes of children diagnosed early and later with autism spectrum disorder. J Autism Dev Disord 2018;48(1):92-102. [doi: 10.1007/s10803-017-3279-x] [Medline: 28905160]
4. Hadders-Algra M. Early diagnostics and early intervention in neurodevelopmental disorders-age-dependent challenges and opportunities. J Clin Med 2021;10(4):861 [FREE Full text] [doi: 10.3390/jcm10040861] [Medline: 33669727]
5. AADPA. ADHD guideline development group: Australian evidence-based clinical practice guideline for attention deficit hyperactivity. 2022. URL: https://adhdguideline.aadpa.com.au/ [accessed 2024-10-25]
6. NICE. National Institute for Health and Care Excellence: Guidelines. In: Attention Deficit Hyperactivity Disorder: Diagnosis and Management. England, UK: National Institute for Health and Care Excellence (NICE) Copyright © NICE 2018; 2019.
7. Whitehouse AJO, Evans K, Eapen V, Wray J. A national guideline for the assessment and diagnosis of autism spectrum disorders in Australia. In: Summary and Recommendations. Australia: Cooperative Research Centre for Living with Autism; 2018.
8. Straub L, Bateman BT, Hernandez-Diaz S, York C, Lester B, Wisner KL, et al. Neurodevelopmental disorders among publicly or privately insured children in the United States. JAMA Psychiatry 2022;79(3):232-242 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.3815] [Medline: 34985527]
9. Boulton KA, Hodge MA, Jewell A, Ong N, Silove N, Guastella AJ. Diagnostic delay in children with neurodevelopmental conditions attending a publicly funded developmental assessment service: findings from the Sydney Child Neurodevelopment Research Registry. BMJ Open 2023;13(2):e069500 [FREE Full text] [doi: 10.1136/bmjopen-2022-069500] [Medline: 36725093]
10. Sapiets SJ, Hastings RP, Totsika V. Predictors of access to early support in families of children with suspected or diagnosed developmental disabilities in the United Kingdom. J Autism Dev Disord 2024;54(4):1628-1641 [FREE Full text] [doi: 10.1007/s10803-023-05996-7] [Medline: 37142908]
11. Boulton KA, Hodge A, Levu K, Ong N, Silove N, Guastella AJ. Access and barriers to supports for children and caregivers attending public child developmental assessment services: findings from the Sydney Child Neurodevelopment Research Registry. Autism Res 2024;17(3):555-567. [doi: 10.1002/aur.3064] [Medline: 38009266]
12. Boulton KA, Guastella AJ, Hodge MA, Demetriou EA, Ong N, Silove N. Mental health concerns in children with neurodevelopmental conditions attending a developmental assessment service. J Affect Disord 2023;335:264-272 [FREE Full text] [doi: 10.1016/j.jad.2023.04.098] [Medline: 37119867]
13. Thapar A, Cooper M, Rutter M. Neurodevelopmental disorders. Lancet Psychiatry 2017;4(4):339-346. [doi: 10.1016/S2215-0366(16)30376-5] [Medline: 27979720]
14. Kerns CM, Kendall PC, Zickgraf H, Franklin ME, Miller J, Herrington J. Not to be overshadowed or overlooked: functional impairments associated with comorbid anxiety disorders in youth with ASD. Behav Ther 2015;46(1):29-39. [doi: 10.1016/j.beth.2014.03.005] [Medline: 25526833]
15. Carlsson E, Miniscalco C, Kadesjö B, Laakso K. Negotiating knowledge: parents' experience of the neuropsychiatric diagnostic process for children with autism. Int J Lang Commun Disord 2016;51(3):328-338. [doi: 10.1111/1460-6984.12210] [Medline: 26833425]
16. Crane L, Batty R, Adeyinka H, Goddard L, Henry LA, Hill EL. Autism diagnosis in the United Kingdom: perspectives of autistic adults, parents and professionals. J Autism Dev Disord 2018;48(11):3761-3772 [FREE Full text] [doi: 10.1007/s10803-018-3639-1] [Medline: 29948530]
17. Crane L, Chester JW, Goddard L, Henry LA, Hill E. Experiences of autism diagnosis: a survey of over 1000 parents in the United Kingdom. Autism 2016;20(2):153-162. [doi: 10.1177/1362361315573636] [Medline: 25810370]
18. Valentine AZ, Brown BJ, Groom MJ, Young E, Hollis C, Hall CL. A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: a map of the current evidence. Clin Psychol Rev 2020;80:101870 [FREE Full text] [doi: 10.1016/j.cpr.2020.101870] [Medline: 32712216]
19. Gibson AN, Kaplan S, Vardell E. A survey of information source preferences of parents of individuals with autism spectrum disorder. J Autism Dev Disord 2017;47(7):2189-2204. [doi: 10.1007/s10803-017-3127-z] [Medline: 28451948]
20. Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, Early Childhood Development Interventions Review Group, for the Lancet Early Childhood Development Series Steering Committee. Nurturing care: promoting early childhood development. Lancet 2017;389(10064):91-102. [doi: 10.1016/S0140-6736(16)31390-3] [Medline: 27717615]

21. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, Secretariat of the LancetFinancial Times Commission. The Lancet and Financial Times commission on governing health futures 2030: growing up in a digital world. Lancet 2021;398(10312):1727-1776. [doi: 10.1016/S0140-6736(21)01824-9] [Medline: 34706260]
22. Alon N, Perret S, Torous J. Working towards a ready to implement digital literacy program. Mhealth 2023;9:32 [FREE Full text] [doi: 10.21037/mhealth-23-13] [Medline: 38023777]
23. 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;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
24. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
25. Harvey PD, Depp CA, Rizzo AA, Strauss GP, Spelber D, Carpenter LL, et al. Technology and mental health: state of the art for assessment and treatment. Am J Psychiatry 2022;179(12):897-914. [doi: 10.1176/appi.ajp.21121254] [Medline: 36200275]
26. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
27. Stewart E, Milton A, Yee HF, Song MJ, Roberts A, Davenport T, et al. eHealth tools that assess and track health and well-being in children and young people: systematic review. J Med Internet Res 2022;24(5):e26015 [FREE Full text] [doi: 10.2196/26015] [Medline: 35550285]
28. 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;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
29. 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;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
30. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. World Psychiatry 2024;23(1):139-149 [FREE Full text] [doi: 10.1002/wps.21183] [Medline: 38214614]
31. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
32. Zervogianni V, Fletcher-Watson S, Herrera G, Goodwin M, Pérez-Fuster P, Brosnan M, et al. A framework of evidence-based practice for digital support, co-developed with and for the autism community. Autism 2020;24(6):1411-1422 [FREE Full text] [doi: 10.1177/1362361319898331] [Medline: 32026715]
33. Grant N, Rodger S, Hoffmann T. Evaluation of autism-related health information on the web. J Appl Res Intellect Disabil 2015;28(4):276-282. [doi: 10.1111/jar.12127] [Medline: 25224690]
34. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
35. Lim SL, Bentley PJ, Kanakam N, Ishikawa F, Honiden S. Investigating country differences in mobile app user behavior and challenges for software engineering. IEEE Trans Software Eng 2015;41(1):40-64. [doi: 10.1109/TSE.2014.2360674]
36. Hoffman L, Benedetto E, Huang H, Grossman E, Kaluma D, Mann Z, et al. Augmenting mental health in primary care: a 1-year study of deploying smartphone apps in a multi-site primary care/behavioral health integration program. Front Psychiatry 2019;10:94 [FREE Full text] [doi: 10.3389/fpsyt.2019.00094] [Medline: 30873053]
37. Torous J. A path towards progress: lessons from the hard things about digital mental health. World Psychiatry 2022;21(3):419-420 [FREE Full text] [doi: 10.1002/wps.21003] [Medline: 36073703]
38. Chittaranjan G, Blom J, Gatica-Perez D. Mining large-scale smartphone data for personality studies. Pers Ubiquit Comput 2011;17(3):433-450. [doi: 10.1007/s00779-011-0490-1]
39. Iovannone R, Dunlap G, Huber H, Kincaid D. Effective educational practices for students with autism spectrum disorders. Focus Autism Other Dev Disabl 2003;18(3):150-165. [doi: 10.1177/10883576030180030301]
40. Roberts AE, Davenport TA, Wong T, Moon H, Hickie IB, LaMonica HM. Evaluating the quality and safety of health-related apps and e-tools: adapting the mobile app rating scale and developing a quality assurance protocol. Internet Interv 2021;24:100379 [FREE Full text] [doi: 10.1016/j.invent.2021.100379] [Medline: 33777705]
41. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
42. Piasecki J, Waligora M, Dranseika V. Google search as an additional source in systematic reviews. Sci Eng Ethics 2018;24(2):809-810 [FREE Full text] [doi: 10.1007/s11948-017-0010-4] [Medline: 29249022]

43. D'Urzo KA, Man KE, Bassett-Gunter RL, Latimer-Cheung AE, Tomasone JR. Identifying "real-world" initiatives for knowledge translation tools: a case study of community-based physical activity programs for persons with physical disability in Canada. Transl Behav Med 2019;9(4):797-809. [doi: 10.1093/tbm/iby088] [Medline: 30184185]
44. Siddique AB, Krebs M, Alvarez S, Greenspan I, Patel A, Kinsolving J, et al. Mobile apps for the care management of chronic kidney and end-stage renal diseases: systematic search in app stores and evaluation. JMIR Mhealth Uhealth 2019;7(9):e12604 [FREE Full text] [doi: 10.2196/12604] [Medline: 31486408]
45. Camacho E, Cohen A, Torous J. Assessment of mental health services available through smartphone apps. JAMA Netw Open 2022;5(12):e2248784 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.48784] [Medline: 36576737]
46. Boulton KA, Coghill D, Silove N, Pellicano E, Whitehouse AJO, Bellgrove MA, et al. A national harmonised data collection network for neurodevelopmental disorders: a transdiagnostic assessment protocol for neurodevelopment, mental health, functioning and well-being. JCPP Adv 2021;1(4):e12048 [FREE Full text] [doi: 10.1002/jcv2.12048] [Medline: 37431407]
47. Frøen JF, Myhre SL, Frost MJ, Chou D, Mehl G, Say L, et al. eRegistries: electronic registries for maternal and child health. BMC Pregnancy Childbirth 2016;16:11 [FREE Full text] [doi: 10.1186/s12884-016-0801-7] [Medline: 26791790]
48. Robotham D, Waterman S, Oduola S, Papoulias C, Craig T, Wykes T. Facilitating mental health research for patients, clinicians and researchers: a mixed-method study. BMJ Open 2016;6(8):e011127 [FREE Full text] [doi: 10.1136/bmjopen-2016-011127] [Medline: 27503859]
49. Samji H, Wu J, Ladak A, Vossen C, Stewart E, Dove N, et al. Review: mental health impacts of the COVID-19 pandemic on children and youth - a systematic review. Child Adolesc Ment Health 2022;27(2):173-189 [FREE Full text] [doi: 10.1111/camh.12501] [Medline: 34455683]
50. Alon N, Torous J. Current challenges for evaluating mobile health applications. J Am Med Inform Assoc 2023;30(3):617-624 [FREE Full text] [doi: 10.1093/jamia/ocac244] [Medline: 36484621]
51. Carpenter-Song E, Acquilano SC, Noel V, Al-Abdulmunem M, Torous J, Drake RE. Individualized intervention to support mental health recovery through implementation of digital tools into clinical care: feasibility study. Community Ment Health J 2022;58(1):99-110 [FREE Full text] [doi: 10.1007/s10597-021-00798-6] [Medline: 33611684]
52. Perret S, Alon N, Carpenter-Song E, Myrick K, Thompson K, Li S, et al. Standardising the role of a digital navigator in behavioural health: a systematic review. Lancet Digit Health 2023;5(12):e925-e932 [FREE Full text] [doi: 10.1016/S2589-7500(23)00152-8] [Medline: 38000876]
53. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the mobile application rating scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
54. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
55. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
56. Fleming T, Poppelaars M, Thabrew H. The role of gamification in digital mental health. World Psychiatry 2023;22(1):46-47 [FREE Full text] [doi: 10.1002/wps.21041] [Medline: 36640378]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
A-MARS: Adapted Mobile App Rating Scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
NDC: neurodevelopmental condition

Edited by N Cahill; submitted 22.03.24; peer-reviewed by RM Garad, J Mistry; comments to author 20.04.24; revised version received 29.04.24; accepted 16.09.24; published 01.01.25.

Please cite as:
Boulton KA, Hilton M, Sutton E, Guastella AJ
Apps and Digital Resources for Child Neurodevelopment, Mental Health, and Well-Being: Review, Evaluation, and Reflection on Current Resources
J Med Internet Res 2025;27:e58693
URL: https://www.jmir.org/2025/1/e58693
doi:10.2196/58693
PMID:

Review

Machine Learning Approaches in High Myopia: Systematic Review and Meta-Analysis

Huiyi Zuo[1], MM; Baoyu Huang[1], MM; Jian He[1], MD; Liying Fang[1], MM; Minli Huang[1], MD

[1] Ophthalmology Department, First Affiliated Hospital of GuangXi Medical University, Nanning, China

Corresponding Author:
Minli Huang, MD
Ophthalmology Department
First Affiliated Hospital of GuangXi Medical University
No 6 Shuangyong Road, Nanning, Guangxi
Nanning, 530000
China
Phone: 86 0771 5356507
Email: 420306@sr.gxmu.edu.cn

Abstract

Background: In recent years, with the rapid development of machine learning (ML), it has gained widespread attention from researchers in clinical practice. ML models appear to demonstrate promising accuracy in the diagnosis of complex diseases, as well as in predicting disease progression and prognosis. Some studies have applied it to ophthalmology, primarily for the diagnosis of pathologic myopia and high myopia-associated glaucoma, as well as for predicting the progression of high myopia. ML-based detection still requires evidence-based validation to prove its accuracy and feasibility.

Objective: This study aims to discern the performance of ML methods in detecting high myopia and pathologic myopia in clinical practice, thereby providing evidence-based support for the future development and refinement of intelligent diagnostic or predictive tools.

Methods: PubMed, Cochrane, Embase, and Web of Science were thoroughly retrieved up to September 3, 2023. The prediction model risk of bias assessment tool was leveraged to appraise the risk of bias in the eligible studies. The meta-analysis was implemented using a bivariate mixed-effects model. In the validation set, subgroup analyses were conducted based on the ML target events (diagnosis and prediction of high myopia and diagnosis of pathological myopia and high myopia-associated glaucoma) and modeling methods.

Results: This study ultimately included 45 studies, of which 32 were used for quantitative meta-analysis. The meta-analysis results unveiled that for the diagnosis of pathologic myopia, the summary receiver operating characteristic (SROC), sensitivity, and specificity of ML were 0.97 (95% CI 0.95-0.98), 0.91 (95% CI 0.89-0.92), and 0.95 (95% CI 0.94-0.97), respectively. Specifically, deep learning (DL) showed an SROC of 0.97 (95% CI 0.95-0.98), sensitivity of 0.92 (95% CI 0.90-0.93), and specificity of 0.96 (95% CI 0.95-0.97), while conventional ML (non-DL) showed an SROC of 0.86 (95% CI 0.75-0.92), sensitivity of 0.77 (95% CI 0.69-0.84), and specificity of 0.85 (95% CI 0.75-0.92). For the diagnosis and prediction of high myopia, the SROC, sensitivity, and specificity of ML were 0.98 (95% CI 0.96-0.99), 0.94 (95% CI 0.90-0.96), and 0.94 (95% CI 0.88-0.97), respectively. For the diagnosis of high myopia-associated glaucoma, the SROC, sensitivity, and specificity of ML were 0.96 (95% CI 0.94-0.97), 0.92 (95% CI 0.85-0.96), and 0.88 (95% CI 0.67-0.96), respectively.

Conclusions: ML demonstrated highly promising accuracy in diagnosing high myopia and pathologic myopia. Moreover, based on the limited evidence available, we also found that ML appeared to have favorable accuracy in predicting the risk of developing high myopia in the future. DL can be used as a potential method for intelligent image processing and intelligent recognition, and intelligent examination tools can be developed in subsequent research to provide help for areas where medical resources are scarce.

Trial Registration: PROSPERO CRD42023470820; https://tinyurl.com/2xexp738

(J Med Internet Res 2025;27:e57644) doi:10.2196/57644

KEYWORDS

high myopia; pathological myopia; high myopia-associated glaucoma; machine learning; deep learning

Introduction

Myopia is currently widely regarded as a significant public health issue, leading to substantial vision loss and serving as a risk factor for a range of other serious ocular diseases. It is estimated that by 2050, 4.758 billion people (49.8% of the world population) and 938 million people (9.8% of the world population) will suffer from myopia and high myopia, respectively [1]. A recent meta-analysis study proposed that the global economic burden due to productivity losses from uncorrected myopia and myopic macular degeneration is estimated to reach US $250 billion [2]. Therefore, the prevention of high myopia as well as the diagnosis and treatment of pathological myopia remain a formidable societal challenge.

High myopia is defined as the spherical equivalent ≤–6.0 diopter [3] when the accommodation of the eye is relaxed. However, increased severity of myopia and elongation of the eye's axial length could alter the posterior segment structures, causing posterior scleral staphyloma, myopic macular degeneration, and optic neuropathy related to high myopia, potentially leading to the loss of best-corrected visual acuity [3]. High myopia-related fundus lesions stand as an important contributing factor to blindness across the world as well as in China [4]. The detection of high myopia hinges primarily on artificial auxiliary techniques, like refraction detection, fundus examination, measurement of axial length, and fundus photography. Nevertheless, manual examination and analysis by ophthalmologists are still essential, necessitating a significant investment of time and effort [5]. Additionally, in regions with limited medical resources, the shortage of ophthalmologists and medical equipment impedes the early and accurate identification of high-risk patients with high myopia, resulting in missed opportunities for optimal treatment. Therefore, forecasting the risk of high myopia and precisely diagnosing pathological myopia are currently major research focus.

With the rapid advances in computing technology and the ongoing refinement of statistical theory, machine learning (ML) has gradually been promoted and applied in clinical practice. For instance, ML can not only improve image quality, reduce misregistration, and simulate attenuation correction imaging in core cardiology [6], but also be used for cancer screening (detection of lesions), characterization and grading of tumors, and prognosis prediction, thus facilitating clinical decision-making [7]. Since fundus images are noncontact, noninvasive, low-cost, easily accessible, and easy to process, ML has been extensively used to diagnose common retinal diseases, including diabetic retinopathy [8-10], macular degeneration [10], and glaucoma [11-13]. ML has been applied to various image-processing tasks. Novel techniques for analyzing fundus images of high myopia and pathological myopia are continuously emerging [14,15]. However, the accuracy of these ML detections has not been systematically studied. Consequently, the present study was executed to comprehensively describe the accuracy of ML in detecting different degrees of lesions in high myopia, furnishing an evidence-based reference for subsequent lesion management.

Methods

Study Registration

This study was implemented as per the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines and prospectively registered with PROSPERO (ID: CRD42023470820). The PRISMA checklist is available in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

We established detailed inclusion and exclusion criteria for this systematic review. To enhance visualization, these criteria are presented in tabular form (Textbox 1).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Study type: (1) case-control, cohort, nested case-control, and case-cohort studies and (2) studies reported in English.
- Machine learning (ML): studies that fully constructed ML models for the prediction or diagnosis of high myopia, the diagnosis of pathological myopia, or the diagnosis of high myopia-associated glaucoma.
- Outcome measures: at least one of the following outcome indicators were reported: receiver operating characteristic (ROC), *c*-index, sensitivity, specificity, accuracy, recovery rate, accuracy rate, confusion matrix, F_1-score, and calibration curve.
- Datasets: (1) some studies lacked independent validation sets, and only *k*-fold cross-validation was leveraged to verify the effect of the constructed mode; and (2) in some publicly available datasets, particularly those involving medical imaging, different studies have reported the efficiency of varying ML methods.

Exclusion criteria

- Study type: (1) meta, review, guide, expert opinion; and (2) studies with too few samples (less than 20 cases).
- ML: literature that only executed the risk factor analysis but did not develop a complete ML mode.
- Outcome measures: none of the following outcomes were reported: ROC, *c*-index, sensitivity, specificity, accuracy, recovery rate, accuracy rate, confusion matrix, F_1-score, and calibration curve.

Data Sources and Search Strategy

PubMed, Cochrane, Embase, and Web of Science were thoroughly retrieved up to September 3, 2023, using the form of MeSH (Medical Subjects Headings) + free term, without any restrictions on region or publication period. The specific search strategy is depicted in Multimedia Appendix 2.

Study Selection and Data Extraction

Duplicates were excluded from the retrieved literature, and titles and abstracts were reviewed to delete obviously irrelevant studies. The full texts of the remaining studies were then downloaded and thoroughly read to determine the final included studies in the systematic review. A standard electronic data extraction spreadsheet was prepared prior to extracting data. The extracted data encompassed the title, first author, type of study, year of publication, author’s country, patient source, target event, number of cases of the target event, the total number of cases, number of training set cases, the total number of training set cases, method of validation set generation, number of events in the validation set, total number of cases in the validation set, type of models, and modeling variables.

Two researchers (HZ and LF) independently screened the literature and extracted data. Upon completion, their findings were cross-checked. A third reviewer (JH) was consulted for resolution in case of any dissents.

Risk of Bias in Studies

The risk of bias in the eligible studies was appraised by two independent reviewers (HZ and LF) using the prediction model risk of bias assessment tool [16]. This tool is comprised of a large number of questions in four domains (participants, predictors, outcomes, and analysis), which reflect overall bias risk and applicability. The 4 domains involve 2, 3, 6, and 9 specific questions, respectively, and each question may be answered by yes or probably yes, no or probably not, or no information. Following the quality evaluation, a cross-check was carried out. In the event of any disputes, a third researcher (JH) was consulted for resolution.

Synthesis Methods

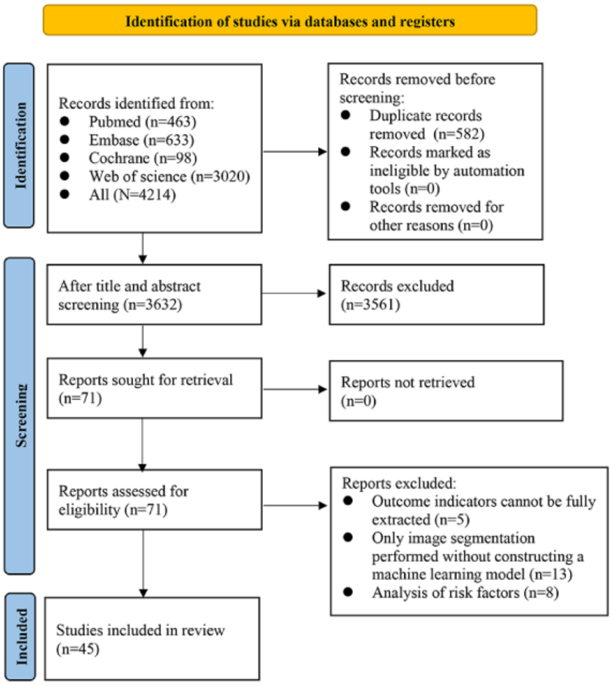
In some of the original studies included in our research, there was not only 1 validation set. Therefore, the number of models included in the meta-analysis does not equal the number of studies. The meta-analysis of sensitivity and specificity was executed using a bivariate mixed-effects model [17]. Sensitivity and specificity were meta-analyzed as per the diagnostic 2×2 table. However, most included studies did not provide the diagnostic 2×2 table. In such cases, the following two approaches were used to calculate the diagnostic 2×2 table: (1) it was computed based on sensitivity, specificity, and precision, combined with the number of cases; and (2) sensitivity and specificity were extracted based on the optimal Youden index, and then combined with the number of cases for calculation. The meta-analysis was implemented using R (version 4.2.0; R Foundation for Statistical Computing).

Results

Study Selection

A total of 4214 records were retrieved from the databases, of which 582 were duplicates. After reading the titles and abstracts, 3561 studies unrelated to ML in high myopia were excluded, leaving 71 studies. Of these, 13 only conducted image segmentation without constructing ML models, 5 did not provide full extractable outcome indicators, and 8 analyzed risk factors. Ultimately, 45 studies were incorporated into this review. The literature screening process is depicted in Figure 1.

Figure 1. Flowchart of literature screening.



Study Characteristics

The included studies were published from 2010 to 2023. Four of the studies [18-21] were about the prediction of high myopia, and the predicted variables were mainly derived from life characteristics, environmental and genetic factors, and routinely interpretable ocular clinical characteristics. Five of the studies [22-26] were about the diagnosis of high myopia, of which 1 study [22] also involved the diagnosis of pathological lesions of high myopia. Six studies focused on the diagnosis of high myopia-associated glaucoma [27-32]. Out of the included studies, 31 studies focused on the diagnosis of pathological myopia, primarily using optical coherence tomography and fundus imaging to construct artificial intelligence models. Of these, 26 studies [4,15,22,33-55] were based on DL (deep learning), while 5 studies [56-60] required manually coded ML for construction. Additionally, it was noted that in the 45 original studies, all 45 studies included binary classification tasks, with 9 studies [4,33,34,38,39,49,50,52,61] additionally incorporating multiclassification tasks. Regarding validation methods, 31 studies provided an external validation set, and 23 used a combination of internal and external validation sets. In terms of the generation method of validation set, 6 studies [23,24,34,40,47,59] used *k*-fold cross-validation, 29 [15,19-22,25-29,35-38,41,42,45,48-58,61] used random sampling, and 6 [4,18,32,33,44,60] applied a combination of *k*-fold cross-validation and random sampling. The detailed characteristics of the eligible studies are shown in Tables 1 and 2.

Table 1. Fundamental features of included studies.

First author	Year of publication	Country of authors	Study type	Patient source	Target events	Total number of cases
Tang et al [33]	2022	China, United States	Retrospective study	Multicenter	Diagnosis of pathologic myopia	1395 fundus photographs, 895 patients
Li et al [56]	2023	China	Nested case-control study	Single center	Diagnosis and prediction of pathological myopia	20,870 patients
Du et al [57]	2021	China	Retrospective study	Single center	Diagnosis of pathologic myopia	313 patients with high myopia and 457 eyes
Foo et al [18]	2023	Singapore	Prospective study	Multicenter	Prediction of high myopia	965 children with 1878 eyes and 7456 fundus photographs
Kim et al [58]	2021	Korea	Retrospective study	Multicenter	Diagnosis of pathologic myopia	860 eyes
Zhang et al [59]	2013	Singapore	Retrospective study	Registry database	Diagnosis of pathologic myopia	2258 patients
Zhu et al [34]	2023	China	Retrospective study	Single center	Diagnosis of pathologic myopia	6078 photographs
Wu et al [35]	2022	China	Retrospective study	Single center	Diagnosis of pathologic myopia	1853 photographs
Ye et al [36]	2021	China	Retrospective study	Single center	Diagnosis of pathologic myopia	1041 patients with pathologic myopia and with 2342 eligible OCT[a] macular images
Wang et al [37]	2023	China	Retrospective study	Single center	Diagnosis of pathologic myopia	7606 patients with 10,347 fundus photographs
Wang et al [19]	2022	China	Prospective, longitudinal, observational study	Wenzhou large-scale survey	Prediction of myopia and high myopia	15,765 patients
Wan et al [4]	2021	China	Retrospective study	Single center	Diagnosis of pathologic myopia	858 photographs
Wan et al [38]	2023	China	Retrospective study	Single center	Diagnosis of pathologic myopia	1750 photographs
Tan et al [22]	2021	Singapore	Retrospective multicohort study	Multicenter + registry database	Diagnosis of high myopia + pathological myopia	125,421 patients with 251,349 photographs
Sun et al [39]	2023	China	Retrospective multicohort study	Multicenter + registry database	Diagnosis of pathologic myopia	1514 fundus photographs
Sogawa et al [40]	2020	Japan	Retrospective study	Single center	Diagnosis of pathologic myopia	910 patients with 910 images
Du et al [41]	2022	Japan	Retrospective study	Single center	Diagnosis of pathologic myopia	1327 patients with 2400 high myopia eyes and 9176 OCT images
Hou et al [60]	2023	China	Prospective cohort study	Single center	Diagnosis of pathologic myopia	576 patients
Li et al [52]	2022	China	Retrospective cohort study	Multicenter	Pathologic myopia	29,230 patients with 57,148 fundus photographs
Li et al [27]	2021	China	Case-control study	Multicenter	Diagnosis of glaucoma in high myopia	2731 participants with 2731 eyes
Chen et al [20]	2019	China	Prospective study	Single center	Prediction of high myopia	1063 patients
Choi et al [23]	2021	Korea	Retrospective study	Single center	Prediction of high myopia	492 patients with 690 eyes
Cui et al [42]	2021	China, Taiwan	Retrospective study	Registry database	Diagnosis of pathologic myopia	800 images

First author	Year of publication	Country of authors	Study type	Patient source	Target events	Total number of cases
Guan et al [24]	2023	China	Retrospective study	Multicenter	Prediction of high myopia	1,285,609 participants
He et al [61]	2022	China	Retrospective study	Multicenter	Diagnosis of pathologic myopia	2866 patients with 3945 OCT images
Hemelings et al [15]	2021	Belgium	Retrospective study	Registry database	Diagnosis of pathologic myopia	1200 photographs
Rauf et al [44]	2021	Pakistan	Retrospective study	Registry database	Diagnosis of pathologic myopia	840 photographs
Park et al [45]	2022	Korea	Retrospective study	Single center	Diagnosis of pathologic myopia	367 eyes
Lu et al [46]	2021	China	Retrospective study	Single center	Diagnosis of pathologic myopia and diagnosis of pathologic myopia	• 17,330 photographs • 17,330 photographs
Lu et al [47]	2021	China	Retrospective study	Multicenter	Diagnosis of pathologic myopia	32,419 patients with 37,659 images
Liu et al [54]	2010	Singapore	Retrospective study	Single center	Pathologic myopia	80 photographs
Li et al [48]	2022	China, United States	Retrospective study	Single center	Diagnosis of pathologic myopia	1139 patients with 5917 images
Lee et al [28]	2023	Korea	Retrospective study	Single center	Diagnosis of glaucoma in high myopia	260 eyes and 260 images
Kim et al [29]	2023	Korea	Retrospective study	Single center	Diagnosis of glaucoma in high myopia	2607 eyes
Jeong et al [30]	2023	Korea	Retrospective cross-sectional study	Single center	Diagnosis of glaucoma in high myopia	274 patients
Huang et al [21]	2022	China	Case-control study	Single center	Prediction of high myopia	1298 patients
Huang et al [49]	2023	China, United Kingdom	Retrospective study	Single center	Diagnosis of pathologic myopia	1131 patients with 3441 images
Du et al [50]	2021	Japan	Retrospective study	Single center	diagnosis of pathologic myopia	4432 eyes and 7020 images
Crincoli et al [51]	2023	Italy	Case-control study	Multicenter	diagnosis of pathologic myopia	84 patients with 84 eyes and 252 photographs
Asaoka et al [31]	2014	Japan	Case-control study	Multicenter	Diagnosis of glaucoma in high myopia	242 patients and 242 eyes
Bowd et al [32]	2023	United States, Germany	Retrospective study	Single center	Diagnosis of glaucoma in high myopia	593 eyes
Zhao et al [25]	2022	China	Retrospective study	Single center	Prediction of high myopia	546 patients
Liu et al [53]	2010	Singapore	Retrospective study	Single center	Diagnosis of pathologic myopia	80 photographs
Dai et al [26]	2020	China	Retrospective study	Single center	Prediction of high myopia	319 patients with 932 images
Baid et al [55]	2019	India	Retrospective study	Registry database	Diagnosis of pathologic myopia	481 photographs

[a] OCT: optical coherence tomography.

Table 2. Fundamental features of included studies.

Total number of cases in training set	Generation of validation set	Total number of cases in validation set	Total number of cases in test set	Model type	Modeling variables
727 fundus photographs	5-fold cross-validation + random sampling	238 fundus photographs	238 fundus photographs	DL[a]	Fundus photographs
2069 patients	Random sampling	1382 patients	Unclear	ACP[b], ML[c]	Clinical features
319 eyes	Random sampling	138 eyes	Unclear	ML-based radiomics analysis method	Fundus photographs
769 children with 1502 eyes and 5945 photographs	Internal validation (5-fold cross-validation + random sampling) + multicenter external validation	196 children with 376 eyes and 1511 fundus photographs	99 children with 189 eyes and 821 photographs	DL	Fundus photographs + clinical features
602 eyes	Random sampling	258 eyes	unclear	SVM[d], ML	Fundus photographs
2258 patients	Stratified 20-fold cross-validation	unclear	unclear	SVM, ML	SNP[e] + clinical features + fundus photographs
4252 photographs	Stratified 20-fold cross-validation	unclear	1826 photographs	DL	Fundus photographs
1483 photographs	Random sampling	unclear	370 photographs	DL	Fundus photographs
1874 photographs	Internal validation (random sampling) + external validation（multicenter）	468 photographs	450 photographs	DL	Fundus photographs
5003 patients with 7389 photographs	Random sampling	775 patients with 821 photographs	1828 patients with 2137 photographs	DL	Fundus photographs
11,350 patients	Internal validation (random sampling) + external validation(prospective)	4415 patients	6168 patients (prognostic cohort)	LR[f], GBDT[g], NN[h]	Clinical features
758 photographs	5-fold cross-validation + random sampling	100 photographs	Unclear	DL	Fundus photographs
1402 photographs	Random sampling	174 photographs	174 photographs	DL	Fundus photographs
226,686 photographs	Internal validation (random sampling) + external validation（multicenter）	11,303 photographs	213,475 photographs	DL	Fundus photographs
400 fundus photographs	Multicenter	400 fundus photographs	714 fundus photographs	DL	Fundus photographs
Unclear	5-fold cross-validation	Unclear	Unclear	DL	Fundus photographs
7865 photographs	random sampling	1311 photographs	Unclear	DL	Fundus photographs
516 patients	10-fold cross-validation + random sampling	60 patients	Unclear	XGBoost[i], SVM, LR	Clinical features + metabolic characteristics
29,213 photographs	Internal validation (random sampling) + external validation（multicenter）	7302 photographs	16,554 photographs	DCNN[j], DL	Fundus photographs
2223 participants with 2223 eyes	Random sampling	508 participants with 508 eyes	Unclear	FCN[k]	OCT[l] images + clinical features
638 patients	Random sampling	425 patients	Unclear	LR	Genetic factors + clinical features
434 patients with 600 eyes and 1200 images	5-fold cross-validation	Unclear	58 patients with 90 eyes and 180 images	CNN[m], DL	OCT images
400 images	Random sampling	200 images	200 images	DL	Fundus photographs
1600 participants	Internal validation (5-fold cross-validation)	Unclear	400 patients	RF[n], LR, SVM	Clinical features
2380 images	Random sampling	680 photographs	340 photographs	DL	OCT images

Total number of cases in training set	Generation of validation set	Total number of cases in validation set	Total number of cases in test set	Model type	Modeling variables
400 photographs	Random sampling	400 photographs	400 photographs	DL	Fundus photographs
400 photographs	10-fold cross-validation + random sampling	40 photographs	400 photographs	DL	Fundus photographs
293 eyes	random sampling	37 eyes	37 eyes	DL	3D OCT images
11,502 photographs	Unclear	3284 photographs	1642 photographs	DL	Fundus photographs
2457 photographs	Unclear	707 photographs	372 photographs	DL	Fundus photographs
32,010 images	Internal validation (5-fold cross-validation) + external validation (multicenter)	Unclear	732 patients with 1000 images	DL	Fundus photographs
40 photographs	Random sampling	Unclear	40 photographs	SVM, DL	Fundus photographs
838 patients with 4338 images	Internal validation (random sampling) + external validation (prospective)	210 patients with 1167 photographs	91 patients with 174 eyes and 412 photographs	DL	OCT macular images
165 images	Random sampling	46 photographs	49 photographs	DL	OCTA[o] and OCT images
1416 eyes	Internal validation (random sampling) + external validation	471 eyes	720 eyes	DL	OCT images
Unclear	Unclear	Unclear	Unclear	Decision tree	OCT images
325 patients	Random sampling	973 patients	Unclear	DL	Genetic + clinical features
2264 images	Internal validation (random sampling) + external validation (prospective)	501 photographs	604 photographs	DL	OCT images
4140 photographs	Random sampling	1036 photographs	1844 photographs	DL	Fundus photographs
176 photographs	Random sampling	25 photographs	51 photographs	DL	OCT images
Unclear	Unclear	Unclear	Unclear	RF	HRT parameters
347 eyes	5-fold cross-validation + random sampling	87 eyes	159 eyes	CNN	OCT images
928 fundus photographs	Random sampling	232 photographs	Unclear	DL	Fundus photographs
40 photographs	Random sampling	Unclear	40 photographs	SVM or DL	Fundus photographs
792 photographs	Random sampling	Unclear	140 photographs	DL	Fundus photographs
374 photographs	Random sampling	80 photographs	27 photographs	CNN	Fundus photographs

[a] DL: deep learning.

[b] ACP: algorithm of conditional probability.

[c] ML: machine learning.

[d] SVM: support vector machine.

[e] SNP: single nucleic polymorphism.

[f] LR: logistic regression.

[g] GBDT: gradient boosted decision tree.

[h] NN: neural network.

[i] XGBoost: extreme gradient boosting

[j] DCNN: deep convolutional neural networks

[k] FCN: fully connected network

[l] OCT: optical coherence tomography.

[m] CNN: convolutional neural networks

[n] RF: random forest

[o] OCTA: optical coherence tomography angiography.

Risk of Bias in Studies

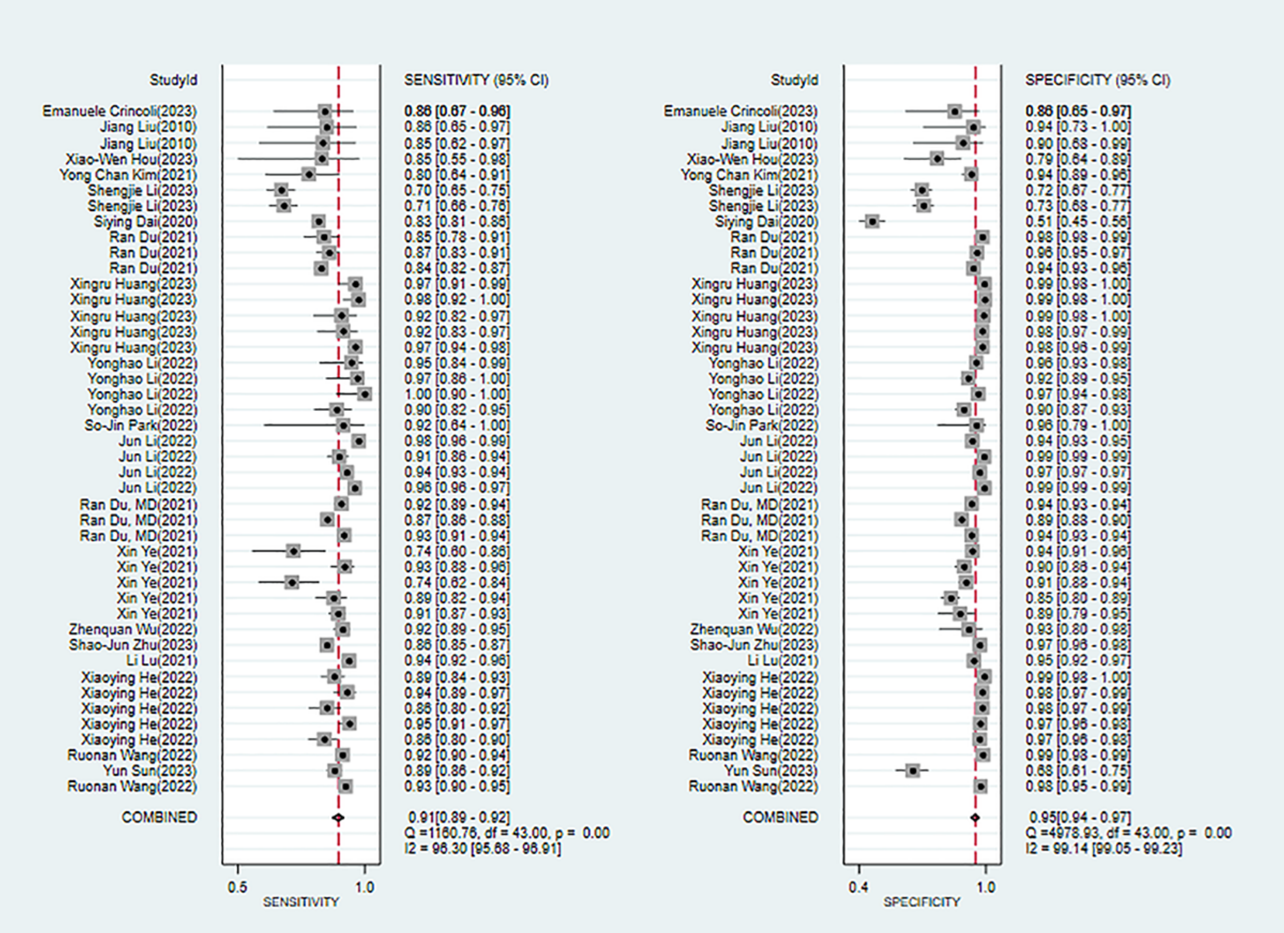
This review incorporated 67 models. There were 36 retrospective studies [4,15,22-26,28-30,32-42,44-50,52-55,57-59,61] that constructed 39 models, indicating a high bias in the choosing of study participants. Five case-control studies [21,27,31,51,56] constructed 13 models, also showing high bias in the selection of study participants. Since the predictors were evaluated in the context of a known outcome in the case-control studies, there was a high bias in the assessment of predictive factors. Twelve studies [19,20,23,24,27,30,31,56-60] constructed 22 models based on manually coded ML, with a high bias in predictive factors. In terms of statistical analysis, 2 studies [21,45] with 5 models did not meet the requirement of having an event per variable>20%, indicating a high risk of bias. In the statistical analysis, 32 models in 34 studies [4,15,18,21-23,25,26,28,29,32-42,44-55,61] could not estimate event per variable due to the use of the DL method. Additionally, 10 studies [19,20,24,27,30,31,56,58-60] with 29 models in ML did not report on the complexity of the data, rendering it difficult to determine their bias risk. Five studies [20,27,30,31,60] with 11 models were identified as having a high risk of bias in statistical analysis because they did not perform cross-validation to adjust the stability of models with different parameters. In summary, in terms of research participants, 14 models had a low risk of bias; 52 models had a high risk of bias, and 1 model had an unclear risk of bias. In terms of predictors, 37 models had a low risk of bias and 30 models had a high risk of bias. In terms of outcomes, all 67 models had a low risk of bias. In terms of statistical analysis, 3 models had a low risk of bias, 16 models had a high risk of bias, and 48 models had an unclear risk of bias.

Meta-Analysis of ML for Binary Classification Tasks

Pathological Myopia

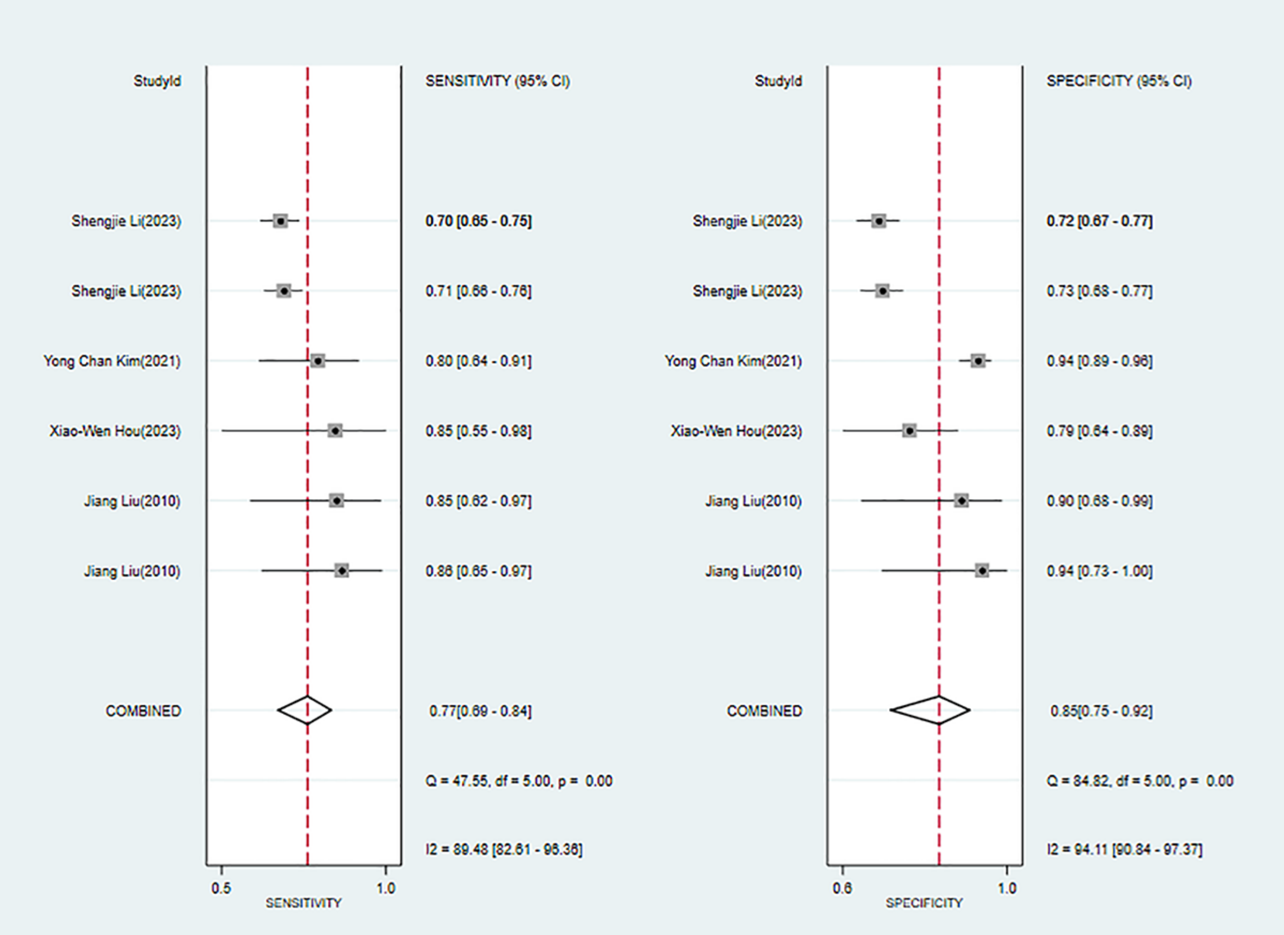
Twenty studies [26,34-37,39,41,45,47-54,56,58,60,61] reported ML for diagnosing pathological myopia. Modeling algorithms included algorithms of conditional probability, support vector machines (SVMs), logistic regression (LR), extreme gradient boosting, convolutional neural networks (CNNs), and deep convolutional neural networks (DCNNs). The overall sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and summary receiver operating characteristic (SROC) were 0.91 (95% CI 0.89-0.92), 0.95 (95% CI 0.94-0.97), 19.7 (95% CI 13.8-28.2), 0.10 (95% CI 0.08-0.12), 201 (95% CI 122-331), and 0.97 (95% CI 0.95-0.98), respectively. The Deek funnel plot indicated no substantial evidence of publication bias in the included ML models. Assuming that the prior probability of pathological myopia was 20% if the result of ML was pathological myopia, then the probability of true pathological myopia would be 83%. If the result of ML was nonpathological myopia, then the probability of true pathological myopia would be 2% (ie, the probability of true nonpathological myopia was 98%; Figure 2 and Figures S1-S3 in Multimedia Appendix 3).

Figure 2. Forest plot for the meta-analysis of sensitivity and specificity of machine learning in detecting pathological myopia [26,34-37,39,41,45,47-54,56,58,60,61]. Note: the pooled sensitivity and specificity of 44 models from 20 machine learning studies on the diagnosis of pathological myopia were 0.91 (95% CI 0.89-0.92) and 0.95 (95% CI 0.94-0.97), respectively.



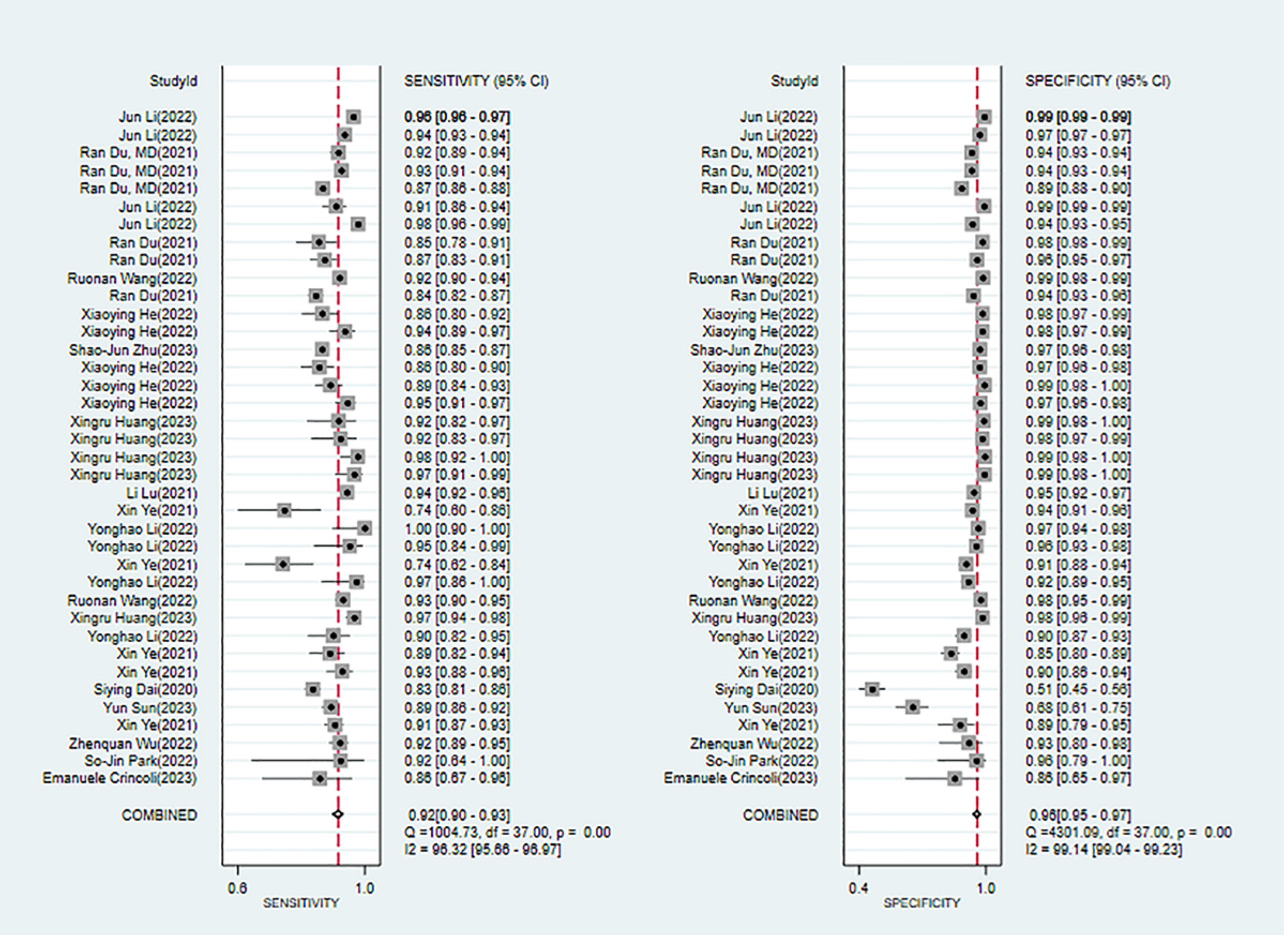
Five studies [53,54,56,58,60] reported conventional ML (non-DL) for diagnosing pathological myopia. Modeling algorithms included algorithms of conditional probability, SVM, extreme gradient boosting, and LR. The overall sensitivity, specificity, PLR, NLR, DOR, and SROC curve were 0.77 (95% CI 0.69-0.84), 0.85 (95% CI 0.75-0.92), 5.2 (95% CI 2.8-9.8), 0.27 (95% CI 0.18-0.39), 20 (95% CI 7-51), and 0.86 (95% CI 0.75-0.92), respectively. The Deek funnel plot indicated the presence of publication bias in the conventional ML (non-DL) models. Assuming that the prior probability of pathological myopia for conventional ML (non-DL) was 20% if the result of conventional ML (non-DL) was pathological myopia, then the probability of true pathological myopia would be 57%. If the result of conventional ML (non-DL) was nonpathological myopia, then the probability of true pathological myopia would be 6% (ie, the probability of true nonpathological myopia was 94%; Figure 3 and Figures S4-S6 in Multimedia Appendix 3).

Figure 3. Forest plot for the meta-analysis of sensitivity and specificity of conventional machine learning (non-deep learning) in detecting pathological myopia [53,54,56,58,60]. Note: the pooled sensitivity and specificity of 6 models from 5 conventional machine learning (non-deep learning) studies on the diagnosis of pathological myopia were 0.77 (95% CI 0.69-0.84) and 0.85 (95% CI 0.75-0.92), respectively.



Fifteen studies [26,34-37,39,41,45,47-52,61] mentioned DL for diagnosing pathological myopia. Modeling algorithms included CNN and DCNN. The overall sensitivity, specificity, PLR, NLR, DOR, and SROC were 0.92 (95% CI 0.90-0.93), 0.96 (95% CI 0.95-0.97), 23.7 (95% CI 16.5-34.0), 0.09 (95% CI 0.07-0.11), 271 (95% CI 168-437), and 0.97 (95% CI 0.95-0.98), respectively. The Deek funnel plot revealed no remarkable publication bias in the DL models. Assuming that the prior probability of pathological myopia for DL was 20% if the result of DL was pathological myopia, then the probability of true pathological myopia would be 86%. If the result of DL was nonpathological myopia, then the probability of true pathological myopia would be 2% (ie, the probability of true nonpathological myopia was 98%; Figure 4 and Figures S7-S9 in Multimedia Appendix 3).

Figure 4. Forest plot for the meta-analysis of sensitivity and specificity of deep learning in detecting pathological myopia [26,34-37,39,41,45,47-52,61]. Note: the pooled sensitivity and specificity of 38 models from 15 deep learning studies on the diagnosis of pathological myopia were 0.92 (95% CI 0.90-0.93) and 0.96 (95% CI 0.95-0.97), respectively.

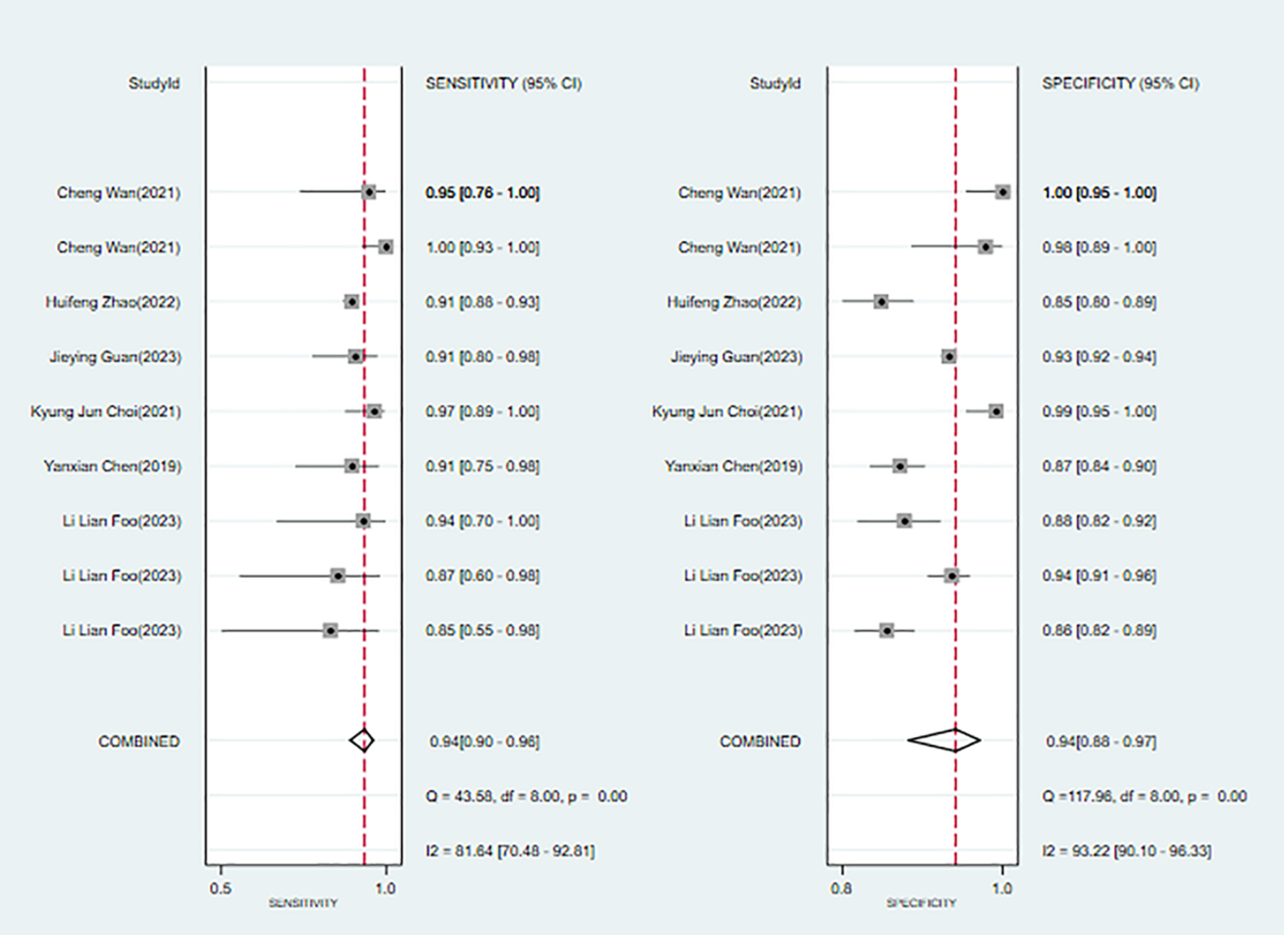


High Myopia

Six studies [4,18,20,23-25] discussed ML for diagnosing and forecasting high myopia. Modeling algorithms included DCNN, CNN, LR, SVM, random forest (RF), and linear mixed models. The sensitivity, specificity, PLR, NLR, DOR, and SROC were 0.94 (95% CI 0.90-0.96), 0.94 (95% CI 0.88-0.97), 16.2 (95% CI 7.7-33.8), 0.06 (95% CI 0.04-0.11), 255 (95% CI 79-822), and 0.98 (95% CI 0.96-0.99), respectively. The Deek funnel plot indicated no substantial evidence of publication bias in the included ML models. Assuming that the prior probability of high myopia for ML was 20% if the result of ML was high myopia, then the probability of true high myopia would be 80%. If the result of ML was non-high myopia, then the probability of true high myopia would be 2% (ie, the probability of true non-high myopia was 98%; Figure 5 and Figures S10-S12 in Multimedia Appendix 3).

Three studies [4,23,25] focused on diagnosing high myopia, while 3 studies [18,20,24] focused on predicting high myopia. Due to the limited number of studies included, we did not perform a meta-analysis for the diagnostic and prediction tasks. In the validation sets of the diagnostic tasks, sensitivity ranged from 0.91 to 1.00 and specificity ranged from 0.85 to 1.00, while in the validation sets of the prediction tasks, these values were 0.85-0.94 and 0.86-0.94, respectively. We found that both diagnostic and prediction tasks demonstrated highly favorable performance.

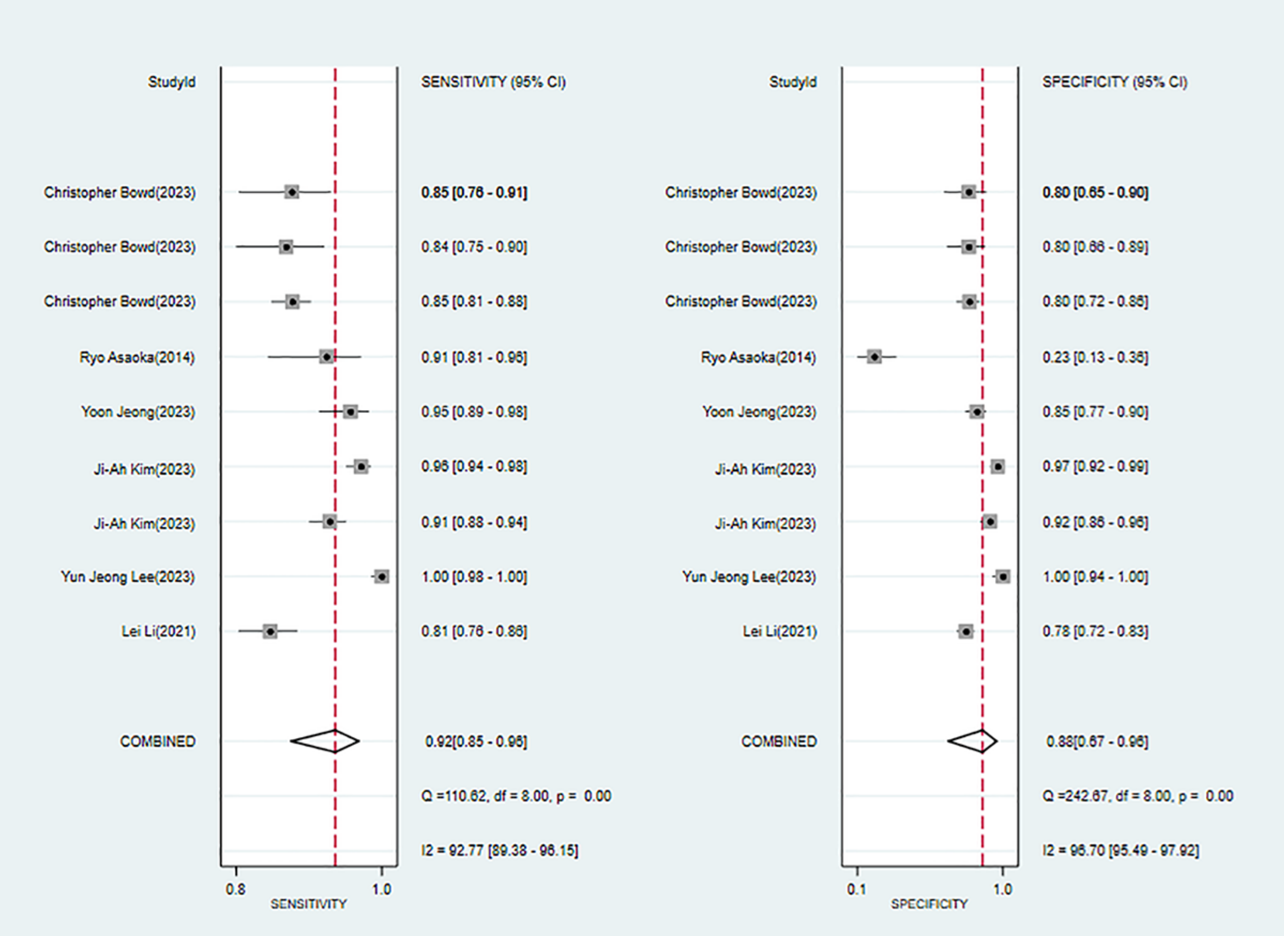
Figure 5. Forest plot for the meta-analysis of sensitivity and specificity of machine learning in detecting high myopia [4,18,20,23-25]. Note: the pooled sensitivity and specificity of 9 models from 6 machine learning studies on the diagnosis and prediction of high myopia were 0.94 (95% CI 0.90-0.96) and 0.94 (95% CI 0.88-0.97), respectively.



High Myopia–Associated Glaucoma

Six studies [27-32] mentioned ML for diagnosing high myopia-associated glaucoma. Modeling algorithms included Lagrange multiplier, fully connected network, radial basis function network, decision tree, RF, and CNN. The sensitivity, specificity, PLR, NLR, DOR, and SROC curve were 0.92 (95% CI 0.85-0.96), 0.88 (95% CI 0.67-0.96), 7.6 (95% CI 2.4-23.8), 0.09 (95% CI 0.04-0.20), 84 (95% CI 13-555), and 0.96 (95% CI 0.94-0.97), respectively. The Deek funnel plot indicated no substantial evidence of publication bias in the included ML models. Assuming that the prior probability of high myopia–associated glaucoma was 20% if the result of ML was high myopia-associated glaucoma, then the probability of true high myopia–associated glaucoma would be 65%. If the result of ML was non-high myopia–associated glaucoma, then the probability of true high myopia–associated glaucoma would be 2% (ie, the probability of true non-high myopia–associated glaucoma was 98%; Figure 6 and Figures S13-S15 in Multimedia Appendix 3).

Figure 6. Forest plot for the meta-analysis of sensitivity and specificity of machine learning in detecting high myopia-associated glaucoma [27-32]. Note: the pooled sensitivity and specificity of 9 models from 6 machine learning studies on the diagnosis of high myopia-associated glaucoma were 0.92 (95% CI 0.85-0.96) and 0.88 (95% CI 0.67-0.96), respectively.



Review of ML for Multiclassification Tasks

Out of the included studies, 9 [4,33,34,38,39,49,50,52,61] used ML for multiclassification tasks. Due to significant variations in the diagnostic differences across these multiclassification tasks, a quantitative analysis was not feasible. Five studies [33,34,38,39,50] focused on fundus images–based DL to detect different types of myopic atrophy maculopathy in high myopia, with an accuracy ranging from 88% to 97%. Two studies [49,61] used optical coherence tomography (OCT) image–based DL to detect different types of myopic traction maculopathy in high myopia, with an accuracy ranging from 91% to 96%. One study [4] used fundus image–based DL to differentiate between normal, low-risk high myopia, and high-risk high myopia, with an accuracy of 99%. One study [52] applied fundus image–based DL to distinguish between normal, fundus tessellation, and pathologic myopia, with an accuracy of 94%, as illustrated in Table 3.

Table 3. Results of machine learning for multiclassification tasks.

First author	Year	Diagnostic purpose	Types of artificial intelligence	Modeling variables	Generation of validation set	Accuracy rate, %
Tang et al [33]	2022	Classification of atrophic macular lesions in myopic	CNNs[a]; DL[b]	Fundus photographs	5-fold cross-validation + random sampling	94
Zhu et al [34]	2023	Classification of atrophic macular lesions in myopic	Neural network; DL	Fundus photographs	Stratified 20-fold cross-validation	90
Wan et al [4]	2021	Normal, low, and high risk of high myopia	DCNNs[c]; DL	Fundus photographs	5-fold cross-validation + random sampling	99
Wan et al [38]	2023	Classification of atrophic macular lesions in myopic	DL	Fundus photographs	Random sampling	95-97
Sun et al [39]	2023	Classification of atrophic macular lesions in myopic	DL	Fundus photographs	External validation (multicenter)	89.2
Li et al [52]	2022	Differential diagnosis of normal, leopard print fundus, and pathological myopia	DCNN; DL	Fundus photographs	Internal validation (random sampling) + external validation (multicenter)	94
He et al [61]	2022	Differential diagnosis of tractive macular degeneration and neovascular macular degeneration in high myopia, and others	DL	OCT[d] images	Random sampling	91-96
Huang et al [49]	2023	Classification of tractive macular degeneration in high myopia	DL	OCT images	Internal validation (random sampling) + external validation (prospective)	96
Du et al [50]	2021	Classification of atrophic macular lesions in myopic	DL	Fundus photographs	Random sampling	88

[a] CNN: convolutional neural network.

[b] DL: deep learning.

[c] DCNN: deep convolutional neural network.

[d] OCT: Optical Coherence Tomography.

Discussion

Summary of the Main Findings

This study comprehensively described the accuracy of ML in detecting high myopia, high myopia-associated glaucoma, and pathologic myopia. ML demonstrated exceptionally favorable performance in detecting high myopia, while DL was highly accurate in diagnosing pathologic myopia.

Comparison With Previous Reviews

Previous studies have also explored the detection accuracy of ML in this field. A systematic review has reported that fundus image– or OCT image–based DL can effectively diagnose and classify myopic maculopathy. Additionally, ML examination of the optic disc area can detect myopic maculopathy that may not be easily identified during clinical examination [14]. A recent meta-analysis based on only 11 studies evaluated the performance of DL in identifying pathological myopia based on fundus images. The SROC, specificity, and sensitivity were found to be 0.9905, 0.959 (95% CI 0.955-0.962), and 0.965 (95% CI 0.963-0.966), respectively [62]. In the previous meta-analysis, the 11 original studies all constructed fundus images-based DL models, and studies on conventional ML (non-DL) were not incorporated. The number of studies included in our review was further expanded, with a total of 20 studies on the performance of ML in diagnosing pathological myopia. Moreover, subgroup analysis was executed between conventional ML (non-DL) and DL. Our finding also indicated that DL demonstrated exceptionally favorable efficiency in detecting pathological myopia.

As the understanding of the etiology of myopia deepens, growing evidence reveals risk factors for the onset or progression of myopia, including age, sex, parental myopia, susceptibility genes, and outdoor activities. For high myopia, early prediction appears to be more beneficial. Among the included studies, one incorporated 135 myopia-related single nucleotide polymorphisms to forecast the progression and onset of high myopia. ML for the prediction of high myopia was mainly based on genetic factors, environmental factors, and ocular clinical characteristics. ML showed an SROC of 0.96, sensitivity of 0.91, and specificity of 0.87, respectively [20], suggesting that ML methods can effectively identify high-risk individuals with high myopia, thus effectively preventing this condition, especially in minors.

Glaucoma is a significant contributor to irreversible vision impairment and blindness all over the world. A 10-year study in Chinese individuals over the age of 40 years found that every 1 mm increase in axial length increased the risk of open-angle glaucoma by 1.72 times. In comparison to emmetropic and hyperopic eyes, highly myopic eyes had a 7.3 times higher risk of developing open-angle glaucoma [63]. Due to the changes in retinal structure caused by myopia, diagnosing glaucoma in myopic patients, especially those with high myopia, is challenging. Six studies were included to evaluate the diagnosis of high myopia glaucoma. Of them, 3 studies [28,29,32] used fundus OCT image-based DL techniques, while the remaining 3 [27,30,31] used non-DL ML (Lagrange multiplier, fully connected network, radial basis function network, decision tree, RF) approaches using OCT parameters, Heidelberg Retina Tomograph parameters, and ocular biometric parameters of patients. The findings indicated that ML yielded highly promising results in the detection of high myopia glaucoma.

It was also noted that different ML methods, conventional ML and DL, showed significant differences in their ability to identify positive or outcome events. Conventional ML is often used to construct models with interpretable clinical features. Lately, various image-based ML methods have emerged. However, a significant challenge in this context is the requirement for manual annotation to facilitate ML. From this standpoint, manual annotation poses a formidable barrier to effectively mitigating the risk of bias. DL, on the other hand, enables intelligent processing of medical images and has been widely applied in various fields, including detecting diabetic retinopathy [8-10], retinopathy of prematurity [64,65], age-related macular degeneration [10], and glaucoma [11-13]. With the rapid development of ML, imaging data are increasingly becoming a valuable source for medical analysis. Multiple studies have demonstrated that images from various sources, including fundus images [66], external eye appearance [67], and refractive images [68], can effectively estimate a patient's spherical refractive error, indicating the potential of imaging data in predicting the risk of myopia. This study also finds that image-based DL is more accurate than conventional ML, providing a theoretical basis for the creation of future intelligent tools.

Additionally, the dataset used in ML demands considerable attention. Many studies are hampered by a limited number of cases, raising concerns about the robustness of the findings. Additionally, validation methods often depend heavily on internal validation, which may not fully capture the model's generalizability. Hence, incorporating comprehensive patient data is essential for building a robust large-scale database, which will enable the development of ML models that are applicable to a broader population. Among the studies included, 7 [15,22,39,42,44,55,59] established ML models based on publicly available large databases.

Limitations

Although our review includes a larger number of studies than previous meta-analyses and provides an evidence-based basis for subsequent studies, this study has limitations. First, there were few studies on the prediction of high myopia, which limits the interpretation of our results, and clinically interpretable variables for predicting high myopia were not explained. Second, we did not conduct a subgroup analysis on the type of ML (conventional ML vs DL) owing to the insufficient number of included studies based on high myopia glaucoma and high myopia. Third, the majority of the models included in this study were assessed as having a high risk of bias, which may impact the interpretation of our results. Most included studies adopted a retrospective design, which might lead to selection bias.

Conclusions

In conclusion, this study comprehensively reviews and meta-analyzes the performance of ML in the diagnosis and prediction of high myopia, high myopia-associated glaucoma, and pathological myopia, providing valuable guidance and references for future research. Challenges exist within the emerging field of myopia prediction. With the development of new analytical methods and the accumulation of real medical datasets, future research holds the promise of improving the prediction of myopia onset and progression. This advancement brings us closer to the ultimate goal of identifying high-risk individuals promptly and implementing targeted interventions in clinical practice.

Acknowledgments

This work was supported by the Joint Project on Regional High-Incidence Diseases Research of Guangxi Natural Science Foundation under grant (2024GXNSFAA010322); Science and Technology Plan of Qingxiu District, Nanning City (2020016); Medical and Health Appropriate Technology Development and Promotion Application Project of Guangxi Zhuang Autonomous Region (S2018093); and Self-Funded Research Project of Health Commission of Guangxi Zhuang Autonomous Region (Z20210589).

Data Availability

All data generated or analyzed during this study are included in this published article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - jmir_v27i1e57644_app1.docx]

Multimedia Appendix 2
Literature search strategy.
[DOCX File , 17 KB - jmir_v27i1e57644_app2.docx]

Multimedia Appendix 3
Additional figures.
[DOCX File , 7517 KB - jmir_v27i1e57644_app3.docx]

References

1. Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. Ophthalmology 2016;123(5):1036-1042 [FREE Full text] [doi: 10.1016/j.ophtha.2016.01.006] [Medline: 26875007]
2. Naidoo KS, Fricke TR, Frick KD, Jong M, Naduvilath TJ, Resnikoff S, et al. Potential lost productivity resulting from the global burden of myopia: systematic review, meta-analysis, and modeling. Ophthalmology 2019;126(3):338-346 [FREE Full text] [doi: 10.1016/j.ophtha.2018.10.029] [Medline: 30342076]
3. Flitcroft DI, He M, Jonas JB, Jong M, Naidoo K, Ohno-Matsui K, et al. IMI—defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. Invest Ophthalmol Vis Sci 2019;60(3):M20-M30 [FREE Full text] [doi: 10.1167/iovs.18-25957] [Medline: 30817826]
4. Wan C, Li H, Cao GF, Jiang Q, Yang WH. An artificial intelligent risk classification method of high myopia based on fundus images. J Clin Med 2021;10(19):4488 [FREE Full text] [doi: 10.3390/jcm10194488] [Medline: 34640509]
5. Ikuno Y. Overview of the complications of high myopia. Retina 2017;37(12):2347-2351. [doi: 10.1097/IAE.0000000000001489] [Medline: 28590964]
6. Miller RJH. Artificial intelligence in nuclear cardiology. Cardiol Clin 2023;41(2):151-161. [doi: 10.1016/j.ccl.2023.01.004] [Medline: 37003673]
7. Majumder A, Sen D. Artificial intelligence in cancer diagnostics and therapy: current perspectives. Indian J Cancer 2021;58(4):481-492 [FREE Full text] [doi: 10.4103/ijc.IJC_399_20] [Medline: 34975094]
8. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017;318(22):2211-2223 [FREE Full text] [doi: 10.1001/jama.2017.18152] [Medline: 29234807]
9. Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. PLoS One 2019;14(6):e0217541 [FREE Full text] [doi: 10.1371/journal.pone.0217541] [Medline: 31170223]
10. González-Gonzalo C, Sánchez-Gutiérrez V, Hernández-Martínez P, Contreras I, Lechanteur YT, Domanian A, et al. Evaluation of a deep learning system for the joint automated detection of diabetic retinopathy and age-related macular degeneration. Acta Ophthalmol 2020;98(4):368-377 [FREE Full text] [doi: 10.1111/aos.14306] [Medline: 31773912]
11. Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. JAMA Ophthalmol 2019;137(12):1353-1360 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.3501] [Medline: 31513266]
12. Hood DC, De Moraes CG. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology 2018;125(8):1207-1208. [doi: 10.1016/j.ophtha.2018.04.020] [Medline: 30032794]
13. Mirzania D, Thompson AC, Muir KW. Applications of deep learning in detection of glaucoma: a systematic review. Eur J Ophthalmol 2021;31(4):1618-1642. [doi: 10.1177/1120672120977346] [Medline: 33274641]
14. Li Y, Foo LL, Wong CW, Li J, Hoang QV, Schmetterer L, et al. Pathologic myopia: advances in imaging and the potential role of artificial intelligence. Br J Ophthalmol 2023;107(5):600-606. [doi: 10.1136/bjophthalmol-2021-320926] [Medline: 35288438]
15. Hemelings R, Elen B, Blaschko MB, Jacob J, Stalmans I, De Boever P. Pathological myopia classification with simultaneous lesion segmentation using deep learning. Comput Methods Programs Biomed 2021;199:105920. [doi: 10.1016/j.cmpb.2020.105920] [Medline: 33412285]
16. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]

17. Hsiao V, Massoud E, Jensen C, Zhang Y, Hanlon BM, Hitchcock M, et al. Diagnostic accuracy of fine-needle biopsy in the detection of thyroid malignancy: a systematic review and meta-analysis. JAMA Surg 2022;157(12):1105-1113 [FREE Full text] [doi: 10.1001/jamasurg.2022.4989] [Medline: 36223097]
18. Foo LL, Lim GYS, Lanca C, Wong CW, Hoang QV, Zhang XJ, et al. Deep learning system to predict the 5-year risk of high myopia using fundus imaging in children. NPJ Digit Med 2023;6(1):10 [FREE Full text] [doi: 10.1038/s41746-023-00752-8] [Medline: 36702878]
19. Wang H, Li L, Wang W, Wang H, Zhuang Y, Lu X, et al. Simulations to assess the performance of multifactor risk scores for predicting myopia prevalence in children and adolescents in China. Front Genet 2022;13:861164 [FREE Full text] [doi: 10.3389/fgene.2022.861164] [Medline: 35480319]
20. Chen Y, Han X, Guo X, Li Y, Lee J, He M. Contribution of genome-wide significant single nucleotide polymorphisms in myopia prediction: findings from a 10-year cohort of Chinese twin children. Ophthalmology 2019;126(12):1607-1614 [FREE Full text] [doi: 10.1016/j.ophtha.2019.06.026] [Medline: 31416661]
21. Huang Z, Xu X, Cheng W. Deep learning-based health management model application in extreme myopia eye vision monitoring and risk prediction. J Oncol 2022;2022:4537021. [doi: 10.1155/2022/4537021] [Medline: 36276280]
22. Tan TE, Anees A, Chen C, Li S, Xu X, Li Z, et al. Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study. Lancet Digit Health 2021;3(5):e317-e329 [FREE Full text] [doi: 10.1016/S2589-7500(21)00055-8] [Medline: 33890579]
23. Choi KJ, Choi JE, Roh HC, Eun JS, Kim JM, Shin YK, et al. Deep learning models for screening of high myopia using optical coherence tomography. Sci Rep 2021;11(1):21663. [doi: 10.1038/s41598-021-00622-x] [Medline: 34737335]
24. Guan J, Zhu Y, Hu Q, Ma S, Mu J, Li Z, et al. Prevalence patterns and onset prediction of high myopia for children and adolescents in southern China via real-world screening data: retrospective school–based study. J Med Internet Res 2023;25:e39507 [FREE Full text] [doi: 10.2196/39507] [Medline: 36857115]
25. Zhao H, Ma Z, Guan Y, Li J, Wei Y. High myopia detection method on fundus images based on curriculum learning. : IEEE; 2022 Presented at: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC); June 27, 2022; Los Alamitos, CA. [doi: 10.1109/compsac54236.2022.00227]
26. Dai S, Chen L, Lei T, Zhou C, Wen Y. Automatic detection of pathological myopia and high myopia on fundus images. : IEEE; 2020 Presented at: 2020 IEEE International Conference on Multimedia and Expo (ICME); July 6-10, 2020; London, United Kingdom p. 6-10 URL: https://ieeexplore.ieee.org/document/9102787 [doi: 10.1109/icme46284.2020.9102787]
27. Li L, Zhu H, Zhang Z, Zhao L, Xu L, Jonas RA, et al. Neural network–based retinal nerve fiber layer profile compensation for glaucoma diagnosis in myopia: model development and validation. JMIR Med Inform 2021;9(5):e22664 [FREE Full text] [doi: 10.2196/22664] [Medline: 34003137]
28. Lee YJ, Sun S, Kim YK, Jeoung JW, Park KH. Diagnostic ability of macular microvasculature with swept-source OCT angiography for highly myopic glaucoma using deep learning. Sci Rep 2023;13(1):5209 [FREE Full text] [doi: 10.1038/s41598-023-32164-9] [Medline: 36997639]
29. Kim JA, Yoon H, Lee D, Kim M, Choi J, Lee EJ, et al. Development of a deep learning system to detect glaucoma using macular vertical optical coherence tomography scans of myopic eyes. Sci Rep 2023;13(1):8040 [FREE Full text] [doi: 10.1038/s41598-023-34794-5] [Medline: 37198215]
30. Jeong Y, Kim YK, Jeoung JW, Park KH. Comparison of optical coherence tomography structural parameters for diagnosis of glaucoma in high myopia. JAMA Ophthalmol 2023;141(7):631-639 [FREE Full text] [doi: 10.1001/jamaophthalmol.2023.1717] [Medline: 37200038]
31. Asaoka R, Iwase A, Tsutsumi T, Saito H, Otani S, Miyata K, et al. Combining multiple HRT parameters using the 'Random Forests' method improves the diagnostic accuracy of glaucoma in emmetropic and highly myopic eyes. Invest Ophthalmol Vis Sci 2014;55(4):2482-2490. [doi: 10.1167/iovs.14-14009] [Medline: 24609628]
32. Bowd C, Belghith A, Rezapour J, Christopher M, Jonas JB, Hyman L, et al. Multimodal deep learning classifier for primary open angle glaucoma diagnosis using wide-field optic nerve head cube scans in eyes with and without high myopia. J Glaucoma 2023;32(10):841-847. [doi: 10.1097/IJG.0000000000002267] [Medline: 37523623]
33. Tang J, Yuan M, Tian K, Wang Y, Wang D, Yang J, et al. An artificial-intelligence-based automated grading and lesions segmentation system for myopic maculopathy based on color fundus photographs. Transl Vis Sci Technol 2022;11(6):16 [FREE Full text] [doi: 10.1167/tvst.11.6.16] [Medline: 35704327]
34. Zhu SJ, Zhan HD, Wu MN, Zheng B, Liu BQ, Zhang SC, et al. Research on classification method of high myopic maculopathy based on retinal fundus images and optimized ALFA-Mix active learning algorithm. Int J Ophthalmol 2023;16(7):995-1004 [FREE Full text] [doi: 10.18240/ijo.2023.07.01] [Medline: 37465510]
35. Wu Z, Cai W, Xie H, Chen S, Wang Y, Lei B, et al. Predicting optical coherence tomography–derived high myopia grades from fundus photographs using deep learning. Front Med 2022;9:842680. [doi: 10.3389/fmed.2022.842680] [Medline: 35308524]
36. Ye X, Wang J, Chen Y, Lv Z, He S, Mao J, et al. Automatic screening and identifying myopic maculopathy on optical coherence tomography images using deep learning. Transl Vis Sci Technol 2021;10(13):10 [FREE Full text] [doi: 10.1167/tvst.10.13.10] [Medline: 34751744]

37. Wang R, He J, Chen Q, Ye L, Sun D, Yin L, et al. Efficacy of a deep learning system for screening myopic maculopathy based on color fundus photographs. Ophthalmol Ther 2023;12(1):469-484 [FREE Full text] [doi: 10.1007/s40123-022-00621-9] [Medline: 36495394]
38. Wan C, Fang J, Hua X, Chen L, Zhang S, Yang W. Automated detection of myopic maculopathy using five-category models based on vision outlooker for visual recognition. Front Comput Neurosci 2023;17:1169464 [FREE Full text] [doi: 10.3389/fncom.2023.1169464] [Medline: 37152298]
39. Sun Y, Li Y, Zhang F, Zhao H, Liu H, Wang N, et al. A deep network using coarse clinical prior for myopic maculopathy grading. Comput Biol Med 2023;154:106556. [doi: 10.1016/j.compbiomed.2023.106556] [Medline: 36682177]
40. Sogawa T, Tabuchi H, Nagasato D, Masumoto H, Ikuno Y, Ohsugi H, et al. Accuracy of a deep convolutional neural network in the detection of myopic macular diseases using swept-source optical coherence tomography. PLoS One 2020;15(4):e0227240 [FREE Full text] [doi: 10.1371/journal.pone.0227240] [Medline: 32298265]
41. Du R, Xie S, Fang Y, Hagino S, Yamamoto S, Moriyama M, et al. Validation of soft labels in developing deep learning algorithms for detecting lesions of myopic maculopathy from optical coherence tomographic images. Asia Pac J Ophthalmol 2022;11(3):227-236 [FREE Full text] [doi: 10.1097/APO.0000000000000466] [Medline: 34937047]
42. Cui J, Zhang X, Xiong F, Chen CL. Pathological myopia image recognition strategy based on data augmentation and model fusion. J Healthc Eng 2021;2021:5549779 [FREE Full text] [doi: 10.1155/2021/5549779] [Medline: 34035883]
43. Burlui D, Badralexe L, Bobocescu E, Brătucu E, Popescu-Băran D, Dragoncea C. [Intestinal carcinoid]. Rev Chir Oncol Radiol O R L Oftalmol Stomatol Chir 1981;30(6):463-467. [Medline: 6211704]
44. Rauf N, Gilani SO, Waris A. Automatic detection of pathological myopia using machine learning. Sci Rep 2021;11(1):16570 [FREE Full text] [doi: 10.1038/s41598-021-95205-1] [Medline: 34400662]
45. Park SJ, Ko T, Park CK, Kim YC, Choi IY. Deep learning model based on 3D optical coherence tomography images for the automated detection of pathologic myopia. Diagnostics 2022;12(3):742 [FREE Full text] [doi: 10.3390/diagnostics12030742] [Medline: 35328292]
46. Lu L, Zhou E, Yu W, Chen B, Ren P, Lu Q, et al. Development of deep learning-based detecting systems for pathologic myopia using retinal fundus images. Commun Biol 2021;4(1):1225 [FREE Full text] [doi: 10.1038/s42003-021-02758-y] [Medline: 34702997]
47. Lu L, Ren P, Tang X, Yang M, Yuan M, Yu W, et al. AI-model for identifying pathologic myopia based on deep learning algorithms of myopic maculopathy classification and "Plus" lesion detection in fundus images. Front Cell Dev Biol 2021;9:719262 [FREE Full text] [doi: 10.3389/fcell.2021.719262] [Medline: 34722502]
48. Li Y, Feng W, Zhao X, Liu B, Zhang Y, Chi W, et al. Development and validation of a deep learning system to screen vision-threatening conditions in high myopia using optical coherence tomography images. Br J Ophthalmol 2022;106(5):633-639 [FREE Full text] [doi: 10.1136/bjophthalmol-2020-317825] [Medline: 33355150]
49. Huang X, He S, Wang J, Yang S, Wang Y, Ye X. Lesion detection with fine-grained image categorization for myopic traction maculopathy (MTM) using optical coherence tomography. Med Phys 2023;50(9):5398-5409. [doi: 10.1002/mp.16623] [Medline: 37490302]
50. Du R, Xie S, Fang Y, Igarashi-Yokoi T, Moriyama M, Ogata S, et al. Deep learning approach for automated detection of myopic maculopathy and pathologic myopia in fundus images. Ophthalmol Retina 2021;5(12):1235-1244. [doi: 10.1016/j.oret.2021.02.006] [Medline: 33610832]
51. Crincoli E, Servillo A, Catania F, Sacconi R, Mularoni C, Battista M, et al. Artificial intelligence's role in differentiating the origin for subretinal bleeding in pathologic myopia. Retina 2023;43(11):1881-1889. [doi: 10.1097/IAE.0000000000003884] [Medline: 37490781]
52. Li J, Wang L, Gao Y, Liang Q, Chen L, Sun X, et al. Automated detection of myopic maculopathy from color fundus photographs using deep convolutional neural networks. Eye Vis 2022;9(1):13 [FREE Full text] [doi: 10.1186/s40662-022-00285-3] [Medline: 35361278]
53. Liu J, Wong DWK, Zhang Z, Lu S, Li H, Saw SM, et al. Automatic classification of pathological myopia in retinal fundus images using PAMELA. 2010 Presented at: SPIE Medical Imaging 2010; February 13-18, 2010; San Diego, CA URL: https://doi.org/10.1117/12.844122 [doi: 10.1117/12.844122]
54. Liu J, Wong DWK, Lim JH, Tan NM, Zhang Z, Li H, et al. Detection of pathological myopia by PAMELA with texture - based features through an SVM approach. J Healthc Eng 2010;1(1):1-11 [FREE Full text] [doi: 10.1260/2040-2295.1.1.1]
55. Baid U, Baheti B, Dutande P, Talbar S. Detection of pathological myopia and optic disc segmentation with deep convolutional neural networks. : IEEE; 2019 Presented at: TENCON 2019—2019 IEEE Region 10 Conference (TENCON); October 17-20, 2019; Kochi, India URL: https://ieeexplore.ieee.org/document/8929252 [doi: 10.1109/tencon.2019.8929252]
56. Li S, Li M, Wu J, Li Y, Han J, Cao W, et al. Development and validation of a routine blood parameters–based model for screening the occurrence of retinal detachment in high myopia in the context of PPPM. EPMA J 2023;14(2):219-233 [FREE Full text] [doi: 10.1007/s13167-023-00319-3] [Medline: 37275550]
57. Du Y, Chen Q, Fan Y, Zhu J, He J, Zou H, et al. Automatic identification of myopic maculopathy related imaging features in optic disc region via machine learning methods. J Transl Med 2021;19(1):167 [FREE Full text] [doi: 10.1186/s12967-021-02818-1] [Medline: 33902640]

58. Kim YC, Chang DJ, Park SJ, Choi IY, Gong YS, Kim HA, et al. Machine learning prediction of pathologic myopia using tomographic elevation of the posterior sclera. Sci Rep 2021;11(1):6950 [FREE Full text] [doi: 10.1038/s41598-021-85699-0] [Medline: 33772040]
59. Zhang Z, Xu Y, Liu J, Wong DWK, Kwoh CK, Saw SM, et al. Automatic diagnosis of pathological myopia from heterogeneous biomedical data. PLoS One 2013;8(6):e65736 [FREE Full text] [doi: 10.1371/journal.pone.0065736] [Medline: 23799040]
60. Hou XW, Yang JLX, Li DL, Tao YJ, Ke CF, Zhang B, et al. Machine learning–based integration of metabolomics characterisation predicts progression of ,mopic retinopathy in children and adolescents. Metabolites 2023;13(2):301 [FREE Full text] [doi: 10.3390/metabo13020301] [Medline: 36837920]
61. He X, Ren P, Lu L, Tang X, Wang J, Yang Z, et al. Development of a deep learning algorithm for myopic maculopathy classification based on OCT images using transfer learning. Front Public Health 2022;10:1005700 [FREE Full text] [doi: 10.3389/fpubh.2022.1005700] [Medline: 36211704]
62. Prashar J, Tay N. Performance of artificial intelligence for the detection of pathological myopia from colour fundus images: a systematic review and meta-analysis. Eye 2024;38(2):303-314 [FREE Full text] [doi: 10.1038/s41433-023-02680-z] [Medline: 37550366]
63. Wang YX, Yang H, Wei CC, Xu L, Wei WB, Jonas JB. High myopia as risk factor for the 10-year incidence of open-angle glaucoma in the Beijing eye study. Br J Ophthalmol 2023;107(7):935-940 [FREE Full text] [doi: 10.1136/bjophthalmol-2021-320644] [Medline: 35193858]
64. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol 2018;136(7):803-810 [FREE Full text] [doi: 10.1001/jamaophthalmol.2018.1934] [Medline: 29801159]
65. Mao J, Luo Y, Liu L, Lao J, Shao Y, Zhang M, et al. Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks. Acta Ophthalmol 2020;98(3):e339-e345 [FREE Full text] [doi: 10.1111/aos.14264] [Medline: 31559701]
66. Han Y, Li W, Liu M, Wu Z, Zhang F, Liu X, et al. Application of an anomaly detection model to screen for ocular diseases using color retinal fundus images: design and evaluation study. J Med Internet Res 2021;23(7):e27822 [FREE Full text] [doi: 10.2196/27822] [Medline: 34255681]
67. Yang Y, Li R, Lin D, Zhang X, Li W, Wang J, et al. Automatic identification of myopia based on ocular appearance images using deep learning. Ann Transl Med 2020;8(11):705 [FREE Full text] [doi: 10.21037/atm.2019.12.39] [Medline: 32617325]
68. Chun J, Kim Y, Shin KY, Han SH, Oh SY, Chung TY, et al. Deep learning-based prediction of refractive error using photorefraction images captured by a smartphone: model development and validation study. JMIR Med Inform 2020;8(5):e16225 [FREE Full text] [doi: 10.2196/16225] [Medline: 32369035]

Abbreviations

CNN: convolutional neural network
DCNN: deep convolutional neural network
DL: deep learning
DOR: diagnostic odds ratio
LR: logistic regression
MeSH: Medical Subjects Headings
ML: machine learning
NLR: negative likelihood ratio
OCT: optical coherence tomography
PLR: positive likelihood ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RF: random forest
SROC: summary receiver operating characteristic
SVM: support vector machine

Edited by A Mavragani; submitted 22.02.24; peer-reviewed by TE Komolafe, S Hansun; comments to author 20.05.24; revised version received 02.07.24; accepted 06.11.24; published 03.01.25.

Please cite as:
Zuo H, Huang B, He J, Fang L, Huang M
Machine Learning Approaches in High Myopia: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e57644
URL: https://www.jmir.org/2025/1/e57644
doi:10.2196/57644
PMID:

Review

Social Media Potential and Impact on Changing Behaviors and Actions in Skin Health Promotion: Systematic Review

Justyna Martyna Brzozowska[1]; Joanna Gotlib[2]

[1]School of Medical & Health Sciences, University of Economics and Human Sciences in Warsaw, Warsaw, Poland

[2]Department of Education and Research in Health Sciences, Faculty of Health Sciences, Medical University of Warsaw, Warsaw, Poland

Corresponding Author:
Justyna Martyna Brzozowska
School of Medical & Health Sciences
University of Economics and Human Sciences in Warsaw
59 Okopowa Street
Warsaw, 01-043
Poland
Phone: 48 604441820
Email: justyna@kardas.pl

Abstract

Background: Social media is used as a tool for information exchange, entertainment, education, and intervention. Intervention efforts attempt to engage users in skin health.

Objective: This review aimed to collect and summarize research assessing the impact of social media on skin health promotion activities undertaken by social media users.

Methods: In accordance with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines, the following scientific databases were searched: Scopus, Web of Science, PubMed, Academic Search Ultimate (via EBSCO), Academic Research Source eJournals (via EBSCO), ERIC (via EBSCO), Health Source: Consumer Edition (via EBSCO), and Health Source: Nursing/Academic Edition (via EBSCO). Using ProQuest Dissertations and Theses, OpenGrey, Grey Literature Report, and MedNar, the search was supplemented with gray literature. Articles on skin care, skin health, skin diseases, skin protection, and educational activities promoting healthy skin on social media were selected for review (search date: February 6, 2023). The following qualification criteria were used: original research; research conducted on social media; and research topics regarding educational activities in skin health promotion, skin care, skin health, skin diseases, and skin protection. To assess the risk of bias, the following tools were used: the Cochrane Collaboration tool for risk-of-bias assessment (randomized controlled trials and quasi-experimental studies) and the Centre for Evidence-Based Medicine checklist (cross-sectional studies).

Results: Altogether, 1558 works were considered, of which 23 (1.48%) qualified, with 3 (13%) studies on acne and 20 (87%) on skin cancer, sunscreen, and tanning. Social media interventions were dealt with in 65% (15/23) of the studies. The review made it possible to investigate cognitive and cognitive-behavioral interventions. In both observational and interventional studies, the most frequently discussed topics were skin exposure and protection against UV radiation and skin cancer. The analyzed research showed that social media is a source of information. Visualization has a strong impact on users. The involvement of social media users is measured through the amount of content shared and contributes to changing attitudes and behaviors regarding skin health.

Conclusions: This review outlined the impact of social media, despite its heterogeneity, on users' skin health behaviors, attitudes, and actions. It identified strategies for digital interventions to promote skin health. In health sciences, a standardized tool is needed to assess the quality of social media digital interventions. This review has several limitations: only articles written in English were considered; ongoing studies were omitted; and there was a small number of interventional studies on acne and a lack of research on daily skin care, education, and antiaging activities on social media. Another limitation, resulting from the topic being too broad, was a failure to perform quantitative data analysis, resulting in the studies that qualified for the review being heterogeneous.

(J Med Internet Res 2025;27:e54241) doi:10.2196/54241

KEYWORDS

skin; social media; prevention; behavioral intervention; skin cancer; sun protection; acne

Introduction

Background

For many people, social media platforms are now an integral part of daily life. Social media browsing is the most popular online activity. According to the statistics, >4.26 billion people used social media in 2021, and this number is likely to grow to almost 6 billion in 2027 [1]. In April 2023, a total of 4.8 billion people, or 59.9% of the world's population, were social media users [2].

Social media is a powerful means of communication, mainly used for an exchange of information and for entertainment. In addition, companies use it for marketing purposes to promote products and services. Currently, there is also a noticeable trend to develop educational platforms, especially on health education. According to Savas et al [3], because of an increase in the number of patients looking for medical information, physicians can use social media such as Facebook as a significant educational tool. Gantenbein et al [4] argue that medical information available on the internet and social media could be useful for most dermatological patients. The authors identified the most important needs of patients: online consultations, medical content on YouTube, and a possibility to chat with a specialist [4]. In turn, according to Banerjee et al [5], social media is a platform for increasing people's commitment and providing information about UV protection, skin cancer prevention, and its early detection.

The aforementioned research confirms the informational and educational role of social media. The aim of information activities is to increase knowledge about skin cancer and increase awareness of the risks of UV radiation. The effect of these activities should be a change in health behaviors (regular skin photoprotection, regular skin self-examination, and regular monitoring of skin lesions by a physician). Engagement in systematic activities by social media users is crucial.

In addition to its educational role, social media is a tool that supports life and work. Rew et al [6] created a prototype of an application based on content concerning social media users' daily activities. The application can assess the condition of the skin and provides advice on improving it, with suggestions for care products.

In addition, social media can be a motivational tool by motivating systematic care and protection activities aimed at healthy and well-groomed skin.

Focusing mainly on the analysis of social media content, previous reviews have discussed their impact on skin care [7] and skin cancer prevention [8]. They have also discussed the effects of social networking sites on health behavior change [9] and on available digital interventions, mainly stressing the importance of sun protection and skin self-examinations [10].

Unlike previous systematic reviews, we sought to understand the nature of motivation to act. We were interested in the following aspects: (1) What dermatological problem that is talked about on social media is the main area of interest for researchers? (2) What drives social media audiences to take goal-directed actions? (3) What is the level of involvement of social media users in achieving this goal? (4) What contributes to perseverance in pursuing a goal? The goal is healthy and well-groomed skin. To achieve this goal, health behavior changes must be made.

We wanted to draw attention to the role of social media as a tool for collecting data and a tool that encourages users to take actions to care for their skin (including preventing the development of skin cancer and alleviating the symptoms of dermatological diseases).

The main parts of skin health preventive measures are promoting skin care, skin protection, and control routines. The involvement and motivation of social media users is important. Therefore, research is needed to conduct a systematic review that evaluates the impact and effectiveness of social media in this field. In addition, such research should not only focus on the information provided by social media but also try to engage users in actions aimed at skin health.

This review helps understand and explores what is already known about social media potential and its impact on users' skin care behaviors and skin health preventive measures. This study suggests the need to optimize social media interventions and develop new studies that will assess long-term change in skin health behaviors. In addition, by increasing the importance of skin protection and its diseases, this review has practical significance. The results of this research can be helpful in practice and in research investigating the role of social media in preventive health care. This review may support the work of dermatologists, cosmetologists, and health educators because it provides information they need, sometimes on complex topics, in a simple and easily available way.

The novelty of this review is that, through the heterogeneity of the research, a broader view of the impact of social media on skin health behaviors is presented. This study provides new observations about the prespecified desired outcome to be achieved. It complements the theory of engagement with digital interventions.

This study differs from previous ones in that it presents social media as an important digital stimulus. It affirms the integration of motivation into specific physical, affective, and cognitive action. The empirical observations concerning, on the one hand, the demonization of the effects of the sun on the skin and, on the other, the image of an attractive tan, which occur in a cause-effect conflict, are surprising.

This review proves that the image, not just the content, is of great importance. There is a noticeable trend in which the image becomes more important than scientific reports. This review also presents the negative side of social media.

Objectives

This paper presents a systematic review of the literature investigating the role of social media platforms as an educational and intervention tool in promoting preventive skin health measures. The aim of this review was to collect and summarize research assessing the impact of social media on skin health promotion activities undertaken by social media users. In the analyzed studies, we focused on the reception of information about the skin (eg, skin cancer, acne, and tanning) by social media users. The literature review comprised (1) studies assessing the impact of the content of social media on its users, (2) social media intervention studies, and (3) studies determining the impact of posts generated by users on their health decisions and their self-monitoring.

The following research questions were formulated:

1. What are the existing data on the use of social media as an educational intervention tool in skin health preventive measures?
2. What activities on social media can be effective in promoting skin health?
3. What intervention strategies are disseminated via social media and what is their role in promoting skin health?

Methods

Design

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines [11,12] and with the Synthesis Without Meta-Analysis reporting guideline [13]. The PRISMA guidelines ensure the highest quality in scientific research. To more fully present the results of the review, a narrative form was additionally adopted.

Information Sources and Search Strategies

This review was conducted using the following databases: Scopus, Web of Science, PubMed, Academic Search Ultimate (via EBSCO), Academic Research Source eJournals (via EBSCO), ERIC (via EBSCO), Health Source: Consumer Edition (via EBSCO), and Health Source: Nursing/Academic Edition (via EBSCO). In addition, the search was supplemented with gray literature using ProQuest Dissertations and Theses, OpenGrey, Grey Literature Report, and MedNar. A reference list search for relevant articles was also conducted.

The search strategy was first constructed by the authors and then consulted with a librarian experienced in scientific information extraction. The developed phrase was as follows:

> *((“social media” OR “Facebook” OR “Instagram” OR “Twitter” OR “Pinterest” OR “YouTube” OR “TikTok” OR “SnapChat”) AND (“skincare” OR “skin care” OR “skin health” OR “skin diagnosis” OR “skin type” OR “beautician” OR “cosmetologist” OR “cosmetician” OR “cosmetic consultation*” OR “cosmetic dermatology” OR “cosmetic product*” OR “dermocosmetic*” OR “cosmetic procedure*” OR “skin cleans*” OR “skin moistur*” OR “skin massage” OR “facial cleans*” OR “facial moistur*” OR “facial massage” OR “photoprotection” OR “sunscreen*” OR “sun protection” OR “skin hydration” OR “skin rejuvenation” OR “skin cancer” OR “skin prevention” OR “acne” OR “rosacea” OR “anti-acne” OR “anti-aging” OR “anti-wrinkle” OR “skin aging” OR “oily skin” OR “dry skin” OR “combination skin” OR “sensitive skin” OR “pigmented skin” OR “pimple*” OR “pustule*” OR “papules” OR “blackheads”))*

The search was conducted on February 6, 2023, and on the same day, the records were imported from the databases into the Mendeley reference manager (Elsevier). The search was not limited to a period to obtain as many results as possible.

Study Selection

After applying the search strategy, the studies were imported and saved to the Mendeley reference manager, and then duplicates were removed. The selection of the studies consisted of 2 stages. The first stage was conducted by one of the authors (JMB), who considered the title and abstract to identify and exclude irrelevant studies. To reach a mutual agreement, the other author (JG) reviewed a randomly selected sample containing 10% of the titles and abstracts. When, based on the title and abstract, the decision whether to include a work was difficult, the full article was then downloaded. The second stage of the selection consisted of an analysis of the full texts of the articles by each author (JMB and JG) separately to determine whether they could be included, and the reasons for exclusion were listed. When the decision to include or exclude a study was unclear, the authors clarified ambiguities and different opinions through discussions. This stage of the article selection process required both authors to be present to resolve voting conflicts. All studies were assessed using the inclusion and exclusion criteria described in the next section.

Inclusion and Exclusion Criteria

All the inclusion and exclusion criteria were defined a priori (Textbox 1).

In particular, the following papers were excluded during the first stage of the selection process: papers written in a language other than English, papers not related to topics of skin care and skin health promotion on social media, conference papers, information from beauty and health blogs, posters of prevention campaigns, and interviews and articles from popular science magazines advertising a specific cosmetic brand. In addition, articles on social media safety rules, product protection, ethics, diet, COVID-19, and tele-dermatology and materials related to patient consent to post photos on social media were eliminated. During the second stage, articles were allocated to the following groups: articles fully meeting the inclusion criteria, comments, case studies, review papers, research protocols, papers in which skin care and skin health preventive measures were not the main topic, papers in which social media was not the main topic (eg, it was used to recruit study participants or to disseminate survey results), papers on plastic surgery and body image perception on social media, papers on social media marketing (eg, dealing with the impact of advertisements encouraging consumers to purchase cosmetic products or with methods of promoting the sale of cosmetic products or assessing the satisfaction with cosmetic products and with brand image building of skin care products), papers on the psychological effects of social media (eg, self-acceptance of social media users; self-confidence; and the relationship among emotions, emotional states, and the use of social media), papers on social media search for medical information by dermatology patients, and papers focusing exclusively on the analysis of the content most frequently shared on social media (ie, posts and videos).

Articles with educational value and assessing the impact of social media skin-related information on its users were included. This review also included studies that investigated social media interventions (articles on behavioral interventions aimed at skin health promotion, which were included during the primary analysis) and studies that investigated the impact of social media data on users (articles on skin health promotion, which were included during the secondary analysis).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Papers based on original research (ie, research articles)
- Papers on skin care, skin health, skin diseases, and skin protection and papers on educational activities directed at skin health preventive measures
- Papers on social media (eg, Facebook, Instagram, Twitter [subsequently rebranded X], Pinterest, YouTube, TikTok, and Snapchat)

Exclusion criteria

- Papers written in a language other than English
- Review papers
- Comment papers
- Case studies
- Conference papers
- Clinical trial reports
- Information from beauty and health blogs, popular science magazines, interviews, and posters from prevention campaigns
- Papers in which social media coverage of skin care, skin health, and preventive actions were not the main topic

Data Extraction

Research information from the included articles downloaded during the search process was rechecked by both authors (JMB and JG) to ensure that it met the inclusion criteria. The following details were extracted from each paper: authors, year of publication, geographic location of data collection (country where the research took place), type of social media platform, form of information, number and characteristics of study participants, type of study or data collection methods, and main topic and purpose of the research (Multimedia Appendix 1 [14-36]). In addition, one author (JMB) developed a form with questions (Textbox 2) in which both authors independently recorded their responses (summarized in Multimedia Appendix 2 [14-36]). The PRISMA checklist is provided in Multimedia Appendix 3.

Textbox 2. Questions related to qualified publications.

What are the authors and the year of publication?

- Is intervention a topic of the study?
- Does the study evaluate the impact of social media on users?
- Does the study deal with public health campaigns?
- Does the study mention skin self-examination?
- Does the study mention the use of cosmetic materials (eg, sunscreen)?
- Does the study contain a statistical analysis?
- What are the main results of the study?
- What are the conclusions or opinions of the authors on the effectiveness of information delivery strategies on social media?

Quality Assessment

The methodological quality of each study included in this review was assessed separately by both authors (JMB and JG), and their opinions were compared. In case of disagreement, the authors discussed their doubts and came to a consensus. The Cochrane Collaboration tool for risk-of-bias assessment was used to assess the risk of bias in randomized controlled trials and quasi-experimental studies [37]. A total of 6 domains were evaluated, with the plus sign (+) representing a low risk of bias, the en dash (–) representing a high risk of bias, and both (+ and –) representing an unclear risk of bias. Blinding of participants and staff was not assessed as this was considered impossible in some studies. A study that received a score of ≥4 plus signs was considered to be of high quality.

For cross-sectional studies, the Centre for Evidence-Based Medicine checklist for critical appraisal was used [38]. It consists of 12 questions with possible answers of "yes," "cannot tell," or "no." A cutoff value of 75% was used for the assessment. This means that, with ≥9 affirmative responses, the study was considered to be of high quality. Otherwise, the study was considered to be of low quality.

Data Synthesis Strategy

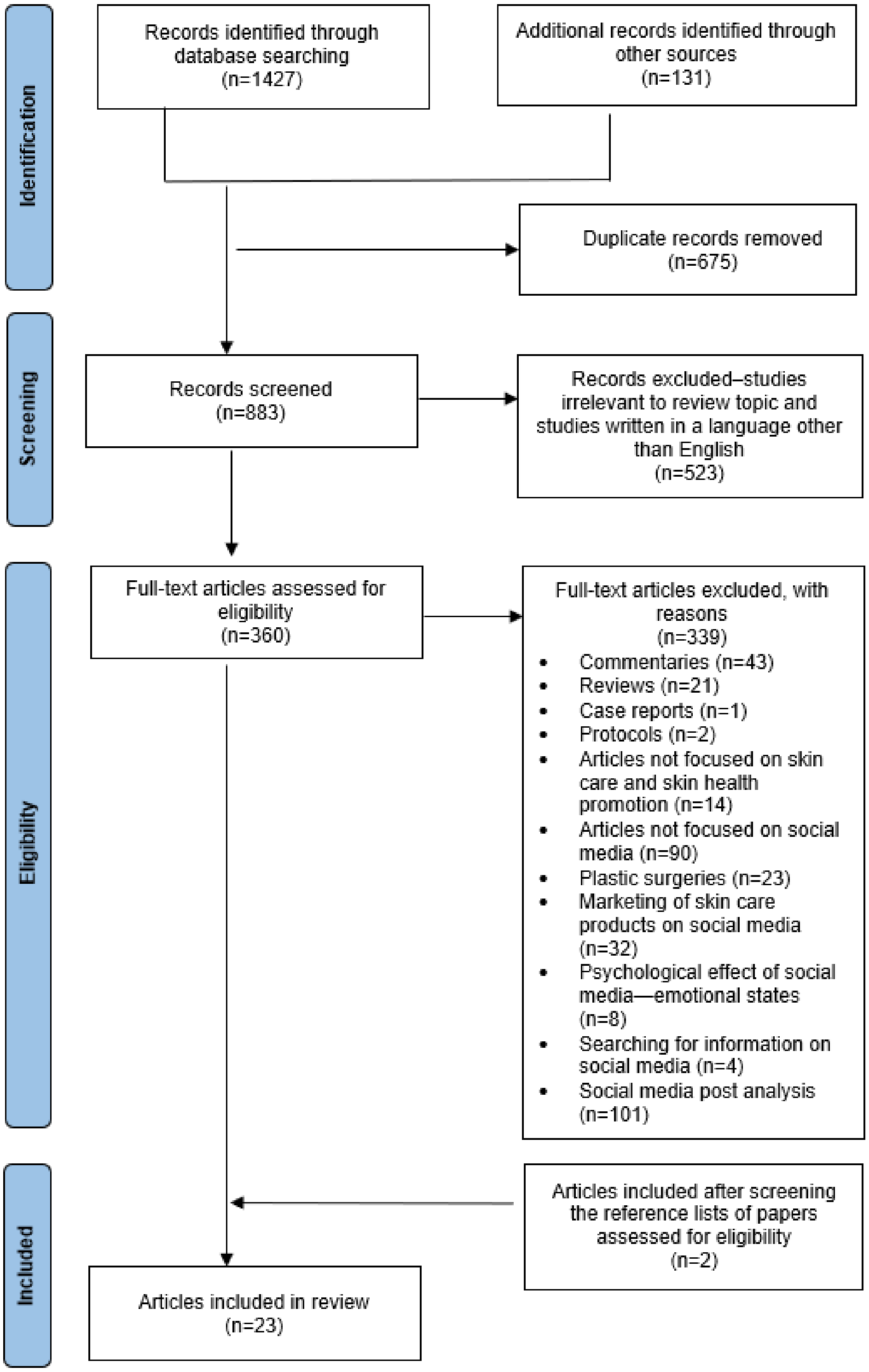
A narrative synthesis of the included studies was performed in accordance with the recommendations by Popay et al [39]. During data analysis, the authors considered the possibility of classifying the included publications according to the health problem. However, after reanalysis of the selected publications, the authors decided that a better conceptual model was to group them according to the type of intervention, if it was used.

Results

Search Results

A literature search revealed 1558 records (Scopus: n=485, 31.13%; Web of Science: n=305, 19.58%; PubMed: n=270, 17.33%; Academic Search Ultimate via EBSCO: n=234, 15.02%; Academic Research Source eJournals via EBSCO: n=26, 1.67%; ERIC via EBSCO: n=10, 0.64%; Health Source: Consumer Edition via EBSCO: n=35, 2.25%; Health Source: Nursing/Academic Edition via EBSCO: n=62, 3.98%; ProQuest Dissertations and Theses: n=22, 1.41%; OpenGrey: n=4, 0.26%; Grey Literature Report: n=0; MedNar: n=105, 6.74%). During the identification stage, 91.59% (1427/1558) of the records were obtained from scientific databases, and 8.41% (131/1558) of the records were obtained from gray literature. After removing 43.32% (675/1558) of duplicates, 883 records remained. After reviewing the titles and abstracts of these 883 records, 523 (59.2%) were excluded because they were irrelevant to the review topic or were written in a language other than English. A total of 40.8% (360/883) of the publications were selected for full-text download. After analyzing the full texts of the articles, 339 were excluded (n=43, 12.7% comments; n=21, 6.2% reviews; n=1, 0.3% case reports; n=2, 0.6% protocols; n=14, 4.1% articles not focused on skin care and skin health promotion; n=90, 26.5% articles not focused on social media; n=23, 6.8% articles about plastic surgeries; n=32, 9.4% articles about marketing of skin care products on social media; n=8, 2.4% articles about the psychological effect of social media—emotional states; n=4, 1.2% articles about searching for information on social media; and n=101, 29.8% articles about social media post analysis). These studies did not meet the inclusion criteria. A total of 21 articles met the inclusion criteria. In addition, after a detailed analysis of the eligible studies and their references, 2 publications were added. In total, 23 publications were included in this review (Figure 1).

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



Study Characteristics

The detailed characteristics and main results of the studies included in this systematic review are presented in Multimedia Appendix 1 [14-36] and Multimedia Appendix 2 [14-36]. Most of the research was conducted using Facebook. Of the 23 studies, most (n=14, 61%) were conducted in the United States [14-27], with a smaller number conducted in other countries: the United Kingdom (n=3, 13%) [28-30], Australia (n=3, 13%) [31-33], the Netherlands (n=1, 4%) [34], Denmark (n=1, 4%) [35], and Saudi Arabia (n=1, 4%) [36]. The highest number of studies was published in 2022 (5/23, 22%) [16,17,21,22,36], 17% (4/23) of the studies were published in 2020 [19,27,28,30], 13% (3/23) were published in 2019 [20,26,32], 13% (3/23) were published in 2018 [14,18,24], 13% (3/23) were published in 2017 [29,31,34], 9% (2/23) were published in 2021 [15,25], 9% (2/23) were published in 2011 [33,35], and 4% (1/23) were published in 2016 [23]. The oldest included publications dated back to 2011 (2/23, 9%) [33,35].

The vast majority of the studies were on skin cancer, sunscreen, and tanning (20/23, 87%), with only 13% (3/23) of the studies on acne [19,27,36]. A total of 17% (4/23) of the studies dealt with skin self-examination [18,19,30,34]. On the other hand, 43% (10/23) of the studies mentioned the use of cosmetic products [14,18,21,22,25,27,28,31,33,36], whereas 35% (8/23) of the studies recommended sunscreen use [14,18,21,22,25,28,31,33]. In addition, 43% (10/23) of the studies discussed the impact of public health campaigns [15-17,20,22,28-30,33,35]. It is worth noting that several publications discussed the results of the same research (eg, 3/23, 13% of the studies by Buller et al [15-17])

Among the 23 studies included in this review, social media interventions were the topic of 15 (65%) [15-18,20-22,24,25,28,29,32-35], among which cognitive effects were assessed by Agha-Mir-Salim et al [28], Damude et al [34], Gough et al [29], and Mingoia et al [32]. On the other hand, both cognitive and behavioral effects were assessed in the studies by Buller et al [15-17], Coups et al [18], Køster et al [35], Morrison et al [20], Myrick et al [21], Pagoto et al [22], Potente et al [33], Stapleton et al [24], and Vraga et al [25]. Comparison groups were used by Agha-Mir-Salim et al [28], Buller et al [15-17], Gough et al [29], Køster et al [35], Mingoia et al [32], Morrison et al [20], Myrick et al [21], Pagoto et al [22], Potente et al [33], and Vraga et al [25].

Quality Appraisal

The risk-of-bias assessment for randomized controlled trials is summarized in Table 1. A total of 35% (8/23) of the studies turned out to be of high quality [15-17,20,22,25,28,32], and 4% (1/23) of the studies were of low quality [29]. A total of 57% (13/23) of cross-sectional studies were of high quality [14,18,19,21,23,24,26,27,30,31,33,35,36], and 4% (1/23) of cross-sectional studies were of low quality [34]. A summary is provided in Multimedia Appendix 2 [14-36].

Table 1. Cochrane risk-of-bias assessment.

Study	Random sequence generation	Allocation concealment	Blinding of participants and personnel or blinding of outcome assessment	Incomplete outcome data	Selecting reporting	Other bias
Agha-Mir-Salim et al [28], 2020	–[a]	+[b]	–	+	+	+
Buller et al [15], 2021	+	+	+	+	+	– (publishing the results of the same study multiple times)
Buller et al [17], 2022	+	+	+	+	+	– (publishing the results of the same study multiple times)
Buller et al [16], 2022	+	+	+	+	+	– (publishing the results of the same study multiple times)
Gough et al [29], 2017	+	–	–	+	+	– (IP addresses were not checked for duplicate users)
Mingoia et al [32], 2019	+	+	–	+	+	+
Morrison et al [20], 2019	+	+	+	+	+	+
Pagoto et al [22], 2022	+	+	+	+	+	+
Vraga et al [25], 2022	+	+	+	+	+	+

[a] High risk of bias.

[b] Low risk of bias.

Data Synthesis

Overview

Among the assumptions we made for this review was that skin health promotion on social media is a broad concept that includes regular activities aimed at protecting the skin against the risk of developing skin cancer, as well as promoting skin care that can ensure good skin condition, prevent skin diseases, slow down the skin aging process, and keep the skin looking nice, which can contribute to one's well-being. However, most of the studies included in this review (20/23, 87%) focused on behaviors associated with skin cancer risk and prevention (exposure to and skin protection from UV radiation). Skin cancer was the main area of interest for researchers, undoubtedly because it is a high-profile problem. There were not enough studies on other skin conditions to draw firm conclusions about the effects of social media. Nevertheless, it is possible to analyze and compare the effect of information and motivational activities on social media on the engagement of users and their behavior changes and actions taken. This will enable social media moderators to develop overall motivational intervention strategies.

This review included observational studies in which social media was used as a tool to disseminate information about the skin, as well as intervention studies in which social media content was manipulated. Despite the fundamental differences in the methodology of these 2 types of research, the main goal was the same—skin health promotion. Due to differences in research methodology, comparison and summary were difficult.

The results are presented in the following sections in narrative form. A division was made into studies that did not use an intervention and studies that did use an intervention. Then, to synthesize and integrate the results well and draw consistent conclusions, observational studies and then interventional studies were summarized and compared. Attention was drawn to the limitations of observational studies compared to interventional studies.

Assessing the Impact of Social Media on Skin Care and Skin Health Promotion Without an Intervention

The studies included in this review argued that social media is a useful source of knowledge about skin health. According to Bahaj et al [36], most individuals aged between 18 and 25 years choose social media as their first source of acne treatment advice. However, much of social media advice does not follow the current guidelines of the American Academy of Dermatology [40]. The aforementioned authors also found that 74.1% of those social media users had a college degree. There was a positive link between educational attainment and willingness to use social media for advice (*P*=.002). In addition, a significant link between the severity of acne and search for treatment advice on social media was found (*P*<.001) [36].

Similarly, in the study by Yousaf et al [27] with 130 participants with acne, 45% (n=58) of them used social media for advice on its treatment. Of these 58 participants, 72% were women (n=42), who were 75% more likely to use social media for advice on acne treatment (prevalence ratio=1.75, 95% CI 1.11-2.76; *P*=.01). In addition, only 31% of the respondents who consulted their problem on social media managed their condition in line with the American Academy of Dermatology clinical guidelines [27]. However, both Bahaj et al [36] and Yousaf et al [27] observed that the advice available on social media did not comply with the guidelines of the American Academy of Dermatology [40]. In the study by Basch et al [14] on skin cancer and tanning, it was found that people who did not fully adhere to proper sunscreen behavior were more likely than those who followed sun protection rules to believe that social media was an accurate source of health information (10.5% vs 3.3%; *P*=.046) or a helpful source of health information (62.5% vs 46.7%; *P*=.02).

According to Mingoia et al [31], the level of social media use was significantly and positively correlated with users' sun exposure (*P*<.001), sunburn (*P*<.001), and dissatisfaction with skin tone and negatively correlated with sun protection (*P*<.001). This meant that viewing photos, posting photos, sharing content, and the number of likes were significantly linked with greater sun exposure, lower sun protection, and greater dissatisfaction with skin tone. It is worth mentioning that the aforementioned study drew attention to the fact that women, to a greater extent than men, are image oriented and more attracted to visualization on social media. For them, the use of images in health messages may be effective in preventive skin health care campaigns [31].

In the study by Stapleton et al [23], 45.8% of participants had a college degree, and 14.8% had a bachelor's degree or higher (a sample of 463 participants). The authors found that higher rates of indoor tanning were associated with Twitter (subsequently rebranded X) and Instagram use among a sample of young adult women. As a result, the authors argued that the aforementioned social media platforms could be a valuable way to provide information about skin cancer prevention among compulsive indoor tanners [23].

In contrast, the sample in the study by Willoughby and Myrick [26] consisted of 502 female college students aged between 18 and 29 years, and it was found that participants who reported more frequent use of social media were more likely to sunbathe outside. Thus, the researchers found that greater use of visual social media platforms such as Instagram, Snapchat, and Pinterest was likely to increase tanning [26].

According to the studies by Bahaj et al [36], Yousaf et al [27], Basch et al [14], Mingoia et al [31], Stapleton et al [23], and Willoughby and Myrick [26], widely shared social media content could influence engagement in risky health behaviors, affecting users' knowledge, attitudes, and actions. Despite the negative impact of social media, the authors noted that it could be a powerful informational, educational, and intervention tool for preventive skin health care campaigns. In addition, the identification of the target group of health advice recipients made it possible to fight disinformation using specific prevention channels (ie, social media platforms). In contrast to the aforementioned studies, the studies by Guckian et al [30] and Martel et al [19] mainly pointed to the positive impact of social media.

Guckian et al [30] found that interventions via social media could have a positive impact on health-related behaviors. According to the authors, social media could motivate patients with distressing cutaneous lesions to visit melanoma screening clinics. However, the authors found that the most common motivating reason for visiting a melanoma screening clinic was skin self-examination. In this study, 65% (162/249) of the patients were regular users of social media, 33% (83/249) saw posts on the internet about cutaneous lesions, 10% (8/83) saw posts from physicians, 36% (30/83) saw posts from health authorities, and 37% (31/83) saw posts from other people (social media users). In addition, 33% (83/249) of patients searched for information about their cutaneous lesions on the internet, whereas one patient did so on social media. In addition, 24% (6/25) of patients said that more posts on cutaneous lesions should be shared [30].

Martel et al [19] found that social media use could be a motivating factor. According to the authors, editing one's own cutaneous lesions on social media, in particular acne and acne scars, contributed to an increase in users' awareness and encouraged them to seek dermatological care. Of 145 people who edited their skin lesions, as many as 128 (88.3%) edited acne and acne scars. This editing increased their awareness of the need for dermatological care (*P*=.02) [19].

Assessing the Impact of Social Media on Skin Care and Skin Health Promotion After an Intervention

Comparative Studies on the Role of Digital Social Media Interventions and Traditional Media in Health Promotion

The study by Agha-Mir-Salim et al [28] was one of those comparing the effects of digital interventions with those of printed materials. The research involved the SunSafe campaign that aimed to raise awareness of sun exposure risk, melanoma, and the need for sun protection among individuals aged 18 to 29 years. Using flyers and Facebook, the authors found that, after the intervention, the average knowledge score improved in both groups (Facebook=1.82; flyers=3.04; $P<.001$). However, printed flyers turned out to be more effective than Facebook posts (95% CI 0.35-2.09; $P=.006$) [28].

Social Media Interventions as an Educational Strategy

Damude et al [34] assessed the knowledge of patients with melanoma about their illness and their opinion on multiple ways of providing information. YouTube instructional videos presented in the research were designed to stress the need for self-examination of the skin and lymph nodes and demonstrate how to do this properly. It turned out that 63% of patients preferred to receive information in multiple ways, 92% of them preferred verbal instructions from their physician, 62% preferred educational YouTube videos, and 43% preferred to receive instructions via brochures. In addition, the authors concluded that there was an urgent need for educational activities focused on melanoma and for supplementing the methods of education with instructional YouTube videos [34].

In the study on skin cancer prevention by Gough et al [29], a trend toward increasing knowledge about preventive actions in response to social media educational interventions was observed. Those interventions increased awareness that fair-skinned people required the most protection and that skin cancer was the most common form of cancer, with melanoma being its most serious type. In addition, the authors noted a trend of increasing awareness of sun and UV ray exposure, skin cancer risk, and the need for sun protection.

Using an educational strategy and cognitive dissonance, Mingoia et al [32] investigated the effects of a social media intervention on the opinion that *tanned skin is desirable and beneficial for appearance*. The educational strategy was aimed at reducing positive attitudes toward sunbathing and improving the ability to critically analyze social media content [32].

The study by Pagoto et al [22] is another example of a dissonance-based social media intervention to promote sun safety. The intervention involved encouraging survey respondents to create social media posts on healthy skin or a healthy lifestyle. According to the authors, participants in the healthy skin intervention group reported a higher motivation to use protective clothing and sunscreen and a decreased motivation to tan. The authors suggested that this intervention, encouraging social media content creation and response, should enhance cognitive dissonance and affect social norms, which in turn might change attitudes and behaviors concerning tanning [22].

The study by Stapleton et al [24] is another example of a dissonance-based intervention focusing on healthy body image. As part of this intervention, a group of Facebook users was encouraged to publish posts and share their opinions on them. The aim of the dissonance-inducing intervention approach was to engage participants in group discussions related to body image [24].

The topic of skin cancer prevention and the use of tanning beds was also addressed by Morrison et al [20]. To educate the audience, the authors used 3 different short videos aimed at discouraging indoor tanning. The first was a humorous video created in collaboration with a physician, the second was a music video with an Instagram star, and the third one was based on facts about skin cancer. The humorous video received the highest engagement, measured via the number of comments, reactions, and shares, whereas the fact-based video received the lowest engagement [20].

Moreover, in the study by Potente et al [33], an ironic music video was also used, the aim of which was to engage young people in skin cancer prevention. Different forms of entertainment, education, and marketing were used on social media to change attitudes and behaviors regarding sun protection [33].

In contrast to the aforementioned studies, Vraga et al [25] used a video with incorrect information about sunscreen and skin cancer. The results indicated that the disinformation video, compared to an educational video, increased belief in sunscreen myths ($P<.001$) and decreased intent to use sunscreen ($P<.01$) [25].

In the study by Myrick et al [21], an intervention focusing on appearance and self-control–related emotions was used. The results indicated that the intervention focused on appearance contributed to a reduction in the time spent viewing Instagram images of tanned women. In contrast, after the intervention involving self-control emotions, an increase in anticipated pride and in a positive attitude toward solar safety was observed [21].

Digital Interventions on Social Media Based on Social Cognitive Theory, Transportation Theory, and Diffusion of Innovations Theory

Buller et al [15-17] conducted a year-long social media health campaign to reduce mothers' permissiveness of indoor tanning with their teenage daughters. Mothers became less lenient toward indoor tanning with their teenage daughters immediately after the campaign and 6 months after it ended [15,16]. According to the researchers, social media campaigns could motivate mothers to communicate with their daughters and share posts about the danger of indoor tanning [15]. In turn, Buller et al [17] measured engagement levels by measuring reactions (eg, sadness and number of likes) and mothers' comments on posts related to the campaign. They noted that users reacted with comments to 76.4% of posts. Immediately after the campaign ended and 32 months after, mothers who posted reactions or comments were less liberal about indoor tanning with their teenage daughters than mothers who did not engage with the posts [17].

Social Media Behavioral Interventions

In the study by Coups et al [18], a behavioral intervention was conducted via Facebook targeting young patients with melanoma and their families. The content of the intervention consisted of daily posts about skin cancer risk, total cutaneous exam, skin self-examination, and preventive measures (ie, sun protection). Preliminary analyses by the researchers showed an increase in intentions to use the total cutaneous exam, skin self-examination, and sun protection. In addition, Coups et al [18] found that posts that included personal stories, quizzes, skin care exercises, and an opportunity to ask a question to a specialist were perceived by survey respondents as particularly engaging.

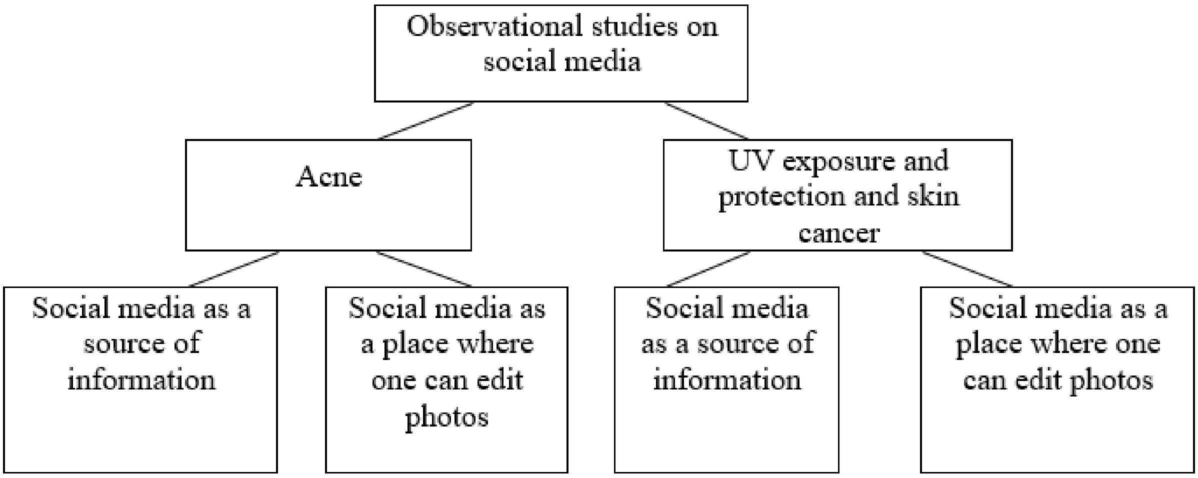
An effective intervention was also demonstrated in the study by Køster et al [35], who reported that, in effect, the use of tanning beds decreased in parallel with antisun campaign activities, with an odds ratio of 0.61 (95% CI 0.54-0.69). The campaign was mainly conducted on social media but also in youth magazines and on the radio. It was supplemented with a music video available on the internet and television on skin damage and the negative effects of using cosmetic beds. In addition, respondents to the postcampaign survey were in favor of legislative changes to restrict the access of children and teenagers aged <18 years to tanning salons [35].

Result Summary

To sum up, social media is a place where observational studies and interventional studies can be carried out. Unlike interventional studies, observational studies do not assess cause and effect relationships, require a large sample, and carry a high risk of error. In both observational and interventional studies, the most frequently discussed topics were skin exposure to and protection against UV radiation and skin cancer. Only 13% (3/23) of observational studies focused on acne [19,27,36]. No skin care studies were found that met the inclusion criteria. Observational studies conducted on social media showed that social media is a source of information [14,23,26,27,30,36] and a place where one can edit one's own photos [19,31]. Observational studies on acne showed that dissatisfaction with the appearance of the skin (presence of skin lesions) stimulates individuals to use social media. In the case of observational studies on exposure and protection against UV radiation, it was shown that the use of social media increases dissatisfaction with appearance and, consequently, increases exposure to UV radiation.

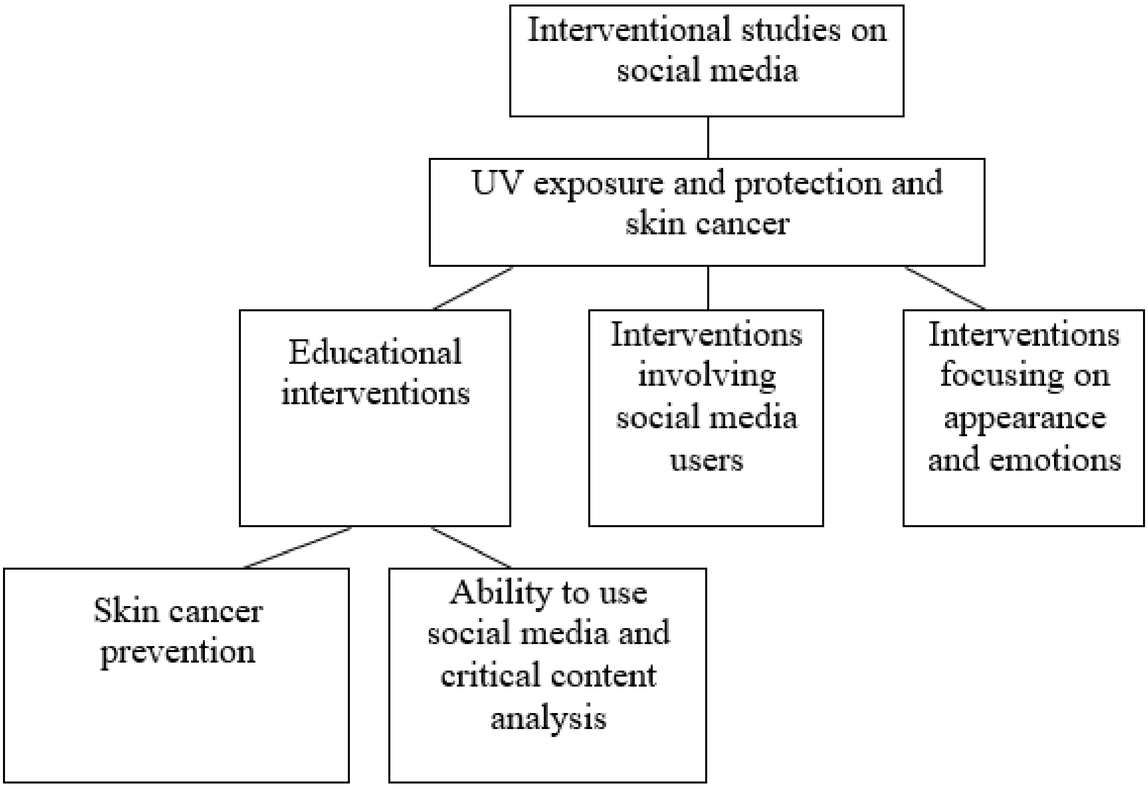
Visualization in social media has a strong impact on users. Showing images contributes to taking actions that are both beneficial (in the presence of skin lesions)—seeking dermatological care—and unfavorable (the desire to have a nice tan)—increased exposure to UV radiation (Figure 2).

Figure 2. Summary of observational studies on social media.



The analyzed social media intervention studies mainly concerned skin exposure to and protection against UV radiation, tanning, skin self-examination, and skin cancer. Most studies (11/23, 48%) used educational interventions on skin cancer prevention [15-18,20,25,28,29,33-35]. One study used an educational intervention on social media literacy and critical content analysis [32]. A total of 13% (3/23) of the studies examined social media users' involvement in generating social media content [15-17,22,24]. These studies pointed out that social media is a motivating factor. Motivation is the readiness to take a specific action, triggered by a need. To take action, one needs the right attitude and the internal conviction that the action is advisable. The aim of educational interventions was to change attitudes toward skin health promotion. However, interventions encouraging participants to create content on social media were intended to engage participants to increase the effectiveness of the intervention and the belief in its purposefulness. A total of 9% (2/23) of the studies focused primarily on body image rather than skin health issues [21,24]. The study by Myrick et al [21] discussed a psychological approach to appearance and emotion self-control (Figure 3). Both in observational and interventional studies, attention was paid to visualization on social media (images, photos, and videos). In addition, attention was paid to the involvement of social media users, which was measured via the amount of content shared, comments, likes, and photos published. Intervention studies showed that educational strategies based on humorous films, interesting personal stories, and opportunities for discussion were more engaging for participants. In observational studies, it was not possible to measure the level of persistence in following recommendations (eg, regular photoprotection). Interventional studies provided this possibility, and their results can be checked after a period.

Figure 3. Summary of interventional studies on social media.



Discussion

Principal Findings

The results of this review indicate that social media is a platform for disseminating health messages on an ongoing basis and for creating interventions targeted at specific groups. However, research on social media impact on education, attitudes, behaviors, and actions concerning skin health promotion was inconclusive and varied. Some studies (2/23, 9%) stressed the negative side of social media, noting that, for example, its acne advice was inconsistent with current guidelines [27,36]. Others (4/23, 17%) drew attention to the fact that social media can promote risky health behaviors concerning sunbathing without sunscreen, which may result in skin cancer [14,23,26,31]. On the other hand, without focusing on digital interventions, some of the studies included in this review (2/23, 9%) dealt with a positive impact of social media on skin care, skin protection, and skin health promotion. They noted that social media increases users' self-awareness and motivates them to take care of their skin [19,30]. According to Benetoli et al [41], social media has positively affected users' interaction with health care professionals. As a result, patients feel more confident and more assertive in their decision-making [41]. Most of the studies discussed in this review (14/23, 61%) argued that social media interventions resulted in positive effects on preventive skin health measures, changing users' behaviors, actions, attitudes, and knowledge levels [15-18,20-22,24,28,29,32-35]. Therefore, it is reasonable to ask how to make social media interventions effective and how to achieve a long-term change in user behaviors.

Agha-Mir-Salim et al [28] investigated the effects of social media digital interventions and found that Facebook was less effective than printed materials in improving awareness of skin cancer and the need for sun protection. However, without information about the level of audience engagement measured through the number of posts and reactions to them, it is difficult to explain the reasons why flyers were more effective than Facebook [28].

User engagement is critical to the success of social media interventions. The assessment of user engagement with social media interventions was presented in 35% (8/23) of the studies, determining the reactions to posts using the number of likes, comments under posts, and shares [17,18,20,22,24,29,32,33]. In addition, the concept of social media engagement plays a key role in health psychology [42]. Systematic reviews evaluating strategies that promote technology-based engagement have shown a potential positive impact of these strategies on health behaviors [43-47]. The success of social media engagement is based on its effect on society. Psychological factors and neural mechanisms are important [48]. Nahum-Shani et al [48] explain how positive engagement develops in response to a digital stimulus facilitating behavior change. By describing the neural basis of decision-making (by explaining its application to digital interventions), the researchers extend the concepts by Samanez-Larkin and Knutson [49] (affect-integration-motivation), creating the concepts of affect-integration-motivation and attention-context-translation of motivation to behavior. These additional elements can be used for targeted engagement of social media users with digital interventions [48].

Behavioral interventions through social media have many advantages, such as a large audience, low costs of disseminating information, and a possibility to create individual content and interact with other users with similar experiences or in a similar situation. This allows for mutual support, the exchange of information, and an opportunity to interact with health professionals by consulting specialists and obtaining answers [18].

In the observational studies on acne by Bahaj et al [36] and Yousaf et al [27], the authors pointed out that much of the advice available on social media is not consistent with the guidelines of the American Academy of Dermatology. In addition, in the observational study on skin cancer by Basch et al [14], it was shown that people who believed that social media was an accurate source of health information were more likely to fail to follow proper sun protection practices. The aforementioned study showed that social media can increase belief in health myths. The study by Vraga et al [25] showed that videos with erroneous information had a stronger impact on users than educational videos. Social media actions should also be aimed at combating misinformation. The content posted should be in agreement with evidence-based medicine and with the guidelines of the American Academy of Dermatology. In addition, interventions should increase the knowledge of social media users about the dangers of improper skin care and change their positive attitude toward inappropriate activities such as too much tanning. Social media interventions should also contribute to an increase in assertiveness toward inappropriate, harmful content, building self-esteem in users and teaching critical analysis of the content shared. In the case of tanning, interventions should describe alternative options, such as the use of safe moisturizing bronzers.

It seems that effective interventions should not only aim at the health consequences of poor skin care and poor skin protection but also focus on monitoring skin texture. In addition, they should make an effort to shape social norms regarding attractiveness (ie, by questioning the ideal of tanned skin and promoting sun protection). The use of dissonance induction approaches in interventions also seems to be important. The goal of those approaches is to encourage participants to persuade others to engage in appropriate behaviors [22].

Behavioral health interventions concerning appearance and body image often use persuasive techniques based on cognitive dissonance theories. Those techniques encourage participants to engage in discussions and cognitive exercises. Participants endorse views that contradict their previous beliefs and, in the end, accept these views as their own [24]. According to the dissonance theory, when there is a conflict between a belief (eg, a perfect tan is good) and behaviors (eg, persuading others that a perfect tan must be avoided and sunscreen should be used), psychological discomfort (cognitive dissonance) is the result, which motivates an individual to change their original beliefs (ie, the person is motivated to use sunscreen because tanned skin is less important than healthy skin) [22,50]. The effectiveness of cognitive dissonance interventions has been confirmed in studies on eating disorders [51] and perfect tans [52].

Pagoto et al [53] proposed a methodology for adaptation of social media behavioral interventions. It requires that parameters be defined to determine whether the intervention is based on social media completely or partially. The objective of a social network should be defined, too. First, it should be decided whether the main objective of the intervention is to engage users, disseminate information, or both. The type of intervention should be related to the objective. One-way communication is adopted mainly for public health campaigns. In 2-way communication, the content is generated by both the intervention moderators and the audience. Two-way communication allows for discussions and for help with problem-solving and behavioral advice. When designing interventions, it is important to choose the social networking platform and the target population. Another important element when designing an intervention is content conversion, with short, catchy posts, videos, photos, graphics, and links to articles. When the main focus of the intervention is user engagement, a plan should be constructed. It should include the size of the group, the frequency of posting, and the number of engaged participants, and it should specify whether the posts are automated. In the design of interventions, it is also reasonable to create chat groups or use microcounseling, which involves the intervention moderator initiating the discussion. The engagement plan may also include recruiting friends or family members for the intervention. In addition, it is important to train users on how to clarify expectations and encourage them to express their opinions and publish posts [53].

Almost all the intervention studies analyzed in this review (13/15, 87%) were conducted after Pagoto et al [53] developed their methodology for adapting social media interventions. After analyzing the presented interventions, it can be assumed that the authors followed some of the advice by Pagoto et al [53] when designing their interventions. Intervention studies that focus on engaging participants have found greater intervention effectiveness.

In summary, the analysis of the studies included in this review indicates that, when designing social media interventions for skin health, it is necessary to pretest them in a pilot project with comparison groups, including control groups, and measure user engagement on a regular basis. In addition, designing interventions requires focusing on the content that is designed to engage the audience. Absorbing content should not only be educational but also entertaining and discourage abnormal health behaviors. The content should be focused on the benefits according to positive psychology (to increase its attractiveness); it should focus on thinking about the future using goal setting and motivation (to have nice and healthier skin). It is also reasonable to include personal stories, exercises to engage the audience, and various puzzles. When designing interventions, the theories of social psychology they are based on should be considered.

The results of this review may be useful in creating future skin health promotion programs on social media. It is aimed at all those who are ready to broaden their knowledge and effectively engage in activities related to skin care and skin protection.

This study aimed to fill the research gap in assessing the usefulness of social media for skin health promotion. Its main assumption was that it would contribute to the discussion on the importance of the information provided about the skin and its impact on changing the attitudes and behaviors of recipients of this information.

Strengths and Limitations

As with any other study, this review had its limitations. The first one was that only articles written in English were analyzed, which meant that important publications written in other languages were not included. The second limitation was the omission of studies that were in progress (not completed). One of them was a Facebook intervention targeted at patients with melanoma and their families (randomized controlled trial) as the report on its results was expected to be completed by November 2023 [54]. Therefore, its analysis was not possible, and this review should be updated at a later time. The third limitation is the small number of interventional studies on acne and a lack of research on rosacea and atopic dermatitis, common conditions dealt with in dermatological and cosmetology clinics. In addition, there was a lack of research on daily skin care, education, and antiaging activities on social media. Some papers, such as those dealing with social media posts on esthetic medicine and more or less invasive plastic surgery, were not included in this review. This prevented a broader analysis and assessment of the impact of social media on users' skin care and health activities. The last limitation was the failure to undertake a quantitative data analysis because of too broad a topic, which resulted in the heterogeneity of the studies that qualified for analysis.

Despite its limitations, this paper presents an up-to-date review of social media impact on users' behavior change and the implementation of interventions mainly in the field of skin cancer prevention. In addition, this review identifies research gaps and topics worth focusing on and indicates what should be improved when conducting social media research.

Conclusions and Practical Implications

Social media remote consultations and digital interventions with the use of new technologies in education and research are a relatively new method of skin health promotion. Despite the limited number of studies and their heterogeneity, this review outlined the impact of social media on user behaviors, attitudes, and actions in terms of skin health. It provided insights into social media importance and identified strategies for digital interventions to promote health. In addition, this review underlines the need to conduct research on skin health promotion via social media and on new methods motivating users to take an active role. There is no doubt that more high-quality studies, consistent and with logical conclusions, on social media using standardized interventions are needed.

The observational studies analyzed in this review (8/23, 35%) showed that social media, according to social media users, is a reliable place to obtain information on skin health. Moreover, there was a noticeable tendency of social media influencing the behavior of social media users. On the basis of unpublished own research (by author JMB), of 150 surveyed social media users (people aged ≤25 years), in 2021, a total of 62% (93 people) believed that information from social media regarding skin cancer prevention, skin care and treatment of acne, and antiaging was very important to them. For 44% of respondents (66 people), information on social media influenced their behavior. In a repeated study in 2023, as many as 86% of respondents (129 people) stated that information on skin cancer prevention, skin care and treatment of acne, and antiaging was very important to them. However, 74% of respondents (111 people) stated that information on social media influenced their behavior.

The studies analyzed in this review showed that more frequent use of social media is significantly associated with more frequent exposure to UV radiation. Social media focuses on visualization. Photos and images are a powerful tool to influence users. In intervention studies, there was a noticeable tendency to encourage people to react to posts and create their own. It can be assumed that interventions involving social media users in generating content increase their motivation to take care of their skin. The aim of social media interventions should be education on promoting skin health based on visualization and motivating social media users to take care of their skin (using photoprotection, limiting sunbathing, using protective clothing, conducting skin self-examinations, and having regular skin checkups with a dermatologist).

The availability of social media worldwide provides unlimited possibilities for transmitting and managing information. Analyzing posts and evaluating user engagement by measuring the number of likes, comments, and followers can be a powerful tool for gathering knowledge and studying patients' opinions. After becoming familiar with the expectations of social media users and assessing information resources and their quality and credibility, appropriately targeted actions should be taken to educate, encourage, and motivate users to be active in skin care and skin health promotion in everyday life.

Social media has become a tool supporting the life and work of society. The model of Society 1.0 was based on hunting, Society 2.0 was based on agriculture, Society 3.0 was based on industry, and Society 4.0 is based on information. Social media implements the assumptions of the Society 4.0 model through information activities. The vision of Society 5.0 is a balance between technological progress and human needs. Model 5.0 enables the collection of data (eg, on health status) from patients on an ongoing basis. Social media can be a tool for digital interventions and for Society 5.0 implementation. The implementation of Digital Society 5.0 depends not only on the cooperation between countries but also on combined efforts of various branches of medicine, business, and science, which is the basis for creating innovations at both the national and global level. Innovations are not possible without cooperation; the use of global data (big data); and the coordination of many data systems and knowledge platforms, including social media.

Proper and specialized use of social media by dermatologists, pharmacists, cosmetologists, and health educators can increase awareness concerning skin condition and skin health preventive measures. It can encourage users to conduct a skin self-examination and visit a dermatologist. This review could also enable dermatologists, cosmetologists, pharmacists, and health educators to understand social media potential and its impact on users. This review presents social media use for intervention studies and indicates its educational and motivational role in actions related to skin health. In addition, those actions have a clinical significance as social media increases user awareness of skin health promotion. As a result, users are not reluctant to seek medical help. Therefore, this review could be a basis for further research on the role of social media interventions in skin health promotion (including sun protection; mole control; and proper skin care for various dermatological diseases, such as acne, rosacea, or atopic dermatitis). It seems that social media, compared to traditional forms of mass communication, has the potential to make a change in skin care and health behaviors and activities among a larger population.

First, future studies should focus on effective social media strategies to promote skin health. Second, research assessing the effects of health and skin care recommendations would be useful. Third, studies assessing the short-term engagement of social media users would be of practical importance, as well as research assessing their commitment to long-term health and skin care habits. The results of this review, in addition to its theoretical value, could have potential applications in online dermatological and cosmetology consultations and in prohealth campaigns concerning increasingly common skin diseases such as acne and melanoma.

This research dealt with studies investigating individually developed social media strategies. It seems that it would still be appropriate to standardize tools assessing the quality of digital health interventions on social media. Such standardization would be helpful both in the design of new digital interventions and in the evaluation of existing ones.

Acknowledgments

The authors would like to thank the Main Library of the University of Siedlce in Poland for access to scientific databases.

Authors' Contributions

JMB contributed to work concept and design, data collection and analysis, visualization, writing the original draft, reviewing and editing the manuscript, critical review, and final approval. JG contributed to data collection and analysis and critical review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the included studies.
[DOCX File , 67 KB - jmir_v27i1e54241_app1.docx]

Multimedia Appendix 2
Findings of the included studies.
[DOCX File , 67 KB - jmir_v27i1e54241_app2.docx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) checklist.
[DOCX File , 31 KB - jmir_v27i1e54241_app3.docx]

References

1. Number of social media users worldwide from 2017 to 2028. Statista. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2023-05-31]
2. Number of internet and social media users worldwide as of October 2024. Statista. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2023-05-31]
3. Savas JA, Huang KE, Tuchayi SM, Feldman SR. Understanding the influence of social media in medicine: lesson learned from Facebook. Dermatol Online J 2014 Sep 16;20(9):16. [Medline: 25244178]
4. Gantenbein L, Navarini AA, Maul LV, Brandt O, Mueller SM. Internet and social media use in dermatology patients: search behavior and impact on patient-physician relationship. Dermatol Ther 2020 Nov;33(6):e14098. [doi: 10.1111/dth.14098] [Medline: 32725746]
5. Banerjee SC, Rodríguez VM, Greene K, Hay JL. Trending on Pinterest: an examination of pins about skin tanning. Transl Behav Med 2019 Jul 16;9(4):737-748 [FREE Full text] [doi: 10.1093/tbm/iby036] [Medline: 29648613]
6. Rew J, Hwang E, Choi YH, Rho S. Monitoring skin condition using life activities on the SNS user documents. Multimed Tools Appl 2017 Apr 22;77(8):9827-9847. [doi: 10.1007/s11042-017-4623-z]
7. Joshi M, Korrapati NH, Reji F, Hasan A, Kurudamannil RA. The impact of social media on skin care: a narrative review. L viv Clin Bull 2022 Jun 17;1-2(37-38):85-96. [doi: 10.25040/lkv2022.01-02.085]

8. De La Garza H, Maymone MB, Vashi NA. Impact of social media on skin cancer prevention. Int J Environ Res Public Health 2021 May 09;18(9):5002 [FREE Full text] [doi: 10.3390/ijerph18095002] [Medline: 34065061]
9. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
10. Niu Z, Bhurosy T, Heckman CJ. Digital interventions for promoting sun protection and skin self-examination behaviors: a systematic review. Prev Med Rep 2022 Apr;26:101709 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101709] [Medline: 35529530]
11. 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]
12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
13. 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]
14. Basch CH, Hillyer GC, Romero RA, MacLean SA, Ethan D. College students' attitudes and behaviors related to sun safety and appearance in relation to health information-seeking behavior and social media use: cross-sectional study. JMIR Dermatol 2018 Dec 19;1(2):e10984. [doi: 10.2196/10984]
15. Buller DB, Pagoto S, Baker K, Walkosz BJ, Hillhouse J, Henry KL, et al. Results of a social media campaign to prevent indoor tanning by teens: a randomized controlled trial. Prev Med Rep 2021 Jun;22:101382 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101382] [Medline: 33996394]
16. Buller DB, Pagoto S, Henry KL, Baker K, Walkosz BJ, Hillhouse J, et al. Persisting effects of a social media campaign to prevent indoor tanning: a randomized trial. Cancer Epidemiol Biomarkers Prev 2022 Apr 01;31(4):885-892 [FREE Full text] [doi: 10.1158/1055-9965.EPI-21-0059] [Medline: 35064063]
17. Buller DB, Pagoto S, Henry KL, Baker K, Walkosz BJ, Hillhouse J, et al. Effects of engagement with a social media campaign for mothers to prevent indoor tanning by teens in a randomized trial. J Health Commun 2022 Jun 03;27(6):394-406 [FREE Full text] [doi: 10.1080/10810730.2022.2113839] [Medline: 35993376]
18. Coups EJ, Manne SL, Pagoto SL, Criswell KR, Goydos JS. Facebook intervention for young-onset melanoma patients and their family members: pilot and feasibility study. JMIR Dermatol 2018 Nov 01;1(2):e3. [doi: 10.2196/derma.9734]
19. Martel J, Powell E, Murina A. The effect of Instagram and photograph editing on seeking dermatologic care. J Cosmet Dermatol 2020 Oct 16;19(10):2732-2735. [doi: 10.1111/jocd.13456] [Medline: 32333461]
20. Morrison L, Chen C, Torres JS, Wehner M, Junn A, Linos E. Facebook advertising for cancer prevention: a pilot study. Br J Dermatol 2019 Oct 04;181(4):858-859 [FREE Full text] [doi: 10.1111/bjd.17993] [Medline: 30972743]
21. Myrick JG, Waldron KA, Cohen O, DiRusso C, Shao R, Cho E, et al. The effects of embedded skin cancer interventions on sun-safety attitudes and attention paid to tan women on Instagram. Front Psychol 2022 Apr 8;13:838297 [FREE Full text] [doi: 10.3389/fpsyg.2022.838297] [Medline: 35465513]
22. Pagoto SL, Waring ME, Groshon LC, Rosen AO, Schroeder MW, Goetz JM. Proof-of-concept feasibility trial of a dissonance-based sun safety intervention for young adult tanners. Ann Behav Med 2022 Aug 02;56(8):830-841 [FREE Full text] [doi: 10.1093/abm/kaab116] [Medline: 35179176]
23. Stapleton JL, Hillhouse J, Coups EJ, Pagoto S. Social media use and indoor tanning among a national sample of young adult nonHispanic White women: a cross-sectional study. J Am Acad Dermatol 2016 Jul;75(1):218-220 [FREE Full text] [doi: 10.1016/j.jaad.2016.01.043] [Medline: 27317521]
24. Stapleton JL, Manne SL, Day AK, Levonyan-Radloff K, Pagoto SL. Healthy body image intervention delivered to young women via Facebook groups: formative study of engagement and acceptability. JMIR Res Protoc 2018 Feb 20;7(2):e54 [FREE Full text] [doi: 10.2196/resprot.9429] [Medline: 29463495]
25. Vraga EK, Bode L, Tully M. The effects of a news literacy video and real-time corrections to video misinformation related to sunscreen and skin cancer. Health Commun 2022 Nov 12;37(13):1622-1630. [doi: 10.1080/10410236.2021.1910165] [Medline: 33840310]
26. Willoughby JF, Myrick JG. Entertainment, social media use and young women's tanning behaviours. Health Educ J 2018 Dec 23;78(3):352-365. [doi: 10.1177/0017896918819643]
27. Yousaf A, Hagen R, Delaney E, Davis S, Zinn Z. The influence of social media on acne treatment: a cross-sectional survey. Pediatr Dermatol 2020 Mar;37(2):301-304 [FREE Full text] [doi: 10.1111/pde.14091] [Medline: 31944359]
28. Agha-Mir-Salim L, Bhattacharyya A, Hart D, Lewandowska M, Spyropoulou E, Stinson L, et al. A randomised controlled trial evaluating the effectiveness of Facebook compared to leaflets in raising awareness of melanoma and harmful sun-related behaviour among young adults. Eur J Cancer Prev 2020 Jan;29(1):89-91. [doi: 10.1097/CEJ.0000000000000519] [Medline: 30998526]

29. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for behavior change: using social media for the dissemination of public health messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]
30. Guckian J, Jobling K, Oliphant T, Weatherhead S, Blasdale K. 'I saw it on Facebook!' Assessing the influence of social media on patient presentation to a melanoma screening clinic. Clin Exp Dermatol 2020 Apr 02;45(3):295-301. [doi: 10.1111/ced.14100] [Medline: 31541480]
31. Mingoia J, Hutchinson AD, Gleaves DH, Corsini N, Wilson C. Use of social networking sites and associations with skin tone dissatisfaction, sun exposure, and sun protection in a sample of Australian adolescents. Psychol Health 2017 Dec;32(12):1502-1517. [doi: 10.1080/08870446.2017.1347788] [Medline: 28691513]
32. Mingoia J, Hutchinson AD, Gleaves DH, Wilson C. The impact of a social media literacy intervention on positive attitudes to tanning: a pilot study. Comput Human Behav 2019 Jan;90:188-195. [doi: 10.1016/j.chb.2018.09.004]
33. Potente S, McIver J, Anderson C, Coppa K. “It's a beautiful day for cancer”: an innovative communication strategy to engage youth in skin cancer prevention. Soc Mark Q 2011 Sep 01;17(3):86-105. [doi: 10.1080/15245004.2011.595604]
34. Damude S, Hoekstra-Weebers JE, van Leeuwen BL, Hoekstra HJ. Melanoma patients' disease-specific knowledge, information preference, and appreciation of educational YouTube videos for self-inspection. Eur J Surg Oncol 2017 Aug;43(8):1528-1535. [doi: 10.1016/j.ejso.2017.06.008] [Medline: 28684059]
35. Køster B, Thorgaard C, Philip A, Clemmensen I. Sunbed use and campaign initiatives in the Danish population, 2007-2009: a cross-sectional study. J Eur Acad Dermatol Venereol 2011 Nov 09;25(11):1351-1355. [doi: 10.1111/j.1468-3083.2010.03960.x] [Medline: 21711466]
36. Bahaj RK, Alsaggaf ZH, Abduljabbar MH, Hariri JO. The influence of social media on the treatment of acne in Saudi Arabia. Cureus 2022 Mar;14(3):e23169 [FREE Full text] [doi: 10.7759/cureus.23169] [Medline: 35444887]
37. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
38. Resources. Center for Evidence Based Management. URL: https://cebma.org/resources/ [accessed 2024-09-10]
39. Popay J, Roberts H, AJ S, Petticrew M, Britten N, Arai L, et al. Guidance on the conduct of narrative synthesis in systematic reviews final report. J Epidemiol Community Health 2005;59(Suppl 1):A7 [FREE Full text]
40. Zaenglein AL, Pathy AL, Schlosser BJ, Alikhan A, Baldwin HE, Berson DS, et al. Guidelines of care for the management of acne vulgaris. J Am Acad Dermatol 2016 May;74(5):945-73.e33 [FREE Full text] [doi: 10.1016/j.jaad.2015.12.037] [Medline: 26897386]
41. Benetoli A, Chen TF, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. Patient Educ Couns 2018 Mar;101(3):439-444. [doi: 10.1016/j.pec.2017.08.015] [Medline: 28882545]
42. Nahum-Shani I, Rabbi M, Yap J, Philyaw-Kotov ML, Klasnja P, Bonar EE, et al. Translating strategies for promoting engagement in mobile health: a proof-of-concept microrandomized trial. Health Psychol 2021 Dec;40(12):974-987 [FREE Full text] [doi: 10.1037/hea0001101] [Medline: 34735165]
43. 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]
44. De Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: systematic review. J Med Internet Res 2014 Mar 25;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]
45. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
46. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
47. 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]
48. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]
49. Samanez-Larkin GR, Knutson B. Decision making in the ageing brain: changes in affective and motivational circuits. Nat Rev Neurosci 2015 May 15;16(5):278-289 [FREE Full text] [doi: 10.1038/nrn3917] [Medline: 25873038]
50. Festinger L. A Theory of Cognitive Dissonance. Redwood City, CA: Stanford University Press; 1957.
51. Stice E, Shaw H, Becker CB, Rohde P. Dissonance-based Interventions for the prevention of eating disorders: using persuasion principles to promote health. Prev Sci 2008 Jun;9(2):114-128 [FREE Full text] [doi: 10.1007/s11121-008-0093-x] [Medline: 18506621]
52. Chait SR, Thompson JK, Jacobsen PB. Preliminary development and evaluation of an appearance-based dissonance induction intervention for reducing UV exposure. Body Image 2015 Jan;12:68-72. [doi: 10.1016/j.bodyim.2014.09.004] [Medline: 25462883]

53. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting behavioral interventions for social media delivery. J Med Internet Res 2016 Jan 29;18(1):e24 [FREE Full text] [doi: 10.2196/jmir.5086] [Medline: 26825969]
54. Manne S, Pagoto S, Peterson S, Heckman C, Kashy D, Berger A, et al. Facebook intervention for young-onset melanoma survivors and families: protocol for a randomized controlled trial. JMIR Res Protoc 2023 Jan 24;12:e39640 [FREE Full text] [doi: 10.2196/39640] [Medline: 36692933]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

Edited by A Mavragani; submitted 02.11.23; peer-reviewed by J Makin, D Buller, A Kinsey; comments to author 20.03.24; revised version received 06.05.24; accepted 09.11.24; published 06.01.25.

Please cite as:
Brzozowska JM, Gotlib J
Social Media Potential and Impact on Changing Behaviors and Actions in Skin Health Promotion: Systematic Review
J Med Internet Res 2025;27:e54241
URL: https://www.jmir.org/2025/1/e54241
doi:10.2196/54241
PMID:

Original Paper

Unpacking Performance Factors of Innovation Systems and Studying Germany's Attempt to Foster the Role of the Patient Through a Market Access Pathway for Digital Health Applications (DiGAs): Exploratory Mixed Methods Study

Sara Gehder[1], MSc; Moritz Goeldner[1], Prof Dr

[1] Working Group for Data-Driven Innovation, Hamburg University of Technology, Hamburg, Germany

Corresponding Author:
Moritz Goeldner, Prof Dr
Working Group for Data-Driven Innovation
Hamburg University of Technology
Am Schwarzenberg-Campus 4
Hamburg, 21073
Germany
Phone: 49 40 428784777
Email: moritz.goeldner@tuhh.de

Abstract

Background: Health care innovation faces significant challenges, including system inertia and diverse stakeholders, making regulated market access pathways essential for facilitating the adoption of new technologies. The German Digital Healthcare Act, introduced in 2019, offers a model by enabling digital health applications (DiGAs) to be reimbursed by statutory health insurance, improving market access and patient empowerment. However, the factors influencing the success of these pathways in driving innovation remain unclear.

Objective: This study aims to identify the key performance factors of the innovation system shaped by the *patient-relevant structural and procedural improvement* (pSVV) pathway within the DiGA model. By examining how this pathway supports the entry of innovative digital health technologies, we seek to uncover the systemic dynamics that influence its effectiveness in fostering patient-centered digital health solutions.

Methods: This study, conducted from May 2023 to November 2024, used a mixed methods approach. A descriptive analysis assessed how DiGA manufacturers use positive health care effects, giving a market overview of the pSVV technology. A qualitative analysis using grounded theory and Gioia methodology provided insights into stakeholder perspectives, focusing on manufacturers and regulatory bodies. A functional-structural analysis examined how components of the innovation system, such as actors, institutions, interactions, and infrastructure, interact and impact the effectiveness of the pathway.

Results: The descriptive analysis showed that only 11 (20%) of the 56 DiGAs available in Germany used the pSVV pathway, with only 1 (2%) provisionally listed DiGA using pSVV as a primary end point; 6 of 9 (67%) pSVV key areas were used. The qualitative analysis revealed that manufacturers prioritize demonstrating medical benefits over pSVV due to evidence requirements and uncertainties around pSVV acceptance. Operational barriers hindered the adoption of pSVV, despite a positive reception among stakeholders. The systemic analysis identified key issues, including a lack of entrepreneurial focus on pSVV, limited regulatory experience, inadequate measurement methods, and entrenched practices prioritizing medical benefits, that hinder market formation and legitimacy.

Conclusions: This study identifies key factors for effectively implementing innovation systems through regulated market access pathways, including content and format security, clearer framework specification, active innovation process management, and market formation stimulation. Addressing these factors can reduce uncertainties and promote wider adoption of digital health technologies. The findings highlight the need for future research on patient empowerment and the development of methodologies beyond traditional therapeutic outcomes.

(J Med Internet Res 2025;27:e66356) doi:10.2196/66356

KEYWORDS

regulatory market access pathways; digital health application; DiGA; patient-relevant structural and procedural improvement; pSVV pathway analysis; qualitative and systemic analysis; policy and stakeholder insights; innovation system analysis

Introduction

Background

Innovation in health care is well known for being challenging. The complexity is driven by factors such as the involvement of diverse stakeholders in highly regulated markets [1,2], the need to address "wicked problems," complex issues without straightforward solutions [3], and substantial "system inertia," meaning resistance to change, which is especially pronounced in health care systems [4]. Policy approaches have been shown to act as inertial forces that can significantly shape the development of innovative technologies in health care [5], particularly digital health technologies, where patient access is greatly limited without broad and scalable market access pathways [6]. Mobile health apps are becoming increasingly available to consumers worldwide [7-9]; however, significant barriers, including regulatory complexity, insufficient evidence of clinical benefit, and data privacy concerns, still prevent their full integration into health care systems, leaving them underused in clinical practice [10-14].

To address these challenges, health policy approaches are being adopted globally to standardize regulatory processes and speed up market access, particularly in the rapidly evolving field of digital health [15-18]. Notable examples of health policy approaches include regulated market access pathways, such as *the fast-track process for digital health applications* pathway in Germany [19] and the *Prise en Charge Anticipée Numérique* (PECAN) pathway in France [20], which aim to facilitate the entry of digital health applications (DiGAs; German: Digitale Gesundheitsanwendung) into their respective markets. These new approaches are gaining significant international attention as pioneering models in the field [6,21].

To illustrate, the German Digital Healthcare Act (DVG), introduced in 2019, established the DiGA pathway to foster the integration of so-called DiGAs into the health care system. This pathway allows DiGAs to be reimbursed by statutory health insurance when prescribed by health care professionals or provided by health insurers. With approximately 90% of the German population covered by statutory health insurance [22], these "apps on prescription" are widely accessible at no cost to patients. Eligible DiGAs are included in the national DiGA Directory [23], which is managed by the Federal Institute for Drugs and Medical Devices (BfArM; German: Bundesinstitut für Arzneimittel und Medizinprodukte) to ensure transparency and oversee the approval process [24,25]. To qualify, a DiGA must be Conformité Européenne (CE)–marked (classes I and IIa or IIb), be based primarily on digital technology intended for direct patient use, and support the detection, monitoring, treatment, or alleviation of diseases, injuries, or disabilities [26]. Following a recent regulatory update in April 2024, digital medical devices classified as class IIb have been added to be eligible for recognition as a DiGA. However, this inclusion comes with specific stipulations; these devices are required to demonstrate medical benefit only, and they are eligible solely for permanent listing, excluding the option for preliminary listing. In addition, a DiGA must meet stringent requirements for safety, quality, functionality, privacy, and data security as defined in the respective DiGA ordinance (DiGAV; German: Digitiale Gesundheitsanwendungen-Verordnung) [24]. To enhance transparency around the assessment criteria and specific requirements established by BfArM, a DiGA guide [27] for manufacturers, service providers, and users was published. This guide provides a concise overview of the applicable ordinances and regulations, detailing how BfArM interprets the normative requirements set forth in the DVG and DiGAV.

Despite some emerging research examining the market dynamics following the implementation of regulated market access pathways, such as the number of products entering the market, time to market, pricing, and study designs used to demonstrate efficacy [28,29], as well as factors contributing to the success of specific DiGA products [30], broader factors influencing successful technology integration remain underexplored. This study addresses these gaps to better inform future policy and innovation strategies. To identify potential success factors, it is crucial to consider not only the regulated market access pathway itself but also the surrounding elements that influence its effectiveness. The broader network of actors, institutions, and resources involved in the creation, diffusion, and use of new technologies is defined as the technology innovation system [31]. A health policy approach, such as the creation of a regulated market access pathway that "brings about processes of change and stimulates technological innovation," can be characterized as a systemic instrument [32]. Systemic instruments are mechanisms designed to shape the conditions under which technological innovation systems operate. Understanding the dynamics that a systemic instrument brings to a technological innovation system requires the consideration of various functional and structural components, all of which must be in place to effectively implement the targeted technological innovation [33]. These components, as outlined by Wieczorek and Hekkert [34], include the structural components, such as actors, institutions, interactions, and infrastructure, as well as the functional components, such as entrepreneurial activities (F1), knowledge development (F2), knowledge diffusion (F3), guidance of the search (F4), market formation (F5), mobilization of resources (F6), and creation of legitimacy (F7) [35].

This exploratory study aimed to identify factors contributing to the success of an innovation system shaped by a regulated market access pathway in the digital health sector. Through an exemplary case study, we sought insights to support ongoing improvements in digital health policy, focusing on optimizing market access pathways for digital health technologies. The central research question is as follows: What are the performance factors of an innovation system created by a regulated market access pathway that supports the market entry of innovative digital health technologies?

To address this question, we selected a specific case from the 2019 German policy innovation DVG, particularly the “patient-relevant structural and procedural improvements” (pSVV; German: Patientenrelevante Struktur- und Verfahrensverbesserung) pathway. The pSVV pathway is a targeted segment within the DiGA pathway, specifically modifying the element of product value assessment, 1 of the 4 “full-stack” elements (regulatory authorization, product value assessment, pricing and reimbursement, and patient access) that define a comprehensive market access pathway [6], promoting innovations aimed to specifically elevate DiGAs that enhance patient–health care provider interactions and support patient health behaviors.

Unlike traditional pharmaceuticals, which require evidence of “therapeutic efficacy” through clinical studies for coverage eligibility [36], DiGAs are allowed to demonstrate their effectiveness through “positive health care effects” defined as either “medical benefit” or pSVV. While the medical benefit criterion focuses on outcomes, such as improved health status, reduced disease duration, prolonged survival, or improved quality of life, the pSVV criterion introduces 9 additional areas to demonstrate the value that is equally eligible as defined in the DiGA guide. These areas include improving care coordination, adherence to guidelines and standards, enhancing patient adherence, facilitating access to care, increasing patient safety, and promoting health literacy and autonomy. This expansion of the traditional product value assessment beyond the traditional therapeutic benefits is further complemented in the DiGA guide by explicitly broadening the acceptable methods of evidence beyond the traditional randomized controlled trials. We selected this case due to the maturity of the evolving DiGA ecosystem [28] in Germany, which offers greater transparency and availability of data on use and uptake compared to other countries. Within this context, the pSVV pathway was chosen as a focal point because it represents a well-defined component within the DiGA pathway, involving a manageable number of stakeholders while offering sufficient accumulated experience to provide valuable insights into the factors influencing digital health innovation.

In addition, the introduction of the pSVV pathway is groundbreaking as it extends beyond traditional treatment parameters, embracing a holistic approach to health care that prioritizes patient empowerment and streamlined care processes [19,37,38]. This aligns with the global trend toward patient empowerment, emphasizing the importance of patient agency and involvement in health care processes [16,39,40].

Our research used a mixed methods approach to provide a comprehensive exploration of the factors at play. We began with a descriptive analysis of the innovation success of the targeted pSVV technology, which helps contextualize the case and assess the effectiveness of the policy intervention [41]. Following this, a qualitative analysis of the perception and use of the market access pathway was conducted to develop theoretical insights into the factors that influence the adoption of specific innovation pathways [42,43]. In addition, a functional-structural analysis of the technological innovation system was performed to uncover broader factors beyond the pathway itself that are essential for successful technological innovation [34].

Contribution

This study contributes to the field by providing a comprehensive analysis of factors that support the successful integration of digital health innovations. Our findings will provide policy makers and stakeholders worldwide with practical insights to improve regulated market access pathways and drive innovation strategies that better meet the needs of health care systems and patient empowerment.

Methods

Setting and Study Design

This exploratory study used a mixed methods approach to investigate factors influencing the success of a regulated market access pathway for digital health technologies. We conducted a descriptive analysis to assess the adoption of pSVV technology, followed by a qualitative analysis to explore stakeholder perspectives on the pSVV pathway, using grounded theory and Gioia methodology to ensure neutrality and objectivity. Finally, we performed a functional-structural analysis to identify additional success factors within the broader technology innovation system. Each method is described in detail subsequently.

This study was carried out between May 2023 and November 2024 with interviews being conducted between May 2023 and November 2023.

First, a descriptive analysis of the use of positive health care effects by DiGA manufacturers was conducted to establish an understanding of the current developmental status of the pSVV technology in the market.

Second, a qualitative analysis was performed to explore the success factors of a regulated market access pathway by subsequently analyzing the perception and use of the pSVV pathway of key stakeholders. Since the pSVV pathway was only introduced in 2019 and has not been studied yet, a qualitative research approach was adopted. This approach allows for deep exploration of complex and less understood topics [44]. Unlike quantitative methods, which rely on larger sample sizes and predefined variables, qualitative methods provide flexibility for exploring nuanced, context-specific insights, particularly valuable given the relatively small number of relevant stakeholders (35 DiGA manufacturers as of early 2023). By using the inductive grounded theory method [45] alongside the Gioia methodology [42,43], we objectively analyzed the views of market participants on the pSVV pathway, avoiding any imposition of preconceived theories and ensuring neutrality in the analysis. For the qualitative analysis, we followed the Consolidated Criteria for Reporting Qualitative Research-32 (COREQ-32) checklist [46].

Building on the dataset from the qualitative analysis, we conducted a functional-structural analysis of the technology innovation system established by the pSVV pathway to identify additional success factors related to regulated market access pathways. This analysis followed the systemic policy framework established by Wieczorek and Hekkert [34], enabling a comprehensive examination of the innovation system's functional dynamics and structural components within the current policy context.

While this study provided valuable insights into potential strategies to improve the pSVV pathway in particular, the international focus of this study precluded detailed insights into Germany-specific factors. To address this, we conducted a subsequent analysis focusing specifically on the question, "What factors facilitate or hinder the implementation of pSVV within the approval process for DiGA?" and "What specific requirements and expectations of relevant stakeholders must be addressed to establish the concept of pSVV in the market?" The results of this specific analysis were published in a German-language journal to inform stakeholders in this context [47].

Sampling and Data Collection

For the execution of our descriptive analysis regarding the state of adoption of pSVV, we evaluated publicly available data on the positive health care effects of DiGA from the DiGA directory [23]. The data were retrieved via an application programming interface provided by BfArM for scientific use.

In our qualitative study, we used a hybrid approach of criterion sampling and snowball sampling for participant recruitment. We aimed to engage with stakeholders who have a direct influence on which digital health products enter the market and, if entry occurs through the DiGA pathway, determine the associated positive health care effects. The inclusion criteria are defined subsequently.

The first inclusion criterion was organizations that are directly involved with the DiGA ordinances and regulations, whether developing it, ensuring or supporting its implementation, or actively using it.

The second criterion was individuals, within these organizations, whose professional roles directly involve the development, implementation, or use of the DiGAVs and regulations.

The third inclusion criterion was additionally individuals who have actively decided against using the DiGA pathway, provided they have extensively engaged with the ordinances and regulations and are potentially qualified to use it (eg, manufacturers of DiGAs not listed as DiGA).

Individuals without direct and active influence over the selection of products entering the DiGA market or the choice of the positive health care effect were excluded.

Consequently, the exclusion criteria led to the exclusion of stakeholder groups in the DiGA market, such as DiGA users, prescribers, and researchers in the field, while the inclusion criteria led to the identification of the stakeholder groups presented in Textbox 1.

Textbox 1. Stakeholder groups included in qualitative analysis.

- Digital health application (DiGA) manufacturers with patient-relevant structural and procedural improvement (pSVV; DiGA with pSVV): This category includes decision makers from manufacturers of DiGAs that are listed in the DiGA directory and have implemented pSVV as a positive health care effect.
- DiGA manufacturers without pSVV (DiGA without pSVV): This category comprises manufacturers who are listed in the DiGA directory but have not used pSVV as a positive health care effect.
- Manufacturers of mobile health applications not listed in the DiGA directory (non-DiGA): This group consists of manufacturers whose CE–certified mobile health apps are not, or are not yet, listed in the DiGA directory.
- Consultants: This group refers to professionals from advisory firms who specialize in assisting manufacturers in the process of getting listed in the DiGA directory.
- Regulatory bodies: This group encompasses public institutions that are deeply involved in the development and management of the DiGA pathway and the evolution of the guiding ordinances and regulations (Digital Healthcare Act [DVG], DiGA ordinance [DiGAV], and DiGA guide).

Due to the very small size of stakeholder groups 1 and 5, yet their significant relevance to our research questions, we contacted all organizations in these groups in accordance with our inclusion criteria. For groups 2 to 4, given the novelty and niche nature of the topic, we opted for snowball sampling [48] to achieve the highest possible participation rate. Individuals were excluded if someone from the same organization had already participated in the study, to avoid multiple responses from the same context. This combined approach resulted in a 66% positive response rate to our participation invitations. Participant recruitment began with initial email contacts, followed by referrals from existing participants in line with snowball sampling. Efforts were focused on ensuring a diverse representation of the identified stakeholder groups.

Our outreach targeted 29 individuals, with 19 (66%) agreeing to participate. Reasons for nonparticipation included lack of response (n=9, 31%) and time constraints (n=1, 4%). The distribution of speakers across groups and their organizational roles are detailed in Table S1 in Multimedia Appendix 1.

For the interviews, a semistructured approach was used with decision makers who had consented to participate and endorsed a data protection declaration. Efforts were made to ensure neutrality, with no prior relationships between interviewers and participants. An interview guide with 7 open-ended questions was prepared to facilitate conversation while ensuring neutrality and objectivity in the analysis. The themes of the questions were developed after preliminary data gathering, which included limited literature research and initial conversations with 2 experts in the field (a managing director from a manufacturer's association and a member of an organization that has 1 DiGA listed in the DiGA directory), adhering to the standards of the Gioia methodology [42]. These experts were not part of the subsequent interviews, ensuring that the development of the interview guide was informed yet independent of the interview process itself (Multimedia Appendix 1). In total, 2 central themes emerged from this process: the general approval process for DiGA and the specifics of the pSVV pathway. For each of these themes, we formulated specific questions, which were refined during the initial interviews to ensure clarity and neutrality. Interviews were scheduled at the convenience of the participants and conducted via Zoom (Zoom Video Communications, Inc), with only the participant and the primary interviewer (SG) present, though 3 selected interviews had a second researcher (MG) present for quality assurance. No interviews were repeated.

Conducted in German, interviews lasted 25 to 35 minutes, audio tracks were recorded, transcribed using the software Whisper Transcription (Good Snooze), and anonymized. Data saturation was reached when no new first-order concepts or second-order themes emerged during analysis, as reflected in field notes and initial coding. This indicated that the data structure, developed through the Gioia methodology, was comprehensive, and additional data collection would not yield further insights. Following the analysis, data triangulation was enhanced through additional discussions with 4 experts from regulatory and digital health sectors to refine the results. This included a conversation with leading representatives from BfArM, which, due to confidentiality agreements, cannot be directly cited in this study. Insights from these discussions were used to cross-verify emerging themes and validate our interpretations, ensuring robustness in our conclusions.

Ethical Considerations

Informed consent was obtained from all participants who were interviewed strictly as representatives of their respective employers. Participants were selected based solely on their relevance to the research question and their experience in the field, with no consideration of personal or socially sensitive attributes. No personal data or socially relevant information were collected, and all data were fully anonymized to ensure privacy and confidentiality. No compensation was provided for participation in the study.

Furthermore, the study did not involve vulnerable populations, health care professionals, medical procedures, health-related data, or any form of deception, manipulation, or intervention that could impact participants' well-being. In addition, there was no risk of physical, psychological, social, or economic harm to participants. Following consultation with the head of the Institute for Ethics in Technology at Hamburg University of Technology and a review by the German Association for Experimental Economic Research eV (certificate number kNjYR7Ag), it was confirmed that specific ethics approval was not required. This approach fully aligns with best practices for ensuring ethical compliance in socioeconomic research.

Descriptive Analysis

For this analysis, we included all DiGA listed in the DiGA directory at BfArM as of November 8, 2024. This study specifically aimed to compare how often medical benefits are used versus pSVV, as published in the DiGA directory [23]. To further refine our understanding of the adoption levels of pSVV, we also examined the use rates across the 9 key areas of pSVV detailed in the DiGA guide [27].

Qualitative Analysis

Transcripts were methodically analyzed in a 3-step procedure using the grounded theory and the well-established Gioia methodology [42,49] to ensure qualitative rigor in inductive research. After familiarization with the primary data, a codebook was developed by generating codes in an iterative process using the software MaxQDA (VERBI Software). SG coded all (19/19, 100%) interviews, and 2 (11%) randomly selected interviews were cocoded by MG. In addition, SG and MG regularly discussed and reflected on the data collection process and preliminary findings throughout all phases of the research. A consensus on all codes was eventually reached by the research team.

After agreeing on a coding scheme that best captured the diversity of the material, the transcripts and codes were analyzed again to identify second-order categories. First-order concepts included concepts, such as "lack of methods is the biggest hurdle for proof of pSVV" or "pSVV good ideas but difficult to operationalize." First-order concepts were combined into second-order themes, such as "study design for pSVV difficult due to lack of measurement instruments." Ultimately, second-order themes were merged hierarchically into five aggregate dimensions that best summarize the theoretical contribution in the data: (1) "key aspects for manufacturers within the regulated market access pathway," (2) "key aspects for study design," (3) "reasons for considering possibilities within new pathway," (4) "hurdles for considering possibilities within new pathway," and (5) "targeted innovative technology important for the health care system" (refer to Figure S1 and Table S3 in Multimedia Appendix 1).

Following the analysis, a triangulation of the results was conducted with 4 individuals from institutions involved in the process: 1 interview participant and 3 additional experts with extensive subject matter expertise, who had not participated in the initial interview study. All 4 experts were selected from the regulatory field, as many of the derived implications pertained to regulatory issues.

Functional-Structural Analysis

The first-order concepts, originally derived from the previous qualitative analysis as elucidated earlier, were subject to an additional analysis using the functional-structural systemic problem analyses framework [34], a methodology specifically tailored for the exploration of technology innovation systems [50,51]. Our analytical focus centered on the innovative pSVV technology, concentrating on the first-order concepts that laid the foundation for dimensions 2 to 5. Notably, the first dimension, "Key aspects for manufacturers within the regulated market access pathway," was excluded from this analysis, as it did not provide insights specific to the pSVV aspects of the DiGA pathway, but rather addressed the pathway in general.

In step 1, to facilitate a comprehensive analysis, all first-order concepts corresponding to dimensions 2 to 5 were systematically organized into a 4×7 matrix. This matrix integrates the 4 structural elements and 7 functional dimensions of an innovation system [34,50,52]. Following the analysis framework, the objective was to elucidate overarching “systemic problems” hindering the intended innovation within the innovation system for pSVV. Refer to Table S4 in Multimedia Appendix 1 for the comprehensive assignment of concepts to the functional and structural dimensions of the innovation system and the derivation of “systemic problems.” The sorting of concepts, derivation of instruments, and clustering into strategic factors were conducted by SG and subsequently agreed upon by the research team.

In step 2, to address the identified “systemic problems,” we first identified “systemic instruments” within the functional-structural framework of Wieczorek and Hekkert [34], tailored to specific systemic issues.

In step 3, we organized the “systemic instruments” identified in step 2 into 3 strategic factors to form actionable clusters. A detailed overview of the derivation process is available in Table S5 in Multimedia Appendix 1.

Results

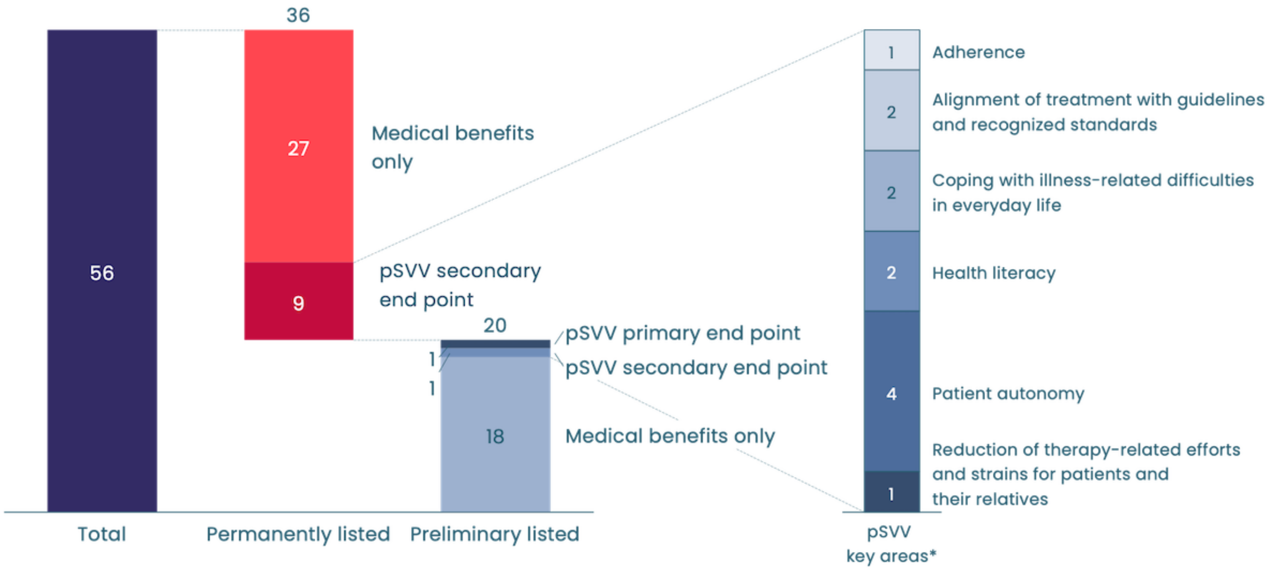
Current State of pSVV Adoption in the DiGA Market: Descriptive Analysis

As of November 8, 2024, out of the 56 DiGAs available in Germany, only 1 (2%) provisionally listed DiGA uses pSVV as its primary end point, with none in the permanently listed category. While 10 (19%) DiGAs incorporate pSVV as a secondary end point, the predominant focus remains on traditional medical benefits. The adoption of pSVV in comparison to medical benefits stands at 11 versus 45, meaning only 20% (11/56) of all currently listed DiGAs use pSVV. Moreover, only 1 (2%) of 56 DiGA exclusively focuses on pSVV as a primary end point [23].

The 11 (20%) of the 56 DiGA currently listed that use pSVV as a primary or secondary end point have collectively used 12 pSVV end points (Figure 1). Among these, 6 (67%) of the 9 possible key areas outlined in the DiGA guide have been used. In total, 4 (44%) of these areas have been used more than once, namely “patient autonomy,” “health literacy,” “coping with disease-related difficulties in everyday life,” and “alignment of treatment with guidelines and recognized standards.”

The 11 (20%) of the 56 DiGA currently listed that use pSVV are owned by 10 organizations, with 1 organization using pSVV end points in clinical trials to prove the positive care effect for 2 of their DiGA products. We estimated the timing of the adoption of the pSVV pathway by examining the time frame of clinical studies and found that 2 DiGA demonstrating a pSVV had already been conducting their clinical studies before the pathway was introduced in 2019 (velibra and vorvida, both developed by GAIA AG).

Figure 1. The number of digital health applications (DiGAs) listed in the DiGA directory in Germany based on positive health care effects and status of listing (as of November 8, 2024), including the breakdown of specific key areas used within patient-relevant structural and procedural improvements (pSVV). *Eleven DiGAs using pSVV use 12 pSVV end points with 1 DiGA using 2 pSVV key areas.

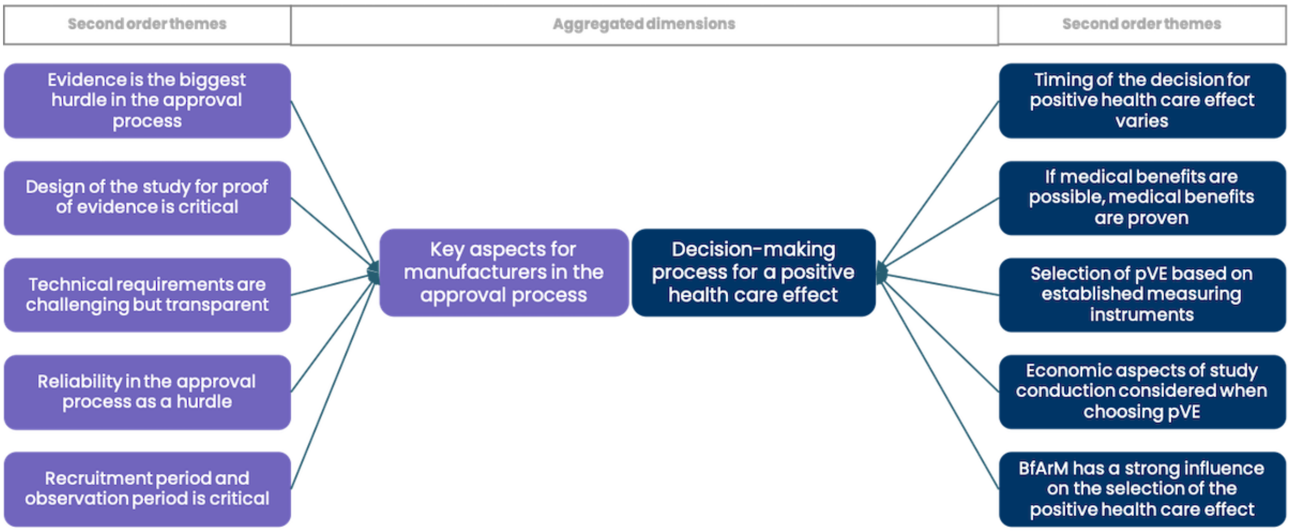


Analyzing Perception and Use of the pSVV Pathway by Key Stakeholders: Qualitative Analysis

Overview

The results are presented based on the data structure developed through the Gioia methodology (Figure S1 in Multimedia Appendix 1), which organizes findings into first-order concepts, second-order themes, and overarching dimensions. In the subsequent sections, findings are systematically presented separating the data structure into 3 segments of aggregated dimensions and further detailed by their corresponding second-order themes. To illustrate our approach, Figures 2-4 provide a visual representation of the respective parts of the data structure. In addition, Table S3 in Multimedia Appendix 1 includes selected codes that were instrumental in forming the data structure.

Figure 2. Graphical illustration of selected parts of the data structure derived from Gioia methodology illustrating key aspects for manufacturers within the regulated market access pathway as well as the key aspects of choosing the health care effect. BfArM: Federal Institute for Drugs and Medical Devices; pVE: positive health care effects.



Key Considerations in the Approval Process for DiGAs: Evidence as the Key Aspect

Being asked about the key considerations a DiGA manufacturer needs to face during the approval process, most (14/16, 88%) of the surveyed manufacturers and consultants concurred on a central theme—evidence takes the forefront, serving as the primary concern for manufacturers as “...(evidence) is the crucial point, which often determines the inclusion or non-inclusion (in the DiGA Directory)” (Consultant_5). In conjunction with this, the DiGA study design for evidence generation is deemed critical. The execution of the study is recognized as one of the primary cost drivers for manufacturers and the approving authority (BfArM) has a strong influence on the final design, as described by a participant, “And one can clearly see that an evidence level is crystallizing, which, within the framework that is possible, is already on the strictest side” (DiGA without pSVV_4).

While technical requirements, such as data privacy, interoperability, and accessibility were also mentioned, discussions in these domains with the BfArM were described as negligible due to the “clear and transparent” (Consultant_2) nature of the requirements. One respondent encapsulated this sentiment by stating the following:

> *The key issue is undoubtedly the matter of evidence. Less so is data privacy and security, as these are issues that can be adequately addressed.* [DiGA without pSVV_3]

Nevertheless, a certain level of planning uncertainty was identified in the process, stemming from updates to the DiGA guide or alterations in statements during consultations with the BfArM, as articulated by a manufacturer:

> *One of the most challenging aspects is the disparity between what is theoretically granted regarding research design and the actual decision practices of the BfArM.* [DiGA without pSVV_2]

In addition, insights into the procedural aspects revealed that both recruitment and observation periods are critically constrained.

Key Aspects of Study Design: Strong Tendency Toward Traditional End Points

Regarding the decision on which positive health care effect a future DiGA will be evaluated for and therefore which end point is used in the study design, significant timing variations are observed. DiGA manufacturers, especially those who began developing their product after the DVG was introduced in 2019, typically determine the positive health care effect early in the development phase, often before initiating the coding process:

> *Ideally, I make the decision before the development begins, as the entire development process needs to be built around this choice.* [DiGA with pSVV_1]

In contrast, manufacturers with established products leveraging the new DiGA approval pathway as a market entry point did not have this advantage.

In the selection of the clinical end point, manufacturers strongly lean toward medical benefits:

> *The medical aspects always take precedence.* [DiGA with pSVV_2]

Demonstrating a positive health care effect without proof of medical benefit is perceived as challenging in 2 ways. One way is to find the appropriate instrument for evidence generation for proving pSVV, “If I look at the list of pSVV, then I wonder for 80% of the things, how am I supposed to prove that?” (DiGA without pSVV_1), while the other lies in finding an agreement with the BfArM on a study design aiming to prove pSVV, “We have to advise relying on the medical benefit...it is not fundamentally that they (BfArM) reject it (pSVV), but we also experience that they...are not so certain about it” (Consultant_5). In addition, it is perceived that other stakeholders in the system are better equipped to evaluate medical benefits compared to pSVV, which could be relevant for price negotiations:

> *If you believe you can demonstrate medical benefits, that is probably the better path towards price negotiations because it's something the system understands and can perhaps evaluate better.* [Consultant_6]

The perceived challenges conclude in a perception of a stronger acceptance of evidence that substantiates a medical benefit, highlighting a potential current bias toward well-known, medically oriented outcomes.

Manufacturers and consultants choose the end points, if possible, based on the example of other DiGA that are already listed and where validated measurement instruments are available:

> *[The selection criteria are] essentially hard facts in the sense that it has worked before, and there are validated measurement instruments.* [Consultant_3]

In addition, both listed DiGA and unlisted health applications prioritize economic considerations:

> *Ultimately, we paid attention to...economic aspects, like the speed at which the study is conducted.* [DiGA without pSVV_2]

At present, pivotal guidance in the process emanates from consultations with the BfArM. The influence of the conversations is underscored by the following statement:

> *So as a manufacturer, you would never do anything where the BfArM advice said otherwise.* [DiGA without pSVV_5]

Reasons for Considering a New Pathway: Additional Evidence

Both, consultants and DiGA manufacturers, currently view the use of pSVV primarily as an opportunity for generating additional evidence alongside medical benefits. This perspective was succinctly expressed by one participant, who mentioned the following:

> *pSVV is included as a secondary endpoint in such a study. If possible, you confirm it, but you would always prefer to demonstrate medical benefits.* [Consultant_3]

DiGA manufacturers and consultants have pinpointed 3 substantial benefits associated with the implementation of the possibilities given within the new pSVV pathway. Potential differentiation from providers of similar DiGA was mentioned:

> *But in the case of [indication X], it could well be that with our DiGA, other DiGA are also approved or provisionally approved. In such instances, I could envision that a pSVV might serve as a consideration for or against a DiGA.* [DiGA without pSVV_4]

Overall positive impacts on sales were also mentioned:

> *Distribution is also a crucial task we undertake when engaging with physicians, ensuring a clear delineation of what is effective beyond standard practices. Particularly with DiGA, the argument hinges on the fact that the pSVV dimension precisely provides the benefit.* [DiGA without pSVV_3]

Moreover, there is an expectation of a positive impact on the official price negotiations occurring later in the process between manufacturers and payers. It is highlighted that "The more evidence you have, the greater the likelihood of maintaining certain price points" (DiGA with pSVV_1) and "Possibly, pSVV can indeed be an enhancing criterion (during the price negotiation process) for the respective product" (Regulatory body_2).

Hurdles for Considering a New Pathway: Operational Insecurities

When asked about the reasons why manufacturers tend to not use the new possibilities within the pSVV pathway, many (5/7, 71%) manufacturers and consultants referred to a lack of clarity on operationalizing pSVV as outlined in the DiGA guide, particularly concerning the methods applicable for assessing the impact:

> *But when I look at them [the pSVV key areas], I wonder with about 80% of them, how am I supposed to prove that? And yes, how do I actually incorporate that into studies?* [DiGA without pSVV_1]

In addition, there is skepticism among DiGA manufacturers and consultants regarding the economic impact of demonstrating pSVV. One participant succinctly expressed this viewpoint by stating the following:

> *Purely from the procedural and financial perspective—if you were to ask our boss, who is responsible for finances—there is a clear answer: this is an absolute disaster.* [DiGA with pSVV_1]

At the same time, introducing and establishing new measurement tools like the use of real-world data is deemed too expensive and risky for DiGA manufacturers:

> *If you had to essentially innovate methods and also deal with the uncertainty...in the worst-case scenario, you would end up with three studies.... That alone costs you a million, just for these studies. No one is willing to foot that bill. And then, if it goes wrong...* [Consultant_3]

In addition, manufacturers and consultants observed a general uncertainty within the system regarding the concept of pSVV:

> *Not that they (BfArM) fundamentally reject it, but we experience that they, how should I say, are not so certain about it.* [Consultant_5]

Moreover, they observed a general lack of knowledge about pSVV that leads to the need for explanation:

> *There is a significant need for explanation, actually, among the different stakeholders. Both with patients and with doctors. What is it exactly? Why is it good? What is in it for me?* [DiGA with pSVV_1]

Consultancies sometimes actively discourage the choice of pSVV in the clinical trial:

> *If you believe you can demonstrate medical benefits, that is probably the better path towards price negotiations because it's something the system*

> *understands and can perhaps evaluate better.* [Consultant_6]

Relevance Perception: Targeted Innovative Technology Is Important for the Health Care System

Looking at the general perception of pSVV as a concept, stakeholders highlighted the systemic benefits, emphasizing the impact on the overall health care system efficiency and supply effects:

> *It...makes sense from the overall health care system perspective.* [Non-DiGA_2]

Patient-centric advantages emerged, revealing the influence of DiGA beyond conventional medical outcomes. A consultant illustrated this by pointing out the positive impact of patient empowerment on intervention success:

> *Individuals with higher health literacy typically respond more effectively to interventions, as they have a better understanding and feel a greater sense of control.* [Consultant_1]

DiGA might also address practical challenges like long waiting times, presenting substantial value beyond traditional medical considerations:

> *But if we now say we would compare ourselves to a traditional intervention...then the pSVV...plays an even greater role. Because when I say, I have something here with a positive effect, and it is more accessible than the therapy, that is already a value where I say, it makes a difference.* [DiGA with pSVV_2]

However, the unique benefits of DiGA, such as the improvement of patient adherence and reduction of therapy-related efforts, often eluded recognition in conventional benefit assessments:

> *pSVV is crucial for the system, but the evidence must focus on it. Thus, the usability of the product must be ensured and tailored to that. The evidence must reflect the particular added value of these applications.* [DiGA without pSVV_5]

Regulatory bodies acknowledged this oversight, emphasizing the need to broaden evaluation criteria to encompass these critical aspects:

> *DiGA can have medical benefits, but they are not solely limited to medical utility.... And these are points that get lost in traditional benefit assessments.* [Regulatory body_1]

In addition, the political and collaborative nature of the development process that led to the pSVV concept was highlighted: “We had many stakeholders involved there. On the one hand, we had colleagues from self-administration, but we also had the manufacturers [stretching the collective effort] in understanding potential, addressing challenges, and bridging the gap between traditional assessment methods and innovative logics” (Regulatory body_1).

Analyzing the pSVV Technology Innovation System: 5 Systemic Problems Identified

Actors’ Presence Problem in “Entrepreneurial Activities” and “Market Formation”

Examining the participants presently engaged in the pSVV innovation system reveals a discernible pattern. Only one of the DiGA entrepreneurs within this emergent system has developed a product with a primary emphasis on pSVV. Notably, some manufacturers have established market longevity before the introduction of the DVG and thus pSVV; therefore, orienting their product toward medical benefits. Moreover, some actors are concurrently navigating evidence-generation endeavors while exploring alternative pathways for market access in Germany, such as individual “selective contracts” with statutory health insurance under §140a SGB V (Sozialgesetzbuch 5; a German law that covers all regulation concerning statutory health insurance and therefore also covers DiGA and selective contracts), which lacks the concept of pSVV. The absence of *entrepreneurial activities* singularly centered on pSVV emerges as a contributory factor hindering the *formation of a distinct market* for dedicated pSVV-related regulated digital medical devices.

Actors’ Capability Problem in “Knowledge Dissemination”

The supervising governmental institution, in its role as advisor for manufacturers before and during the admission process, has limited experience with pSVV but rather a strong focus on medical benefits. This emphasis can be attributed to its dual role, as the BfArM in Germany oversees admissions not only for DiGAs but also for traditional pharmaceuticals. The lack of experience with pSVV in this key role as a knowledge multiplicator and also in this specific innovation system hampers *knowledge dissemination*. In addition, practitioners prescribing DiGA as well as end users of the products have limited experience with products focusing on pSVV, which leads to a need for significant efforts to explain the concept while the demand from potential users is unclear.

Infrastructure Presence Problem in “Knowledge Development” and “Guidance of Search”

The suitability of existing measurement methods designed for therapeutic effects in assessing pSVV is often limited, contributing to its restricted adoption. Simultaneously, the lack of positive examples for products in the pSVV category hampers *knowledge development*. In addition, the absence of positive price examples not only further constrains manufacturers’ support for pSVV but also contradicts the *guidance of search*.

Institution Intensity Problem in “Knowledge Development” and “Resources Mobilization”

In the *institutional* dimension, it was noted in particular that the existing *instructions* (DiGA guide) pose challenges to the operationalization of pSVV. The *established practices* and formed *expectations* regarding pSVV are perceived as counterproductive, particularly for cost-sensitive manufacturers who cannot afford to allocate their limited resources toward pSVV despite existing barriers. The limitations of the DiGA guide concerning pSVV, namely lack of examples or standards and possible end points for evidence demonstration, are identified as a challenge in *knowledge development*, while the *established practices*, inclined toward medical benefits, hinder *resource allocation* for the advancement of pSVV.

Institution Intensity and Interaction Intensity Problem in “Guidance of the Search,” “Market Formation,” and “Creation of Legitimacy”

Established practices in the approval process that make it difficult to enter the market without proven medical benefits diverge from the introduced laws and regulations, creating a misalignment. Interactions within the approval process, especially with the official advisory body, tend to emphasize medical benefits. Notably, consultancies actively guide manufacturers toward prioritizing medical benefits over pSVV, thus steering the guidance of the search away from pSVV and hindering market formation.

Positive expectations on the value of using pSVV that would support the *creation of legitimacy* remain mainly hypothetical. The economic effect, specifically, was formulated in the subjunctive by interviewees. Coupled with the pervasive *established practices*, this poses a significant obstacle to *creating legitimacy* for pSVV.

Discussion

Principal Findings

Our research explored the factors that need to be considered when bringing a specific digital health technology innovation to market through the implementation of a regulated market access pathway as a health policy approach.

For this example case study, the pSVV pathway, the descriptive analysis revealed that only a few DiGAs currently in the market have demonstrated a pSVV. Of the 11 currently listed DiGAs that have entered the market with a pSVV as one element of a positive health care effect since the introduction of the pathway in 2019, only 1 DiGA has a primary focus on pSVV (ie, pSVV as the primary end point of the positive health care effect). Notably, 2 of the products (developed by the same DiGA manufacturer) that demonstrate pSVV as a secondary end point conducted their clinical studies before 2019. This indicates that few manufacturers have taken the initiative to develop a product with a special focus on improving structures and procedures for patients, even though the introduction of the pSVV pathway was intended to encourage this. To date, pSVV is mostly used as an additional end point to medical benefits. To provide a basis for interpreting the results in further analyses on the performance factors of innovation systems shaped by regulated market access pathways, our initial conclusion is that the innovation in DiGAs targeting pSVV, which the pSVV pathway was designed to foster, has not yet been fully realized within the system. This is particularly daunting, as previous studies emphasize the need to focus more on products that address structural and procedural improvements for patients [53-55] and consider the concept of pSVV to be a “fundamental mechanism of action for DiGAs” [56].

Our qualitative analysis of stakeholders’ perspectives during market access reveals that evidence generation is the most critical component of the DiGA approval process. While multiple factors are considered in determining a positive health care effect, our findings suggest that, at the operational level, manufacturers of DiGAs primarily focus on demonstrating medical benefits as evidence for positive health care outcomes (Figure 2). Economic considerations, in light of the lack of established methodologies for measuring the defined key areas and uncertainty about the acceptability of the proposed study design, including pSVV, lead to more cautious use. Bearing the cost for methodological innovation is not considered attractive among manufacturers, particularly as DiGAs with pSVV so far do not negotiate substantially higher prices than DiGAs without pSVV [28]. Consistent with the findings of the descriptive analysis, the use of pSVV is often considered “incidentally” as an addition to medical benefit, only when a methodology for demonstrating one of the pSVV key areas exists for a specific product (Figure 3). It can be concluded that the specific challenges associated with proving the efficacy of pSVV, particularly the lack of established methodologies, currently outweigh its potential benefits. Previous studies have highlighted content and format security [57] as a critical factor, particularly for pioneers in medical technology, which also appears to hold true for the market entry of innovative products in DiGAs. The fact that only one company has attempted to pioneer a product specifically targeting benefits under pSVV may be attributed to the innovators’ need for content and format security, in this case specifically the lack of standards to generate evidence for the targeted technology innovation. Without such security, the costs for pioneers become prohibitively high.

However, the implementation of the pSVV concept is recognized as vital for the health care system’s evolution itself (Figure 4).

Our qualitative analysis highlights a common gap in health care innovation between the perceived importance of new technologies and their actual implementation. This gap is often influenced by systemic factors, such as regulatory conditions and inherent system inertia, alongside a general openness to innovation among stakeholders [58]. For the pSVV pathway, while there is strong stakeholder support for introducing products that demonstrate pSVV, significant operational hurdles, particularly concerning content and format security, have delayed its adoption. These barriers suggest that, despite a willingness to embrace this innovation, the pathway’s practical challenges currently limit its broader impact (Figure 5).

Figure 3. Graphical illustration of selected parts of the data structure derived from Gioia methodology illustrating hurdles of DiGA manufacturers for considering possibilities within the new pathway as well as reasons for doing so. DiGA: digital health application; pSVV: patient-relevant structural and procedural improvement.

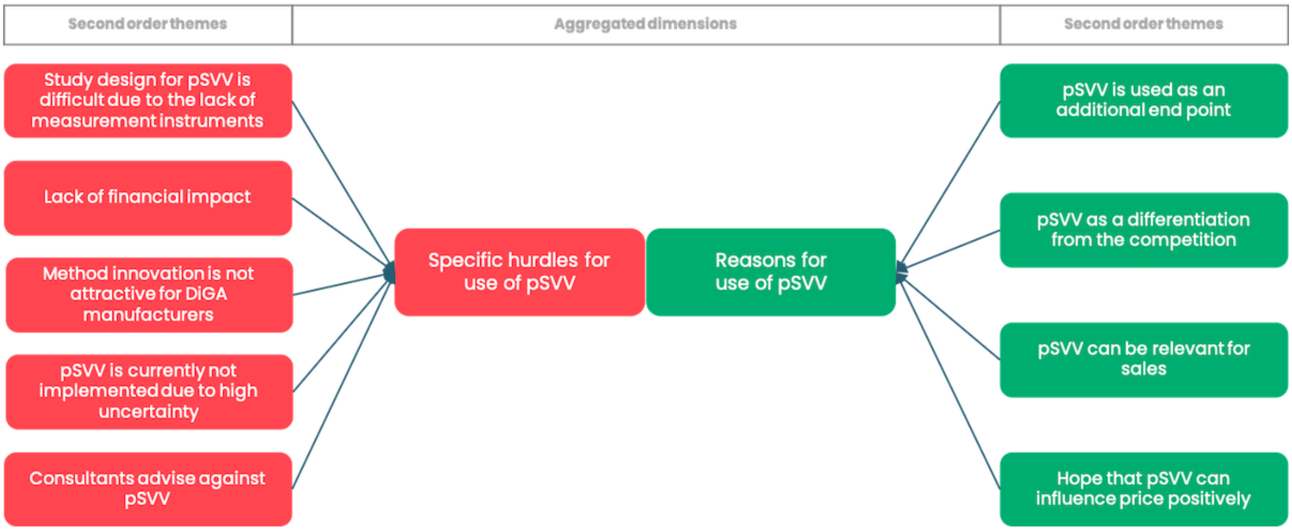
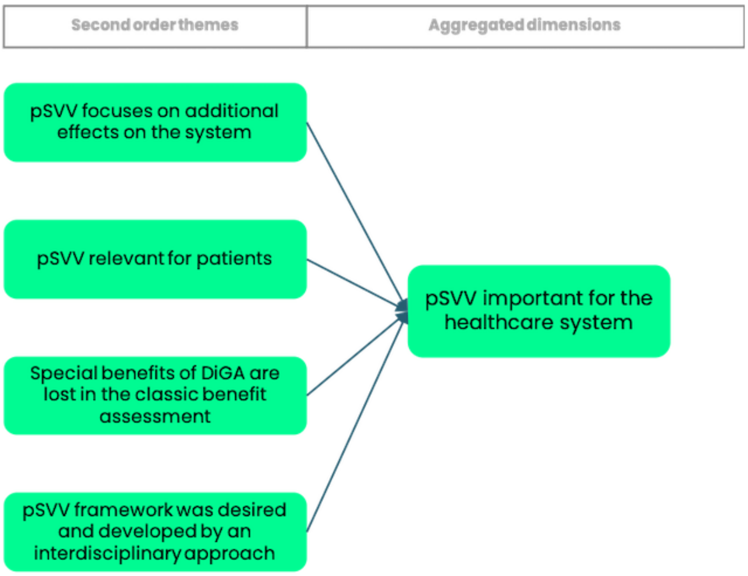


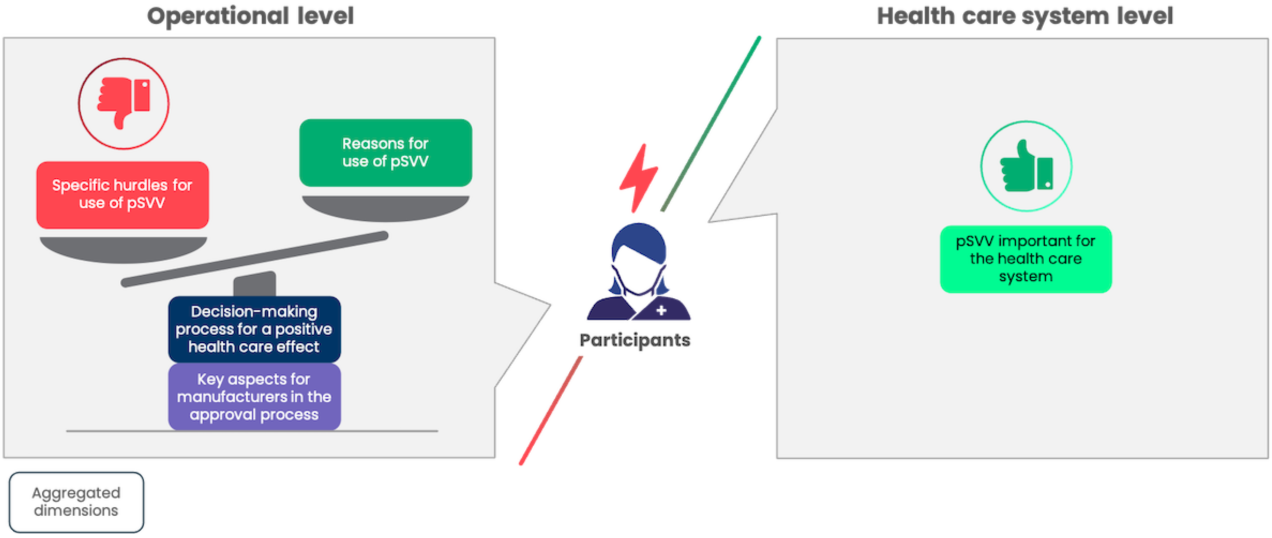
Figure 4. Graphical illustration of a selected part of the data structure derived from Gioia methodology illustrating the perception of the importance of patient-relevant structural and procedural improvements (pSVVs) in the health care system. DiGA: digital health application.



To further explore performance factors in the innovation system shaped by the pSVV pathway, we conducted an analysis of the innovation system's maturity level in relation to pSVV. Our findings indicate 3 central strategic factors of an innovation system in which we can expedite the adoption of innovative technology: *framework specification*, *active innovation processmanagement*, and *market formation stimulation*

While previous work has highlighted the general need for frameworks [16,59], our emphasis is on the degree of specificity required for the successful implementation of a market access pathway for DiGAs. In our example, the specification of the pSVV pathway, defined in the DiGA guide, requires 2D adjustments. First, the creation of a knowledge infrastructure would enable entrepreneurs to navigate the DiGA market, including pSVV, better if the development of exemplary models for pSVV products and especially congruent methodologies for demonstrating the evidence base of pSVV would exist. In addition, a multitude of studies highlight the ongoing challenges associated with identifying appropriate study designs that are capable of measuring effects solely attributable to the use of digital products [60-62]. We conclude that the presence of a specific level of knowledge infrastructure, which underpins the innovation driven by the introduction of a regulated market access pathway, can affect how widely it is adopted [6,63].

Figure 5. Underlying conflict leading to the underuse of patient-relevant structural and procedural improvements (pSVVs).



Second, studies in the field of regulated medical products have indicated that the uncertainty about the accepted content and format during the approval of new products can make the pathway into the market unattractive for first movers [57]. Our analysis has shown that the DiGA guide, which serves as the most relevant instruction at the operational level in this system, requires timely updates to keep pace with advancing knowledge development and ensuring up-to-date format and content security. Simultaneously, for the successful implementation of a pathway, *active innovation process management* is crucial. Providing training to key actors in the defined process, such as the BfArM or consultants, has been shown to be essential for empowering stakeholders to support and advance pSVV as an innovation. In addition, as previous work has highlighted the importance of education and awareness as well as patient and health care provider support while training for physicians on digital health solutions is limited [25,64], the inclusion of user groups and prescribing physicians can be focused to sharpen the objectives of pSVV. Conversely, it is crucial to ensure that involved actors are not overly familiar with traditional processes, as this can lead to inflexible adherence to these methods. The involvement of key actors with an excess of experience in medical benefits (ie, experienced regulatory assessor previously involved in assessing pharmaceutics) could potentially impede innovation speed if their participation is overly predominant.

Furthermore, targeted *market formation stimulation* can aim to introduce manufacturers into the DiGA market, focusing primarily on addressing pSVV key areas [65]. Creating positive examples can accelerate innovation at multiple levels. In addition, stimulating constructive exchange regarding the pSVV concept can expedite knowledge generation and distribution, contributing to faster market formation.

In summary, our study revealed that the lack of established methodologies for proving pSVV has led to significant uncertainty for manufacturers regarding the acceptable formats and evidence required, contributing to slower adoption. To complement these findings on the specific pathway level, the functional-structural analysis identified 3 key factors critical to enhancing the total innovation system: clearer framework specification to provide precise guidelines, active innovation process management to support stakeholders in navigating the regulatory process, and market formation stimulation to foster positive examples and increase engagement with the pSVV pathway. These factors are essential for overcoming both the operational and systemic barriers to successfully introducing innovative digital health technologies into a market.

As initiators of the systemic innovation of pSVV, the regulatory body can implement additional measures to promote its further establishment. This includes stimulating the development of a knowledge infrastructure by involving all stakeholders in the innovation system. Academia can actively contribute to advancing the definition of the pSVV and particularly developing suitable methodologies for evidence generation [66-68]. Existing mechanisms to promote scientific progress, such as dedicated research funds (eg, in Germany by the Federal Ministry of Education and Research or the Innovation Fund of the Federal Joint Committee), could be set up for this objective [69].

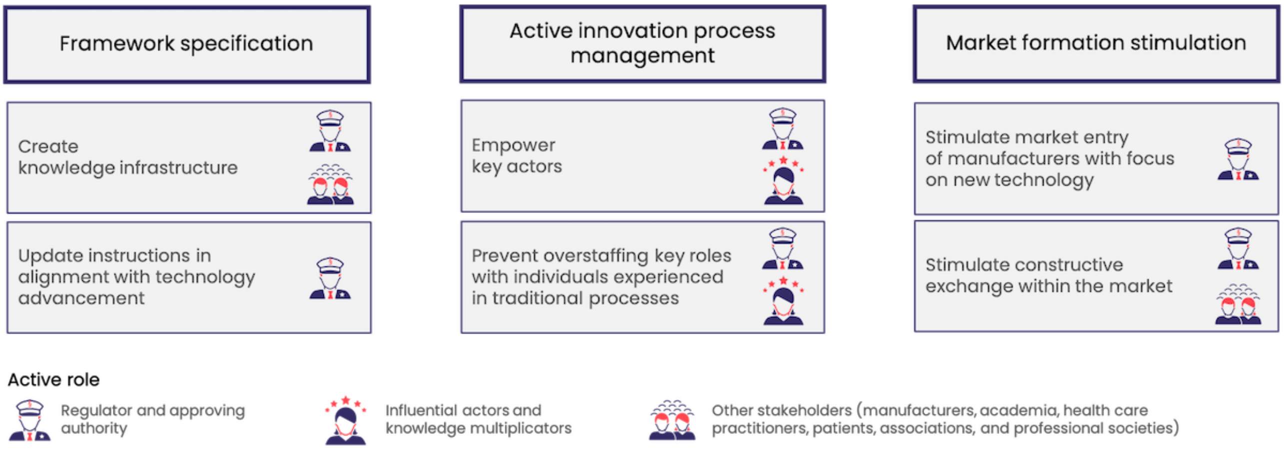
Initiating a public debate on pSVV can engage relevant stakeholders, such as professional societies and associations in a discourse to further develop the pSVV framework [13,66]. Economic measures, such as granting financial support for conducting clinical trials assessing pSVV or even tax incentives for implementation, can provide additional incentives for manufacturers driving innovation and initiating the pSVV market until these products are established in the standard of care. Furthermore, spaces can be created for key stakeholders in the system to educate themselves about pSVV. This could include offering policy laboratories or training programs for knowledge multipliers.

Consultancies and the BfArM, as influential players in the system, can provide room for their employees to further educate themselves about the possibilities of pSVV, to prevent disproportionately biased inclinations toward medical benefits. Ensuring a balanced wealth of experience regarding the use of medical benefits within the workforce can promote openness to innovative processes. It is critical to ensure that the experience of the staff responsible for evaluating the proof of benefit of

DiGAs is balanced in assessing medical quality [70]. Previous work has shown that the Food and Drug Administration might actively aid in similar innovation processes [71]. This assistance is rendered through the administration's in-depth knowledge to improve study designs and by championing innovative methods. Although the breadth of authority and scope of the regulated processes vary, BfArM or other regulatory agencies involved in the approval process and specifically in evaluating coverage eligibility can also contribute their expertise. For manufacturers seeking to use the potential of pSVV in their DiGAs, it is important to prioritize evidence of positive health care effects early in the development process. Identifying the "active ingredient" and its mechanism is crucial for crafting studies effectively assessing DiGA [60].

Besides proposing solutions for individual stakeholder groups within the health care system (Figure 6), the collaboration and coordination among all involved parties should be underscored as crucial for success. As highlighted on an international scale, effective collaboration between regulatory agencies, academia, app developers, payers, and health care providers is essential to truly leverage advances in science and technology and to translate the vision into reality, guided by the right incentives and guidelines [69,71,72].

Figure 6. Central strategic factors for progress exploration in the innovation system derived from an exemplary analysis of the patient-relevant structural and procedural improvement pathway.



Our analysis suggests that the strategic factors identified are valuable considerations for the establishment of a regulatory market access pathway, to ensure its widespread and successful adoption. Drawing on previous research, the strategic factor of content and format security [57] is highlighted, alongside the adoption of recognized and accepted methods for measuring proposed parameters [59,61,68,73-76]. This is particularly relevant in digital health technology aimed at patient empowerment, where significant progress is still needed [68,74]. Furthermore, it is advantageous to strike an optimal balance between leveraging established processes and frameworks allowing room for innovation [77]. As indicated by the strategic factor of innovation process management, it is crucial to carefully select and train key players and individuals in pivotal roles, equipping them with the specific skills and mindset needed to drive innovation. Finally, the strategic factor of market formation stimulation points to the critical need to consider the practicality of a framework in enabling a viable business model, particularly within the context of systemic innovation [21,60,78].

Limitations

Acknowledging the limitations of this study, a conclusive assessment of the economic value of pSVV is precluded, as no product with pSVV as a primary end point has undergone final price negotiation yet (Table S2 in Multimedia Appendix 1). In addition, it could be questioned whether pSVV can ever be the primary end point of a DiGA, as all DiGAs must also be CE–certified medical devices with the corresponding intended purpose. We have not explored this discussion in detail, as it can certainly be conducted more effectively with selected examples of products focusing on pSVV, once these are available in the market. Furthermore, our qualitative analysis, conducted through semistructured interviews, did not encompass all functional-structural dimensions of innovation systems. Entities, such as academia or associations were not mentioned by any interviewees and were thus not part of the evaluation. In addition, our sampling strategy did not include patients or physicians, as the study concentrated on identifying reasons for the scarce number of products targeting the pSVV key focus areas. This decision acknowledges the limited familiarity of end users and prescribers with the pSVV concept at this point and their marginal impact on market entry decisions shortly after the introduction of a new framework designed to facilitate market access.

Conclusions

This exploratory study examined critical factors for bringing digital health technology innovation to market through a regulated market access pathway, with a focus on the pSVV pathway as an example case. Key success factors identified included content and format security, the need for precise framework specification, active innovation process management, and market formation stimulation. These elements collectively help reduce uncertainties for manufacturers and promote broader adoption of digital health innovations. Our findings revealed that although the pSVV pathway was intended to encourage a focus on pSVVs, the current use remains limited, with most manufacturers prioritizing medical benefits as evidence of positive health care outcomes.

The mixed methods approach enriched the analysis by allowing a deep exploration of stakeholder perspectives through qualitative analysis, complemented by the functional-structural analysis of the technology innovation system. The combination of grounded theory and Gioia methodology enabled an objective and detailed examination of how market participants view and use the pSVV pathway, while the systemic analysis highlighted additional success factors in the broader innovation system context.

Our findings contribute valuable insights for stakeholders in the digital health sector, providing recommendations for policy adjustments that can strengthen innovation pathways. With this work, we contribute by identifying critical areas for improvement, including the need for clearer guidelines, the development of a knowledge infrastructure, and enhanced training for stakeholders to help overcome operational barriers. In addition, fostering public discourse on pSVV and introducing economic incentives, such as financial support for clinical trials, may further drive innovation. These recommendations are valuable for health care systems and regulatory bodies seeking to support similar pathways and advance patient-centered digital health solutions.

Furthermore, the impartial perspectives gathered from stakeholders, combined with a holistic analysis of the innovation system, contribute to identifying clear fields for action and proposing solutions relevant beyond this specific case. Our findings provide transferable insights for health care systems embarking on similar systemic innovations, as seen in countries like France, Belgium, and Austria, where comparable concepts to pSVV are being introduced. The implications of our research extend to other systemic innovation initiatives in health care systems facing significant adoption challenges that could otherwise lead to initiative failure. Ongoing efforts from national and international scholars are critical in supporting the success of these initiatives. In Germany, for example, a reimbursement pathway for digital nursing care applications (DiPA), a “regulatory sibling” to DiGA, has been established. While speculative, similar challenges may emerge for DiPA manufacturers, particularly concerning the nursing benefit that aligns with both medical benefit and pSVV, possibly contributing to the current absence of DiPA in the marketplace and highlighting gaps in this systemic innovation approach.

Acknowledgments

The authors are grateful to the 2 anonymous reviewers whose insightful comments and suggestions contributed substantially to the quality of this manuscript. Furthermore, the authors extend their heartfelt thanks to all the participants of this study, whose time, effort, and invaluable contributions made this research possible.

The study was part of a doctorate program at our university. Publishing fees were supported by the Funding Programme Open Access Publishing of the Hamburg University of Technology.

The authors used generative artificial intelligence (ChatGPT, version 4o; OpenAI) to incorporate minor improvements in sentence structure and wording throughout the publication; the contents and conclusions have purely been developed by the authors [79].

Data Availability

The datasets generated and analyzed during this study are not publicly available due to data protection reasons but are available from the corresponding author on reasonable request.

Authors' Contributions

SG conceived and designed the study, interviewed the participants coded the interviews, analyzed the data, and drafted the manuscript. MG transcribed the interviews using Whisper Transcription (Good Snooze). MG cocoded on randomly selected interviews and took part in reflexive discussion and consensus building together with SG. MG supervised the project. SG and MG interpreted the results and finalized the manuscript. All authors approved the final manuscript.

Conflicts of Interest

SG is employed at a statutory health insurance company in Germany. She declares that her work at the statutory health insurance is unrelated to this topic. MG has previously worked as a consultant for digital health application (DiGA) manufacturers. He declares that his previous work is unrelated to the results of the conducted work.

Multimedia Appendix 1

Semistructured interview questionnaires, overview of interview participants, full data structure based on Gioia methodology, negotiated digital health application prices and health care effects, selected interview codes generated using grounded theory, functional-structural analysis of patient-relevant structural and procedural improvement as a systemic innovation, and derivation of strategic fields using the functional-structural framework by Wieczorek and Hekkert [34].

[DOCX File , 651 KB - jmir_v27i1e66356_app1.docx]

References

1. Haring M, Freigang F, Amelung V, Gersch M. What can healthcare systems learn from looking at tensions in innovation processes? A systematic literature review. BMC Health Serv Res 2022 Oct 28;22(1):1299 [FREE Full text] [doi: 10.1186/s12913-022-08626-7] [Medline: 36307839]
2. Siopis G, Moschonis G, Eweka E, Jung J, Kwasnicka D, Asare BY, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with hypertension: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023;5(3):e144-e159 [FREE Full text] [doi: 10.1016/S2589-7500(23)00002-X] [Medline: 36828607]
3. Hunter DJ, Bengoa R. Meeting the challenge of health system transformation in European countries. Policy Soc 2023 Mar;42(1):14-27 [FREE Full text] [doi: 10.1093/polsoc/puac022]
4. Coiera E. Why system inertia makes health reform so difficult. BMJ 2011 Jun 23;342:d3693. [doi: 10.1136/bmj.d3693] [Medline: 21700652]
5. Gilbert S, Anderson S, Daumer M, Li P, Melvin T, Williams R. Learning from experience and finding the right balance in the governance of artificial intelligence and digital health technologies. J Med Internet Res 2023 Apr 14;25:e43682 [FREE Full text] [doi: 10.2196/43682] [Medline: 37058329]
6. Coder M, McBride L, McClenahan S. Core elements of national policy for digital health technology evidence and access. NPJ Digit Med 2024 Aug 13;7(1):212 [FREE Full text] [doi: 10.1038/s41746-024-01209-2] [Medline: 39138261]
7. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
8. Kotecha D, Chua WW, Fabritz L, Hendriks J, Casadei B, Schotten U, et al. European Society of Cardiology smartphone and tablet applications for patients with atrial fibrillation and their health care providers. Europace 2018 Feb 01;20(2):225-233 [FREE Full text] [doi: 10.1093/europace/eux299] [Medline: 29040548]
9. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023 Mar 10;6(1):38 [FREE Full text] [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]
10. Weik L, Fehring L, Mortsiefer A, Meister S. Understanding inherent influencing factors to digital health adoption in general practices through a mixed-methods analysis. NPJ Digit Med 2024 Feb 27;7(1):47 [FREE Full text] [doi: 10.1038/s41746-024-01049-0] [Medline: 38413767]
11. 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]
12. Larson RS. A path to better-quality mHealth apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414 [FREE Full text] [doi: 10.2196/10414] [Medline: 30061091]
13. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
14. Sedhom R, McShea MJ, Cohen AB, Webster JA, Mathews SC. Mobile app validation: a digital health scorecard approach. NPJ Digit Med 2021 Jul 15;4(1):111 [FREE Full text] [doi: 10.1038/s41746-021-00476-7] [Medline: 34267296]
15. Enforcement policy for digital health devices for treating psychiatric disorders during the coronavirus disease 2019 (COVID-19) public health emergency. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Office of Product Evaluation and Quality. 2020 Apr. URL: https://downloads.regulations.gov/FDA-2020-D-1138-0068/attachment_1.pdf [accessed 2024-12-18]
16. Essén A, Stern AD, Haase CB, Car J, Greaves F, Paparova D, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):31 [FREE Full text] [doi: 10.1038/s41746-022-00573-1] [Medline: 35304561]
17. Ju JH, Sim B, Lee J, Lee JY. Reimbursement of digital therapeutics: future perspectives in Korea. Korean Circ J 2022 Apr;52(4):265-279 [FREE Full text] [doi: 10.4070/kcj.2022.0014] [Medline: 35388995]
18. Nomura A. Digital health, digital medicine, and digital therapeutics in cardiology: current evidence and future perspective in Japan. Hypertens Res 2023 Sep;46(9):2126-2134 [FREE Full text] [doi: 10.1038/s41440-023-01317-8] [Medline: 37258624]
19. Ludewig G, Klose C, Hunze L, Matenaar S. Digital health applications: statutory introduction of patient-centred digital innovations into healthcare. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021;64(10):1198-1206 [FREE Full text] [doi: 10.1007/s00103-021-03407-9] [Medline: 34529096]
20. Martin T, Hervias A, Armoiry X, Martelli N. Early access programs for medical devices in France: Overview of recent reforms and outcomes (2015-2022). Health Policy 2024 Oct;148:105146 [FREE Full text] [doi: 10.1016/j.healthpol.2024.105146] [Medline: 39154454]
21. Stern AD, Matthies H, Hagen J, Brönnecke JB, Debatin JF. Want to see the future of digital health tools? Look to Germany. Harvard Business Review. 2020 Dec 2. URL: https://hbr.org/2020/12/want-to-see-the-future-of-digital-health-tools-look-to-germany [accessed 2024-12-18]

22. Hullegie P, Klein TJ. The effect of private health insurance on medical care utilization and self-assessed health in Germany. Health Econ 2010 Sep;19(9):1048-1062. [doi: 10.1002/hec.1642] [Medline: 20665706]
23. DiGA-Verzeichnis. Bundesinstitut für Arzneimittel und Medizinprodukte. URL: https://diga.bfarm.de/de/verzeichnis [accessed 2024-11-08]
24. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020 Jul 10;3(1):94 [FREE Full text] [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
25. 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]
26. Lauer W, Löbker W, Höfgen B. Digital health applications (DiGA): assessment of reimbursability by means of the "DiGA Fast Track" procedure at the Federal Institute for Drugs and Medical Devices (BfArM). Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1232-1240 [FREE Full text] [doi: 10.1007/s00103-021-03409-7] [Medline: 34529095]
27. The fast-track process for digital health applications (DiGA) according to section 139e SGB V. A guide for manufacturers, service providers and users. Bundesinstitut für Arzneimittel und Medizinprodukte. URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.html [accessed 2024-12-18]
28. Goeldner M, Gehder S. Digital health applications (DiGAs) on a fast track: insights from a data-driven analysis of prescribable digital therapeutics in Germany from 2020 to mid-2024. J Med Internet Res 2024 Aug 29;26:e59013. [doi: 10.2196/59013] [Medline: 39208415]
29. Schmidt L, Pawlitzki M, Renard BY, Meuth SG, Masanneck L. The three-year evolution of Germany's Digital Therapeutics reimbursement program and its path forward. NPJ Digit Med 2024 May 24;7(1):139 [FREE Full text] [doi: 10.1038/s41746-024-01137-1] [Medline: 38789620]
30. Schramm L, Carbon CC. Critical success factors for creating sustainable digital health applications: a systematic review of the German case. Digital Health 2024;10:20552076241249604 [FREE Full text] [doi: 10.1177/20552076241249604] [Medline: 38665883]
31. Woiceshyn J, Eriksson P. How innovation systems in Finland and Alberta work: lessons for policy and practice. Innovation 2014;16(1):19-31. [doi: 10.5172/impp.2014.16.1.19]
32. Smits R, Kuhlmann S. The rise of systemic instruments in innovation policy. Int J Foresight Innov Policy 2004;1(1/2):4. [doi: 10.1504/ijfip.2004.004621]
33. van Mierlo B, Leeuwis C, Smits R, Woolthuis RK. Learning towards system innovation: evaluating a systemic instrument. Technol Forecast Soc Change 2010 Feb;77(2):318-334. [doi: 10.1016/j.techfore.2009.08.004]
34. Wieczorek AJ, Hekkert MP. Systemic instruments for systemic innovation problems: a framework for policy makers and innovation scholars. Sci Public Policy 2012 Feb 06;39(1):74-87. [doi: 10.1093/scipol/scr008]
35. Hekkert MP, Suurs RA, Negro SO, Kuhlmann S, Smits RE. Functions of innovation systems: a new approach for analysing technological change. Technol Forecast Soc Change 2007 May;74(4):413-432. [doi: 10.1016/j.techfore.2006.03.002]
36. General methods: version 6.1 of 24 January 2022. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. 2022. URL: https://www.iqwig.de/methoden/general-methods_version-6-1.pdf [accessed 2024-12-18]
37. Dadgar M, Joshi KD. 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). [doi: 10.17705/jais1.00485]
38. Fürstenau D, Gersch M, Schreiter S. Digital therapeutics (DTx). Bus Inf Syst Eng 2023 Apr 06;65:349-360. [doi: 10.1007/s12599-023-00804-z]
39. Goodday SM, Geddes JR, Friend SH. Disrupting the power balance between doctors and patients in the digital era. Lancet Digit Health 2021 Mar;3(3):e142-e143 [FREE Full text] [doi: 10.1016/S2589-7500(21)00004-2] [Medline: 33516653]
40. Ryll B. Digital health applications-patient-centric care with disruptive potential. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1207-1212 [FREE Full text] [doi: 10.1007/s00103-021-03421-x] [Medline: 34477883]
41. Loeb S, Dynarski S, McFarland D, Morris P, Reardon S, Reber S. Descriptive analysis in education: a guide for researchers. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. 2017. URL: https://files.eric.ed.gov/fulltext/ED573325.pdf [accessed 2024-12-18]
42. Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research: notes on the Gioia Methodology. Org Res Methods 2012 Jul 24;16(1):15-31. [doi: 10.1177/1094428112452151]
43. Magnani G, Gioia D. Using the Gioia Methodology in international business and entrepreneurship research. Int Bus Rev 2023 Apr;32(2):102097. [doi: 10.1016/j.ibusrev.2022.102097]
44. Tracy SJ. Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact. Hoboken, NJ: John Wiley & Sons; 2019.
45. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: SAGE Publications; 1990.
46. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Quality Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

47. Gehder S, Goeldner M. Structural benefits on prescription? Exploring potentials and challenges of the pSVV concept in the approval process of digital health applications in Germany. Z Evid Fortbild Qual Gesundhwes 2024 Dec 16 (forthcoming) [FREE Full text] [doi: 10.1016/j.zefq.2024.11.003] [Medline: 39690098]
48. Gabler S. Schneeballverfahren und verwandte Stichprobendesigns. ZUMA Nachrichten 1992;16(31):47-69. [doi: 10.1007/978-3-322-89527-1_4]
49. Qureshi HA, Ünlü Z. Beyond the paradigm conflicts: a four-step coding instrument for grounded theory. Int J Qual Methods 2020 Jun 02;19. [doi: 10.1177/1609406920928188]
50. Hekkert M, Suurs R, Negro S, Kuhlmann S, Smits R. Functions of innovation systems: a new approach for analysing technological change. In: Foxon T, Köhler J, Oughton C, editors. Technological Forecasting and Social Change. Cheltenham, UK: Edward Elgar Publishing; May 2007:413-432.
51. Bergek A, Jacobsson S, Carlsson B, Lindmark S, Rickne A. Analyzing the functional dynamics of technological innovation systems: a scheme of analysis. Res Policy 2008 Apr;37(3):407-429. [doi: 10.1016/j.respol.2007.12.003]
52. Hekkert MP, Janssen MJ, Wesseling JH, Negro SO. Mission-oriented innovation systems. Environ Innov Soc Transitions 2020 Mar;34:76-79. [doi: 10.1016/j.eist.2019.11.011]
53. Wangler J, Jansky M. What potential and added value do DiGA offer for primary care?-Results of a survey of general practitioners in Germany. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2022 Dec;65(12):1334-1343 [FREE Full text] [doi: 10.1007/s00103-022-03608-w] [Medline: 36269336]
54. Mäder M, Timpel P, Schönfelder T, Militzer-Horstmann C, Scheibe S, Heinrich R, et al. Evidence requirements of permanently listed digital health applications (DiGA) and their implementation in the German DiGA directory: an analysis. BMC Health Serv Res 2023 Apr 17;23(1):369 [FREE Full text] [doi: 10.1186/s12913-023-09287-w] [Medline: 37069592]
55. Lantzsch H, Eckhardt H, Campione A, Busse R, Henschke C. Digital health applications and the fast-track pathway to public health coverage in Germany: challenges and opportunities based on first results. BMC Health Serv Res 2022 Sep 21;22(1):1182 [FREE Full text] [doi: 10.1186/s12913-022-08500-6] [Medline: 36131288]
56. Schlieter H, Kählig M, Hickmann E, Fürstenau D, Sunyaev A, Richter P, et al. Digital health applications (DiGA) in the area of tension between progress and criticism : discussion paper from the "digital health" specialist group of the German Informatics Society. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2024 Jan;67(1):107-114 [FREE Full text] [doi: 10.1007/s00103-023-03804-2] [Medline: 38086924]
57. Stern AD. Innovation under regulatory uncertainty: evidence from medical technology. J Public Econ 2017 Jan;145:181-200 [FREE Full text] [doi: 10.1016/j.jpubeco.2016.11.010] [Medline: 28652646]
58. Omachonu VK, Einspruch NG. Innovation in healthcare delivery systems: a conceptual framework. Innov J 2010;15(1).
59. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
60. Torous J, Stern AD, Bourgeois FT. Regulatory considerations to keep pace with innovation in digital health products. NPJ Digit Med 2022 Aug 19;5(1):121 [FREE Full text] [doi: 10.1038/s41746-022-00668-9] [Medline: 35986056]
61. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-a call for innovative evidence generation approaches. NPJ Digit Med 2020 Aug 27;3(1):110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
62. 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]
63. Bertha A, Alaj R, Bousnina I, Doyle MK, Friend D, Kalamegham R, et al. Incorporating digitally derived endpoints within clinical development programs by leveraging prior work. NPJ Digit Med 2023 Aug 10;6(1):139 [FREE Full text] [doi: 10.1038/s41746-023-00886-9] [Medline: 37563201]
64. Giunti G, Guisado-Fernandez E, Belani H, Lacalle-Remigio JR. Mapping the access of future doctors to health information technologies training in the European Union: cross-sectional descriptive study. J Med Internet Res 2019 Aug 12;21(8):e14086 [FREE Full text] [doi: 10.2196/14086] [Medline: 31407668]
65. Wilts H, Bringezu S, Bleischwitz R, Lucas R, Wittmer D. Challenges of metal recycling and an international covenant as possible instrument of a globally extended producer responsibility. Waste Manage Res 2011 Jul 19;29(9):902-910. [doi: 10.1177/0734242x11415311]
66. Cohen AB, Mathews SC, Dorsey ER, Bates DW, Safavi K. Direct-to-consumer digital health. Lancet Digit Health 2020 Apr;2(4):e163-e165 [FREE Full text] [doi: 10.1016/S2589-7500(20)30057-1] [Medline: 33328077]
67. Facey K, Lo Scalzo A, Ørtenblad L. EUnetHTA HTA core model: patient and social aspects. EUnetHTA HTA Core Model v3.0. Maastricht, The Netherlands: European Union; 2016. URL: https://www.eunethta.eu/hta-core-model-3-0/ [accessed 2024-03-01]
68. Stern AD, Brönneke J, Debatin JF, Hagen J, Matthies H, Patel S, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health 2022 Mar;4(3):e200-e206 [FREE Full text] [doi: 10.1016/S2589-7500(21)00292-2] [Medline: 35216754]
69. Brückner S, Brightwell C, Gilbert S. FDA launches health care at home initiative to drive equity in digital medical care. NPJ Digit Med 2024 Aug 21;7(1):204 [FREE Full text] [doi: 10.1038/s41746-024-01198-2] [Medline: 39169161]

70. Donabedian A. The quality of care: how can it be assessed? JAMA 1988 Sep 23;260(12):1743-1748 [FREE Full text] [doi: 10.1001/jama.1988.03410120089033]
71. Hamburg MA. Innovation, regulation, and the FDA. New Engl J Med 2010 Dec 02;363(23):2228-2232. [doi: 10.1056/nejmsa1007467]
72. Diao JA, Venkatesh KP, Raza MM, Kvedar JC. Multinational landscape of health app policy: toward regulatory consensus on digital health. NPJ Digit Med 2022 May 11;5(1):61 [FREE Full text] [doi: 10.1038/s41746-022-00604-x] [Medline: 35545663]
73. 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]
74. Brönneke JB, Herr A, Reif S, Stern AD. Dynamic HTA for digital health solutions: opportunities and challenges for patient-centered evaluation. Int J Technol Assess Health Care 2023 Nov 17;39(1):e72. [doi: 10.1017/s0266462323002726]
75. Masanneck L, Gieseler P, Gordon WJ, Meuth SG, Stern AD. Evidence from ClinicalTrials.gov on the growth of Digital Health Technologies in neurology trials. NPJ Digit Med 2023 Feb 10;6(1):23 [FREE Full text] [doi: 10.1038/s41746-023-00767-1] [Medline: 36765123]
76. Gordon WJ, Coravos AR, Stern AD. Ushering in safe, effective, secure, and ethical medicine in the digital era. NPJ Digit Med 2021 Mar 25;4(1):56 [FREE Full text] [doi: 10.1038/s41746-021-00424-5] [Medline: 33767377]
77. Rouleau G, Wu K, Ramamoorthi K, Boxall C, Liu RH, Maloney S, et al. Mapping theories, models, and frameworks to evaluate digital health interventions: scoping review. J Med Internet Res 2024 Feb 05;26:e51098 [FREE Full text] [doi: 10.2196/51098] [Medline: 38315515]
78. Kvedar JC, Mittermaier M, Pritzker J. The industry impact of the American Medical Association's Digital Medicine Payment Advisory Group (DMPAG). NPJ Digit Med 2022 Dec 24;5(1):193 [FREE Full text] [doi: 10.1038/s41746-022-00743-1] [Medline: 36566288]
79. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. Opinion paper: “so what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manage 2023 Aug;71:102642. [doi: 10.1016/j.ijinfomgt.2023.102642]

Abbreviations

BfArM: Federal Institute for Drugs and Medical Devices (German: Bundesinstitut für Arzneimittel und Medizinprodukte)
CE: Conformité Européenne
COREQ-32: Consolidated Criteria for Reporting Qualitative Research-32
DiGA: digital health application (German: Digitale Gesundheitsanwendung)
DiGAV: digital health application ordinance (German: Digitiale Gesundheitsanwendungen-Verordnung)
DiPA: digital nursing care application
DVG: Digital Healthcare Act
PECAN: Prise en Charge Anticipée Numérique
pSVV: patient-relevant structural and procedural improvement (German: Patientenrelevante Struktur- und Verfahrensverbesserung)

Edited by A Mavragani; submitted 10.09.24; peer-reviewed by S Greene, K Mouloudj; comments to author 04.11.24; revised version received 08.11.24; accepted 13.11.24; published 06.01.25.

Please cite as:
Gehder S, Goeldner M
Unpacking Performance Factors of Innovation Systems and Studying Germany’s Attempt to Foster the Role of the Patient Through a Market Access Pathway for Digital Health Applications (DiGAs): Exploratory Mixed Methods Study
J Med Internet Res 2025;27:e66356
URL: https://www.jmir.org/2025/1/e66356
doi:10.2196/66356
PMID:

Viewpoint

Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine

Kuo Zhang[1*], MD, PhD; Xiangbin Meng[2*], MD, PhD; Xiangyu Yan[3*], PhD; Jiaming Ji[4], PhD; Jingqian Liu[5], PhD; Hua Xu[6], PhD; Heng Zhang[7], MSc; Da Liu[8], MD, PhD; Jingjia Wang[9], MD, PhD; Xuliang Wang[9], MD, PhD; Jun Gao[9], MD, PhD; Yuan-geng-shuo Wang[9], MD, PhD; Chunli Shao[9], MD, PhD; Wenyao Wang[9], MD, PhD; Jiarong Li[10], PhD; Ming-Qi Zheng[8*], MD, PhD; Yaodong Yang[4*], PhD; Yi-Da Tang[9], MD, PhD

[1]Department of Cardiology, State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[2]Pengcheng Laboratory, Shenzhen, Guangdong, China

[3]School of Disaster and Emergency Medicine, Tianjin University, Tianjin, China

[4]Institute for Artificial Intelligence, Peking University, Beijing, China

[5]China Telecom, Beijing, China

[6]Division of Emerging Interdisciplinary Areas, Hong Kong University of Science and Technology, Hong Kong, China (Hong Kong)

[7]Institute for Artificial Intelligence, Hefei University of Technology, Hefei, Anhui, China

[8]Department of Cardiology, the First Hospital of Hebei Medical University, Graduate School of Hebei Medical University, Shijiazhuang, Hebei, China

[9]Department of Cardiology and Institute of Vascular Medicine, Key Laboratory of Molecular Cardiovascular Science, Ministry of Education, Peking University Third Hospital, Beijing, China

[10]Henley Business School, University of Reading, RG6 6UD, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Yi-Da Tang, MD, PhD
Department of Cardiology and Institute of Vascular Medicine
Key Laboratory of Molecular Cardiovascular Science, Ministry of Education
Peking University Third Hospital
49 North Garden Road
Beijing, 100191
China
Phone: 86 88396171
Email: tangyida@bjmu.edu.cn

Abstract

Large language models (LLMs) are rapidly advancing medical artificial intelligence, offering revolutionary changes in health care. These models excel in natural language processing (NLP), enhancing clinical support, diagnosis, treatment, and medical research. Breakthroughs, like GPT-4 and BERT (Bidirectional Encoder Representations from Transformer), demonstrate LLMs' evolution through improved computing power and data. However, their high hardware requirements are being addressed through technological advancements. LLMs are unique in processing multimodal data, thereby improving emergency, elder care, and digital medical procedures. Challenges include ensuring their empirical reliability, addressing ethical and societal implications, especially data privacy, and mitigating biases while maintaining privacy and accountability. The paper emphasizes the need for human-centric, bias-free LLMs for personalized medicine and advocates for equitable development and access. LLMs hold promise for transformative impacts in health care.

(J Med Internet Res 2025;27:e59069) doi:10.2196/59069

KEYWORDS

large language models; LLMs; digital health; medical diagnosis; treatment; multimodal data integration; technological fairness; artificial intelligence; AI; natural language processing; NLP

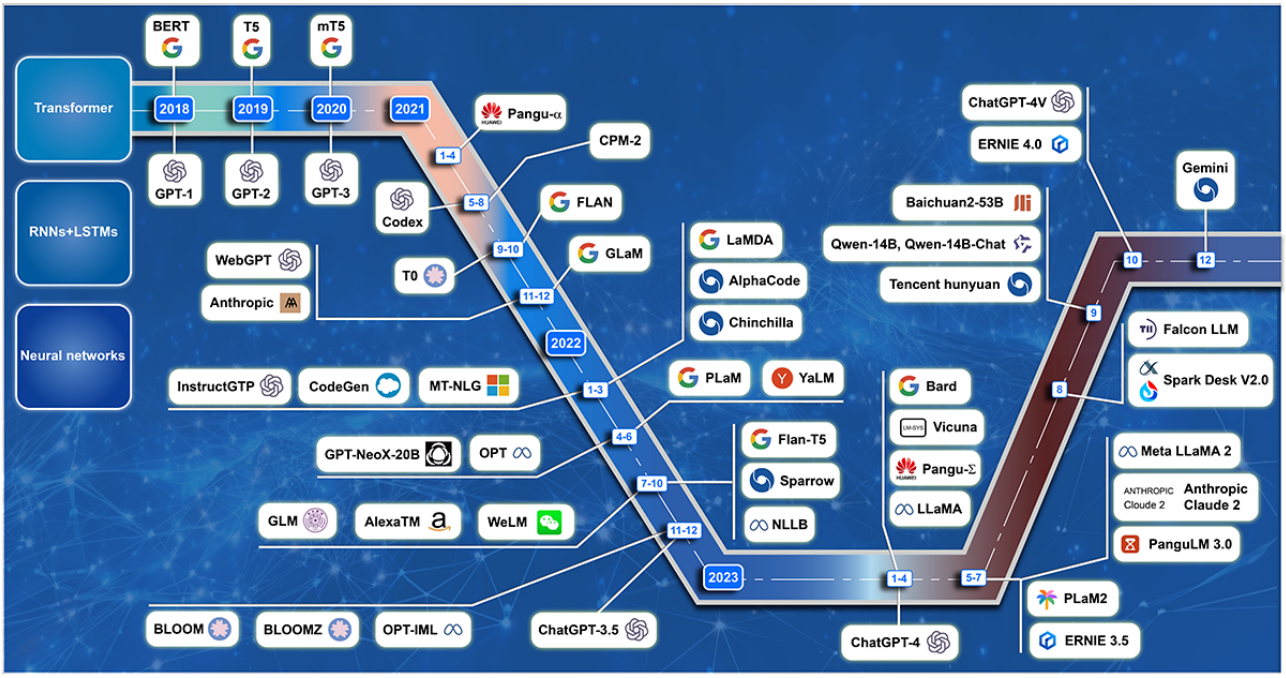
Introduction

Recent advancements in artificial intelligence (AI) have catalyzed the development and significant breakthroughs of large language models (LLMs), placing them at the forefront of AI research [1-4]. LLMs are deep learning models that generate human-like text by predicting the next word in a sequence based on statistical patterns learned from vast text data. These models leverage deep learning algorithms to interpret and generate natural language, using extensive corpus data to enhance pretrained language models, a cornerstone of natural language processing (NLP) [5,6]. Characterized by their immense scale, these models often consist of hundreds of millions to billions of parameters and are trained on vast textual datasets [7,8]. Their ability to efficiently process natural language data with minimal human intervention, capturing intricate grammatical structures, lexical nuances, and semantic contexts, is noteworthy. Globally recognized LLMs include the ChatGPT series, BERT (Bidirectional Encoder Representations from Transformer), PaLM, LaMDA, and Meta’s Llama series, with China contributing models such as Baidu’s “Wenxin Yiyan,” 360’s LLM, Alibaba’s “Tongyi Qianwen,” and SenseTime’s LLM [9]. The evolution of LLMs represents over 7 years of relentless technological innovation and research, marking a significant milestone in AI development since the inception of the Turing machine.

LLMs primarily function to comprehend, generate, and interact through language. In NLP tasks, such as text classification, named entity recognition, and sentiment analysis, their proficiency is unparalleled [10-12]. Beyond these applications, LLMs are expanding their influence. In mathematics, they assist in solving complex problems and contributing to mathematical proofs [13]. In software development, their capabilities include automatic code generation, debugging assistance, and complex algorithm explanation [14]. Intriguingly, LLMs are venturing into artistic creation, exhibiting talent in generating poetry, stories, and music [15,16].

In the medical domain, LLMs are poised to revolutionize clinical decision support. They can assist health care professionals in diagnosing diseases with enhanced accuracy and speed, provide treatment recommendations, and facilitate the analysis of medical records by processing large volumes of medical data [17-20]. They are instrumental in swiftly navigating vast medical literature, providing health care professionals with essential research, guidelines, and information, thus saving time and grounding medical treatments in current knowledge [21-25]. Additionally, LLMs can interact directly with patients, offer medical consultations, and handle document processing efficiently [26-28]. For example, health care professionals use LLMs to assist in diagnosing diseases by quickly processing and interpreting large volumes of patient data such as electronic health records and imaging results. Clinicians also leverage LLMs for treatment planning, where the models suggest potential treatment options based on the latest medical guidelines and patient-specific data. Moreover, LLMs are used in streamlining administrative tasks, such as generating and managing medical documentation, allowing clinicians to spend more time with their patients. Their role in drug research and development is also emerging, aiding in new drug discoveries through detailed analysis of chemical and biological data [29,30]. As such, LLMs are reshaping research methodologies and applications across various fields, particularly in medicine, equipping doctors with advanced tools for more accurate and efficient diagnosis and treatment, while offering patients more convenient and effective medical services. The potential for broader applications of LLMs in the medical field is vast, and there is a strong rationale to expect their significant impact on future health care advancements (Figure 1).

Figure 1. Timeline of mainstream LLMs commercially available to the public. The technological evolution of LLMs, highlighting several key technologies and models. It includes RNNs and LSTMs from the 1990s, Google's Transformer model introduced in 2017, Google's BERT model released in 2018, and the GPT series by OpenAI. Specific emphasis is placed on three major milestones: the first open-source LLM—GPT-2, and the first widely acclaimed LLM—GPT-3. These developments signify major advancements in LLMs within the field of natural language processing. BERT: Bidirectional Encoder Representations from Transformers; LLM: large language model; LSTM: long short-term memory network; RNN: recurrent neural network.



LLM Technical Background and Hardware Infrastructure

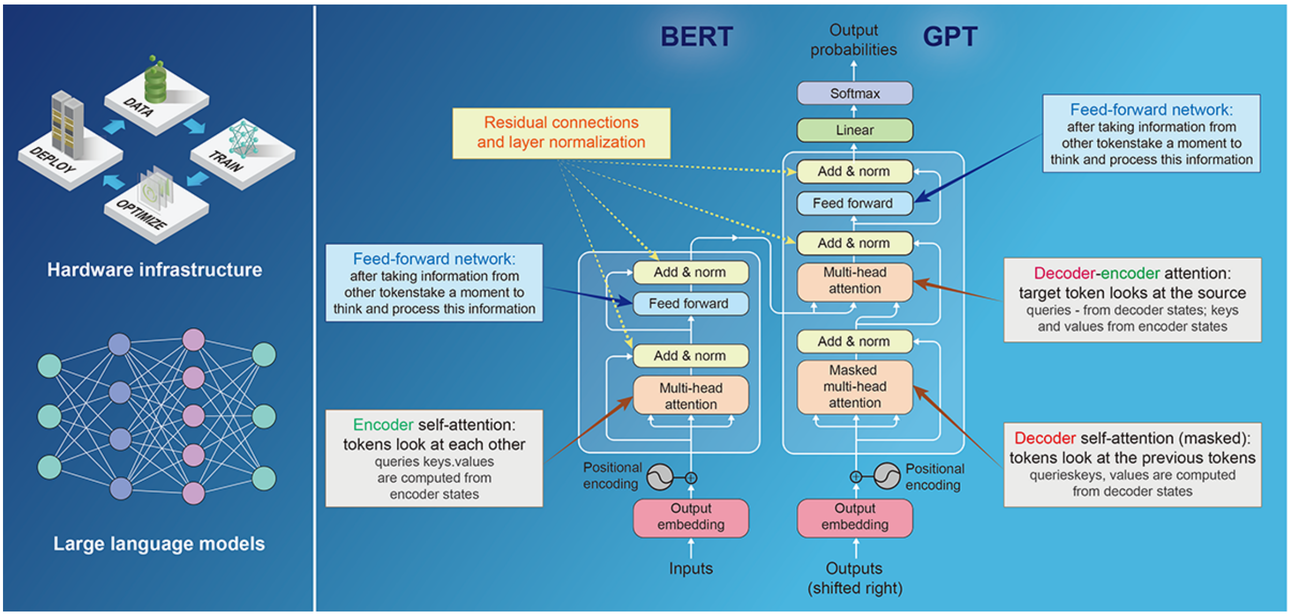
The evolution of LLMs, like OpenAI's GPT-3 and Google's BERT, has been monumental, driven by advancements in AI chip computing power and large, high-quality datasets [31]. The Transformer model, introduced by Google in 2016, underpins this progress, predicting words in sentences based on statistical correlations [32,33]. Notably, GPT-3 in 2020 showcased the significance of model size and data quality.

The operation and training of LLMs, such as ChatGPT, require substantial hardware infrastructure [34]. This includes graphics processing units (GPUs) or tensor processing units (TPUs) with thousands of cores, extensive RAM (several terabytes), over 48 GB of VRAM on GPUs, high-performance solid-state drives, and fast, low-latency networks (10 to 100 Gbps) [35,36]. Effective cooling systems and reliable power supplies are also essential. Compatibility with software frameworks, like TensorFlow and PyTorch, is necessary for optimizing training and deployment. The training of GPT-3, for instance, costs around US $1.4 million, and operational costs for models, like ChatGPT, can reach up to US $700,000 daily, with significant energy consumption.

Future technology advancements are expected to reduce the costs and improve the efficiency of LLMs. Progress in GPU and TPU technologies, along with hardware tailored for LLM training, will drive efficiency. Compact model structures through knowledge distillation, model pruning, transfer learning, energy-efficient practices, distributed training, and edge computing are anticipated. Semisupervised and self-supervised learning methods will also play a role in training models with fewer labeled datasets [37,38]. ChatGPT's recent updates showcase improvements in response speed, handling complex queries, multimodal functionality, global language support, and enhanced privacy and security measures [39].

In health care, deploying large-scale medical models faces unique challenges due to data security and privacy concerns. Hospitals typically have CPUs for general computing, with limited access to GPUs. Medical LLMs, generally smaller than general-purpose LLMs, still require substantial investment in operational hardware [40,41]. For instance, a model with 13 billion parameters might cost under US $138,000 while larger models for entire hospitals may require advanced GPU solutions costing around RMB 10 million. Effective deployment demands careful consideration of model scale, computational resources, data security, and cost control (Figure 2).

Figure 2. The architectural designs of LLMs: a study of self-attention mechanisms and structural variations. The image depicts the hardware infrastructure for LLMs and their implementation in the BERT and GPT models. On the left, there is a network diagram showing servers and computing devices needed to run these models, labeled with hardware such as TPU and GPU. On the right, the structure of BERT and GPT is compared in detail, including positional encoding, self-attention mechanisms, feed-forward networks, addition and normalization layers, and the computation of output probabilities. Although these models have different approaches to processing text, both are large neural network models based on deep learning and self-attention mechanisms. BERT: Bidirectional Encoder Representations from Transformers. GPU: graphics processing unit; LLM: large language model. TPU: tensor processing unit.



Advancing the Integration of LLMs in Health Care: The Imperative for Evidence-Based Research and Collaborative Evaluation

Overview

In the contemporary health care landscape, the paradigm of evidence-based medicine is instrumental in shaping medical decision-making processes. This methodology integrates top-tier research evidence with clinical expertise and aligns it with patient values and expectations, thereby informing patient care decisions. Evidence-based medicine ensures that medical interventions are grounded in scientific evidence rather than solely relying on a physician's experience or intuition, enhancing patient safety and the efficacy of treatments [42-45].

The integration of LLMs into the medical field introduces a significant challenge: the current scarcity of evidence-based medical research concerning the application of LLMs in health care settings [46]. Although LLMs have shown remarkable efficacy in various sectors, the unique context of medicine, with its direct implications for human life and health, necessitates a cautious approach to the introduction of untested technologies or methods into clinical practice [47]. Despite their robust data processing capabilities, LLMs present a potential risk for prediction errors in clinical environments. The medical domain, with its complex interplay of biology, physiology, and pathology, might be challenging for machine learning models to fully encapsulate, especially considering the intricacies and variability inherent in medical data [48]. Furthermore, the realm of medical decision-making often requires a high level of expertise and experience, aspects that may not be entirely replicable by LLMs. The consequences of medical decisions far surpass those in other sectors, where a misdiagnosis or incorrect treatment recommendation could directly jeopardize a patient's life. Hence, it is imperative to back any new technological innovation, including LLMs, with solid scientific evidence before they are implemented in medical practice.

Currently, empirical studies examining the application of LLMs in the medical field are limited. This scarcity of research implies an inability to definitively assess the accuracy, reliability, and safety of LLMs within a health care context. Model reliability refers to the consistency and dependability of a model's outputs across different datasets or under varying conditions. In medical applications, the reliability of LLMs is critical, as it directly affects the accuracy of diagnoses and treatment recommendations, where any inconsistency could have serious consequences for patient care. To comprehensively understand the potential benefits and risks associated with LLMs in medicine, a more robust body of clinical research is required. This research should encompass randomized controlled trials, observational studies, and extensive collaborative research, which are critical to evaluating the clinical utility of LLMs accurately [49].

To accelerate the empirical evaluation of LLMs in the medical field, fostering collaboration between medical institutions, research organizations, and technology companies is essential. This interdisciplinary collaboration ensures the comprehensiveness and quality of the research, facilitating the rapid advancement and application of LLM technologies. To enhance the transparency, trustworthiness, and ethical application of LLMs in health care, it is crucial to address the societal implications, particularly in terms of data privacy. Publicizing research findings and fostering interdisciplinary collaboration among doctors, researchers, and ethicists will be key to ensuring that LLMs are used responsibly and equitably. Furthermore, the integration of robust data privacy measures and adherence to ethical standards must be a priority to prevent potential misuse or unintended consequences that could undermine public trust. Such an approach ensures that LLMs' application in the medical field is underpinned by scientific rigor, is safe, and genuinely benefits both patients and the health care system.

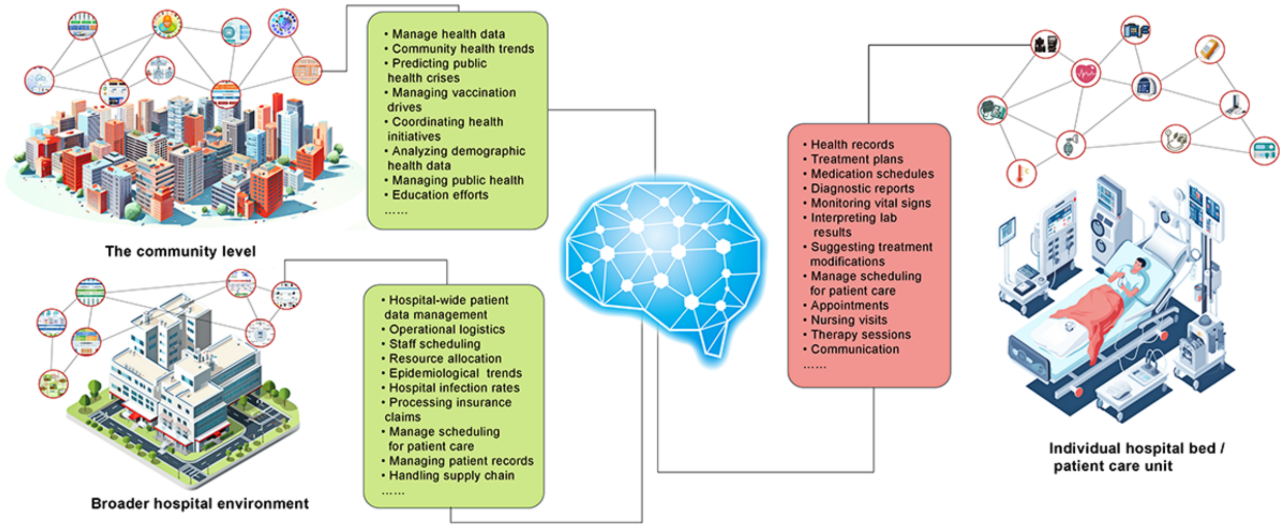
Integrated Application of LLMs in Medical System

As we witness ongoing advancements in medical technology, the integration of LLMs with other tools and platforms within health care systems becomes increasingly crucial [50]. This fusion provides health care professionals with powerful tools to process, analyze, and effectively use vast amounts of health care data [23,51-54]. The integration of LLMs, such as ChatGPT, into medical systems has the potential to drive transformative progress in health care delivery. First, LLMs can potentially enhance diagnostic accuracy and clinical decision-making by analyzing comprehensive medical data to identify relevant information and suggest potential diagnoses based on presented symptoms [55-57]. Second, their proficiency in text processing and generation assists medical professionals in efficiently summarizing medical literature, facilitating research, and improving communication between health care providers and patients [58-61]. The rapid adoption of readily available LLMs, such as ChatGPT, within the medical community, signifies recognition of their potential to transform health care delivery [62-66].

However, the application of LLMs in clinical settings is not without challenges [67]. A primary concern is the generalizability of these models. Although LLMs have shown outstanding performance in numerous standard tasks, the complexity and diversity of the medical field suggest that these models may be susceptible to prediction errors in real clinical scenarios. Such errors can have serious implications, particularly when they influence critical health and life decisions. Additionally, the medical field encompasses a vast array of domain-specific knowledge that might exceed the training scope of LLMs, potentially leading to misunderstandings in complex medical scenarios.

Despite these challenges, the potential benefits and impact of LLMs in health care are considerable. LLMs can notably enhance the efficiency of medical workflows by automating routine processes such as appointment scheduling, diagnosis, and report generation [68]. Their data-driven recommendations provide powerful decision support to doctors, assisting them in making more accurate and timely decisions. Current digital health workflows often burden physicians with extensive data entry, querying, and management tasks, leading to information overload and fatigue. LLMs can alleviate these burdens by automating these tasks, thereby saving valuable time for health care providers. Moreover, by analyzing and integrating patients' medical data, LLMs can offer tailored diagnoses and treatment recommendations, improving the overall quality of health care delivery. LLMs also play a crucial role in enhancing doctor-patient interactions. Leveraging NLP technology, they can better comprehend patients' needs and concerns, offering more personalized medical advice [69]. This not only boosts patient satisfaction but also enhances the overall effectiveness of medical services. The potential of LLMs to optimize digital health care workflows is undeniable. With further technological advancements and empirical research, LLMs are expected to play an increasingly significant role in the future of health care (Figure 3).

Figure 3. Integration of LLMs in health care systems across different scales. LLMs can assist in monitoring and analyzing patient health records, treatment plans, and laboratory results at the individual bed level while managing care schedules and facilitating doctor-patient communication. At the hospital level, LLM helps manage patient data, operational logistics, staff scheduling, and resource allocation, while analyzing epidemic trends and hospital infection rates. At the community level, LLM can be used to predict public health crises, manage vaccination campaigns, coordinate community health initiatives, and analyze population health data to improve health policy. LLM: large language model.



Multimodal LLMs in Real-World Medical Scenarios

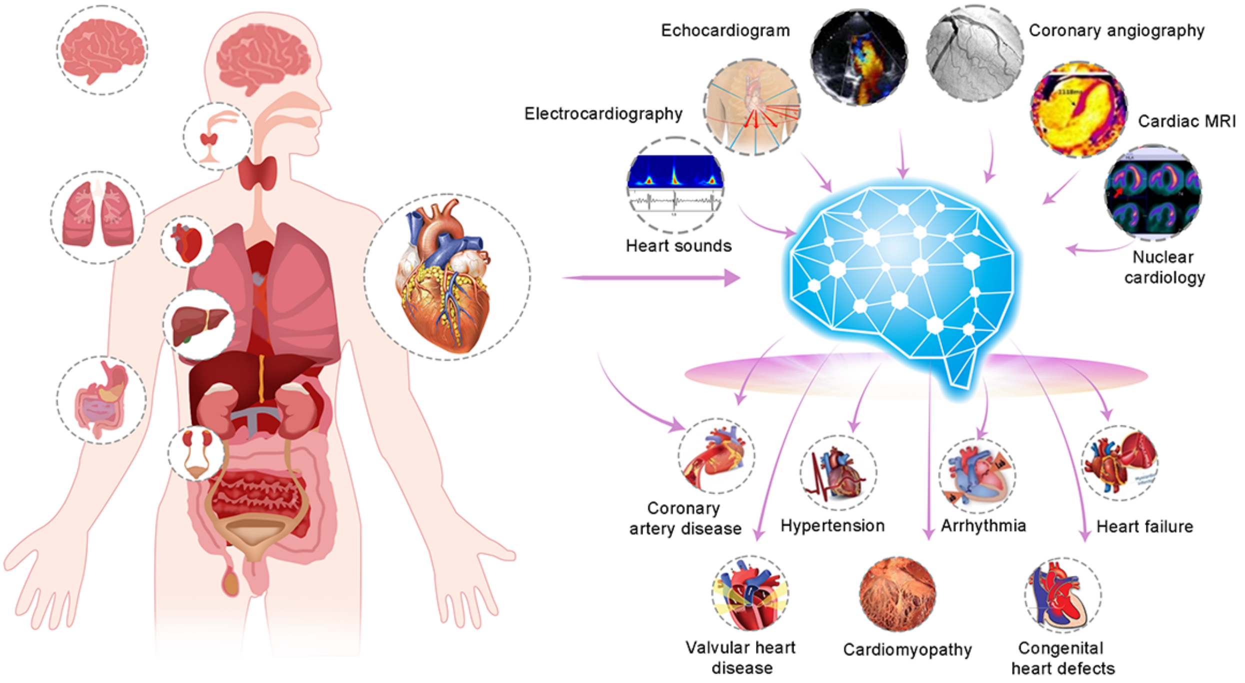
The advent of multimodal LLMs is bringing about a paradigm shift in the medical field by offering the capability to process and generate diverse data types such as text, images, sounds, and videos. This integration of multiple data types enables LLMs to provide more comprehensive and accurate predictions, thereby unlocking unprecedented potential [70-73]. To understand their role, it is essential to define what multimodal LLMs entail. Multimodal LLMs excel in processing, interpreting, and generating a wide array of data types, which significantly enhances their predictive capabilities. For instance, in the medical field, combining textual data from patient records with imaging data from magnetic resonance imaging (MRI), computed tomography scans, and x-rays allows these models to provide more nuanced and precise diagnoses. Additionally, integrating audio data from patient interviews or video data from medical procedures can further enrich the model's understanding, leading to more accurate and personalized treatment recommendations. By leveraging the strengths of various data types, multimodal LLMs can offer a holistic view of a patient's condition, which is often crucial for complex diagnoses and treatment planning.

The utility of LLMs is increasingly becoming a focal point in medical imaging [74-76]. For instance, when a patient undergoes an MRI or computed tomography scan, an LLM can swiftly analyze and integrate the image data with the patient's textual medical records, thereby providing more comprehensive and detailed diagnostic insights. Additionally, LLMs have the capability to automatically identify and highlight crucial areas in medical images, thus providing clinicians with clear references that aid in identifying potential issues [77]. Moreover, LLMs can generate automated image reports, offering initial interpretations and treatment suggestions based on the analyzed image data, significantly boosting the efficiency and accuracy of medical diagnoses and treatments.

Multimodal LLMs are revolutionizing the field of telemedicine, transforming the dynamics of doctor-patient interactions [55,78]. For instance, LLMs have been successfully integrated into MRI analysis, where they can rapidly interpret imaging data and provide diagnostic recommendations. This has significantly reduced the time required for diagnosis and improved accuracy. However, the use of LLMs is not without its challenges. A notable example is Google BARD, which recently demonstrated racial bias in patient diagnosis, disproportionately affecting minority groups. This case highlights the dual-edged nature of LLMs in health care—they offer substantial benefits in efficiency and accuracy, yet they also pose significant risks if not properly validated and monitored for biases. Furthermore, the integration of LLMs with smart sensors and devices enables the continuous monitoring of patient's physiological data, such as heart rate and blood pressure, facilitating early detection and intervention for any health anomalies, thus significantly bolstering patient health management.

In summary, multimodal LLMs offer a novel and efficacious approach to diagnosis, treatment, and health care management. Their robust capabilities in data processing and integration allow medical professionals to deliver more precise and efficient services to patients. At the same time, these models enable patients to access medical advice and care with greater convenience. As these technologies continue to evolve and improve, their significance and impact in the medical field are expected to grow exponentially (Figure 4).

Figure 4. The importance of multimodal large language models in medical applications. The central heart represents the cardiac health status of the human body. The surrounding circular icons depict various cardiac conditions including coronary artery disease, hypertension, arrhythmia, heart failure, valvular heart disease, cardiomyopathy, and congenital heart defects. These conditions are detected and analyzed through different medical imaging and diagnostic technologies such as electrocardiography, heart sounds, echocardiogram, coronary angiography, cardiac MRI, and nuclear cardiology. The results from these diagnostics are processed by an AI system to determine the type and severity of cardiac disease, assisting physicians in formulating treatment plans. MRI: magnetic resonance imaging.



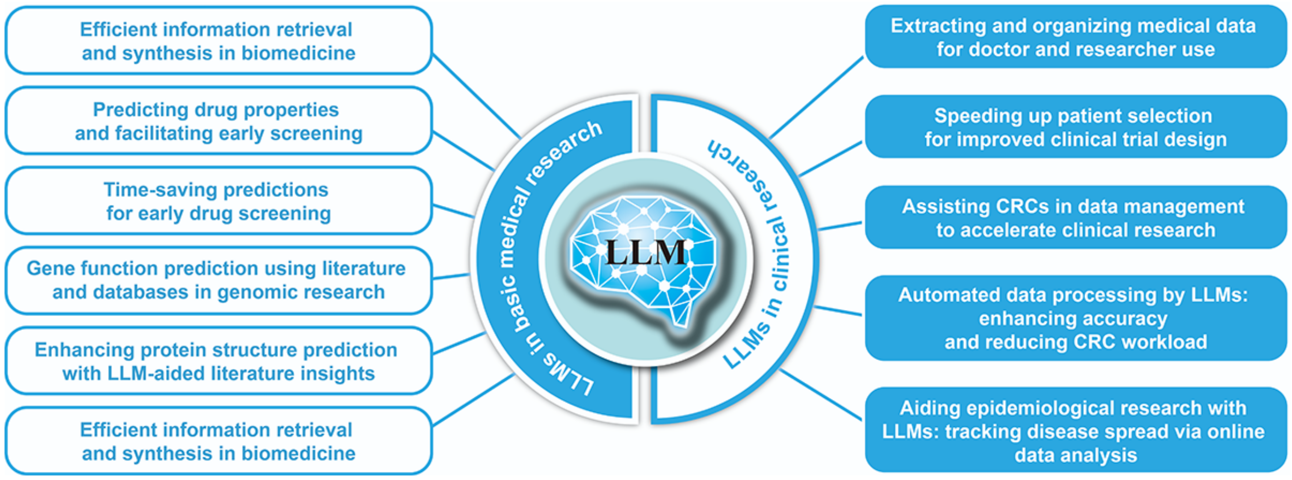
The Key Role of LLMs in Medical Research

In the field of fundamental medical research, the capabilities of LLMs in AI are being increasingly recognized [79-82]. LLMs can swiftly retrieve and organize crucial information from vast biomedical literature, providing researchers with an efficient tool to access and synthesize the latest research findings on specific drugs, diseases, or genes [83]. In drug discovery, LLMs can predict the activity, toxicity, and pharmacokinetic properties of new compounds, facilitating early-stage drug screening [84]. These predictions not only save time but also facilitate the early-stage screening of potential drug molecules. LLMs can use existing literature and databases to predict the potential functions of newly discovered genes, a crucial aspect of genomic research, given the daily discovery and study of new genes. While protein structure prediction depends primarily on specialized models, such as AlphaFold, LLMs can enhance these models by supplying pertinent information from literature, thereby increasing prediction accuracy. In epidemiological research, LLMs can aid researchers in tracking and predicting disease spread by analyzing social media and other web-based text data, offering data support for public health decision-making. Finally, in bioinformatics applications, LLMs can assist researchers in predicting patterns, functional domains, and similarities to known biological sequences. Despite their extensive applications in biomedicine, LLMs cannot entirely replace laboratory experiments or in-depth biomedical expertise. Instead, they should be considered powerful supplementary tools, rather than replacements.

LLMs play a pivotal role in clinical research. They aid doctors and researchers by extracting essential information from medical records, and by organizing and categorizing data for easier analysis and application. For instance, they can expedite the selection of suitable patients for enrolment, thereby enhancing the design and implementation of clinical trials. In the role of a clinical research coordinator, these models assist with data entry, verification, and analysis, thereby accelerating the clinical research process. Through automated data processing and real-time analysis, LLMs can ensure data accuracy and completeness, while also reducing the workload of clinical research coordinators. This, in turn, speeds up the clinical research process and enhances research quality.

Although LLMs have revolutionized biomedicine by simplifying literature searches, aiding drug discovery, annotating gene functions, and supporting epidemiological studies, they experience certain drawbacks. Their ability to swiftly parse large datasets and make predictions may be counterbalanced by potential limitations in real-world validation [85]. For example, while they can predict a drug molecule’s properties, the actual biological response may vary. Similarly, despite gene function predictions being well grounded, they may not fully encapsulate the breadth of gene interactions. Moreover, using LLMs to analyze epidemiological trends without correlating them to underlying data could misdirect public health interventions. Therefore, while LLMs are undeniably beneficial to biomedicine, it is essential to adopt a balanced approach, combining their computational prowess with rigorous experimental validation and expert review, to fully harness their potential without sacrificing scientific rigor (Figure 5).

Figure 5. The crucial role of LLMs in medical science: bridging basic research and clinical trials. This illustration highlights the versatile roles of LLMs in medical research. LLMs analyze medical texts to uncover trends and inform research directions, facilitate hypothesis generation, and enhance clinical trial designs. They personalize medicine through data-driven treatment plans and use predictive modeling to inform clinical trial outcomes. LLMs also streamline research by integrating data and maintaining regulatory compliance. They assist in medical communication and education and evaluate the societal impact of clinical research. LLM: large language model.



Great Challenges of LLMs in Medical Scenarios and Feasible Roadmap

The integration of technology in health care invariably brings a mix of anticipation and challenges, particularly given its direct impact on human life and health. As a leading exemplar of current AI technology, LLMs present a complex array of opportunities and challenges in the medical field, warranting thorough exploration and discussion [86-88].

Handling medical data, some of the most private and sensitive information about individuals, is a significant challenge for LLMs. As LLMs are increasingly integrated into health care, ethical considerations surrounding data privacy and societal impact must be prioritized during their development and deployment. The key lies not only in using this data to enhance medical efficiency but also in implementing robust data protection frameworks to prevent misuse, leakage, and unauthorized access. Furthermore, addressing these ethical challenges requires ongoing dialogue among technologists, health care professionals, policy makers, and the public to ensure that LLM deployment aligns with societal values and legal standards [80,89]. A potential technical solution involves anonymizing patient data, ensuring that neither processing nor transmission stages can be linked to specific individuals. Concurrently, medical organizations and technology providers must establish robust data management and access protocols, ensuring clear authorization and purpose for each data access.

Interpretive challenges loom large with LLMs in medicine. Medical decision-making is distinct from other fields due to its complexity and direct implications for patients' lives and health. When LLMs provide diagnostic or treatment suggestions, it is vital that the rationale behind these recommendations is transparent and comprehensible [90-92]. This brings us to the concept of interpretability in machine learning, which refers to the ability to understand and explain how a model makes its decisions. In the context of health care, interpretability is a significant challenge because clinicians must trust and validate the outputs of LLMs, especially when these models influence critical medical decisions [93]. Developing mental models can aid LLMs in presenting their decision-making logic in a manner that is more accessible to human users. Leveraging deep learning and other machine learning technologies, LLMs can extract disease pathophysiological mechanisms from a vast corpus of medical literature and data, providing a scientific basis for their outputs. To further enhance interpretability, LLMs could use visual tools, like graphics and animations, to clarify the logic and evidence underpinning their decisions for both physicians and patients [94,95].

The issue of technical bias and the possibility of generating misleading information or "hallucinations" are inherent challenges in LLMs. In this context, hallucinations refer to instances where LLMs produce outputs that are factually incorrect or misleading, often because the model attempts to generate an answer despite lacking sufficient context or knowledge. These hallucinations can be especially problematic in medical scenarios, where inaccurate information can have severe consequences. The data sources for these models, often anonymized consultation data and digital materials, are not uniform and vary in quality, sometimes containing erroneous samples. Fine-tuning LLMs based on such data may lead to biased or skewed medical recommendations [96,97]. Addressing this requires rigorous data auditing and the establishment of continuous bias-correction mechanisms. To mitigate the risk of hallucinations, knowledge enhancement methods, such as integrating a knowledge retrieval library or search enhancement tools, can be beneficial. The LLM's responses can be cross-referenced with retrieved data to filter out inconsistencies with reality. Another approach involves reinforcement learning based on human feedback, where high-quality feedback is provided to fine-tune and correct model outputs in collaboration with medical experts [98,99].

The potential of AI to create "information cocoons" through personalized content, potentially reinforcing biases, is another critical aspect that needs to be addressed, especially in the medical domain [100]. AI technologies, including LLMs, in medicine require stringent scrutiny and continuous evaluation to align with the field's unique characteristics and ethical standards. Ensuring privacy protection, eliminating biases and discrimination, and establishing clear accountability are essential. The use of LLMs should be guided by respect for life, aiming to enhance patient well-being and treatment outcomes, without compromising individual interests. A continuous monitoring and evaluation system is crucial for assessing the effectiveness of LLMs and managing potential risks. Regulations should be regularly updated to keep pace with AI advancements, ensuring medical safety and patient rights. By prioritizing safety, fairness, and effectiveness, we can fully leverage LLMs and other AI technologies to facilitate a transformative revolution in medicine, while upholding human values and rights.

In the era of information and intelligence within the medical field, the application of LLMs harbors immense potential [101]. However, the accompanying challenges are equally noteworthy and merit careful consideration. The ongoing discourse should emphasize not only the deeper integration of LLMs into medical practice but also their alignment with both the professional needs of health care providers and the experiential needs of patients [102,103].

Incorporating the theory of mind into LLMs can significantly enhance their utility in the medical field. This concept, which involves understanding others' thoughts, feelings, and intentions, is crucial for fostering trust and empathy within health care interactions. Medicine is not solely a science; it is also an art, deeply influenced by each patient's unique emotional, value-based, and experiential landscape. An AI system endowed with the capability to appreciate and respond to these individual differences can offer more personalized and compassionate medical advice [104,105]. By using the theory of mind, LLMs can gain deeper insights into patients' inherent needs and respond with more attentive and empathetic advice [106-108]. When LLMs can emulate the thoughts and feelings of both doctors and patients, their outputs transcend mere data; they become imbued with empathy and human care, enhancing the patient's treatment experience and fostering stronger trust and communication between doctors and patients. For example, in interactions with terminal patients, LLMs could suggest more compassionate communication strategies, aiding both doctors and patients in navigating these sensitive and complex situations.

LLMs can be synergistically combined with other advanced technologies, such as virtual reality and augmented reality, to transform medical consultations into more immersive and informative experiences. This integration can provide patients with a deeper understanding of their health conditions, empowering them to make more informed decisions regarding their treatment. The evolution of LLMs is also contingent upon the development of efficient and precise algorithms capable of adeptly handling complex medical data, which is essential for accurate and timely medical decision-making. As technology progresses, the use of LLMs in the medical field is expected to become increasingly intelligent, efficient, and personalized, thereby enhancing not only the quality of medical services but also the overall patient experience and driving the evolution and transformation of the health care industry.

In our pursuit of technological progress, we must adhere to a fundamental principle: ensuring that technology is accessible to all. This is particularly pertinent in the context of LLM adoption, where it is crucial not to overlook those who may be marginalized by the technology gap [109,110]. Whether addressing the needs of rural farmers or urban older adults, every individual should have the opportunity to benefit from LLMs. This broad adoption must span various geographical regions and encompass diverse languages and cultural contexts, catering to users speaking English, Chinese, or local dialects [111,112]. Achieving this objective is not solely a technological challenge but also a social imperative. We must ensure that the design and application of LLMs overcome language and cultural barriers, truly reaching and benefiting a diverse global populace. Additionally, addressing technology accessibility issues is vital. For individuals in technologically underserved areas or older adults unfamiliar with new technologies, simpler access methods and more user-friendly interfaces are needed to facilitate effortless use of LLMs.

While the potential of LLMs in health care is significant, realizing this potential requires ongoing research, innovation, and dedication. Continuous efforts are necessary to refine LLM technology continually and ensure its broad adoption across all sectors of society. We firmly believe that with sustained commitment, LLMs will catalyze transformative changes in health care, benefiting society at large. By championing technological inclusivity, we can not only enhance the quality and efficiency of medical services but also promote overall societal health and well-being.

Economic Considerations in the Deployment of LLMs in Health Care

LLMs require significant computational resources for training and maintenance, which translates to substantial financial costs. In the medical domain, these costs can be particularly prohibitive due to the need for specialized data, high levels of accuracy, and continuous updates to ensure model relevance and safety.

Training a state-of-the-art LLM, such as GPT-3, requires access to extensive hardware infrastructure, including thousands of GPUs or TPUs, large amounts of RAM, and high-speed data storage solutions [113,114]. According to estimates, the training cost of models, like GPT-3, can reach up to US $1.4 million, with operational costs amounting to several hundred thousand dollars per day when deployed at scale. In a medical context, where accuracy and reliability are paramount, these costs are even higher due to the additional requirements for data security, privacy, and compliance with health care regulations.

Several studies have documented the economic challenges associated with deploying LLMs in health care [115,116]. For instance, the cost of implementing LLMs in hospital settings, including the necessary infrastructure upgrades, staff training, and ongoing maintenance, has been reported to be a major barrier to widespread adoption. Moreover, the need for regular updates to the models, which involves retraining them with new medical data, adds to the operational expenses [1].

As technology advances, it is expected that the costs associated with LLMs will decrease, making them more accessible to a broader range of health care providers. The development of more energy-efficient hardware, combined with advances in machine learning techniques, is likely to contribute to this trend. However, until these cost reductions are realized, careful planning and resource allocation will be essential for any institution looking to implement LLMs in their health care practice.

Conclusions

The era of digitalization and informatization underscores the transformative potential of LLMs in medicine. The evolution of this technology signifies a paradigm shift in medical services, offering unique opportunities and challenges to the medical community. LLMs, with their advanced NLP capabilities, have a wide range of applications including emergency triage, older people care, and the enhancement of digital medical workflows. As the diversity of medical data expands, LLMs' ability to process multimodal data will play a crucial role in enabling more precise, personalized medical diagnoses and treatments.

Despite the promising trajectory of LLMs in the medical field, ensuring their safety and effectiveness in clinical practice remains a critical challenge. Currently, the regulation of LLMs in health care is still in its early stages, with several frameworks being developed to address the unique risks and challenges they pose. Regulatory bodies, such as the US Food and Drug Administration, European Medicines Agency, and China's National Medical Products Administration, have begun to formulate guidelines that apply to AI-driven medical devices including LLMs. These guidelines typically focus on the validation of the models through rigorous clinical trials, ensuring that they meet specific safety, efficacy, and ethical standards before they can be deployed in clinical settings. However, the growth potential of LLMs in the medical arena is significant. They can enhance patient experiences through the integration of virtual reality and augmented reality, offer comprehensive medical advice through multimodal research, and humanize doctor-patient interactions using the theory of mind. With ongoing advancements in algorithms and computational power, we anticipate considerable improvements in LLMs' processing speed and accuracy.

However, the path to technological advancement is not always linear. To ensure the benefits of LLMs are accessible to all, it is imperative to promote equitable development and address the digital divide, particularly for economically and technologically disadvantaged regions and groups. This goal requires the collective efforts of health care professionals, computer science experts, government regulatory bodies, patients, and their families. Such a collaborative approach will ensure that the application of LLM technology in the medical field genuinely contributes to the betterment of humanity, significantly enhancing health and well-being.

Acknowledgments

This study was funded by the National Key R&D Program of China (2020YFC2004705), the National Natural Science Foundation of China (81825003, 81900272, 91957123, 82270376, 623B2003), and the Beijing Nova Program (Z201100006820002) from Beijing Municipal Science & Technology Commission. This study was also supported by the S&T Program of Hebei (22377719D and 22377771D), the Key Science and Technology Research Program of Hebei Provincial Health Commission (20230991), the Industry University Research Cooperation Special Project (CXY2024020), the Hebei Province Finance Department Project (LS202214, ZF2024226), the Hebei Provincial Health Commission Project (20190448), the Key Discipline Construction Project of Shanghai Pudong New Area Health Commission (PWZxk2022-20), the Science and Technology Plan Project of Jiangxi Provincial Health Commission ([SK] P220227143), Tianjin University Science and Technology Innovation Leading Talent Cultivation Project (2024XQM-0024), and Tianjin Natural Science Foundation (23JCQNJC01430). This study was also supported by the China Hebei International Joint Research Center for Structural Heart Disease, and the Hebei Province Finance Department Project (LS202101).

Conflicts of Interest

None declared.

References

1. Arora A, Arora A. The promise of large language models in health care. Lancet 2023;401(10377):641. [doi: 10.1016/S0140-6736(23)00216-7] [Medline: 36841609]
2. Ayers JW, Zhu Z, Poliak A, Leas EC, Dredze M, Hogarth M, et al. Evaluating artificial intelligence responses to public health questions. JAMA Netw Open 2023;6(6):e2317517 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.17517] [Medline: 37285160]
3. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
4. Li R, Kumar A, Chen JH. How chatbots and large language model artificial intelligence systems will reshape modern medicine: fountain of creativity or pandora's box? JAMA Intern Med 2023;183(6):596-597. [doi: 10.1001/jamainternmed.2023.1835] [Medline: 37115531]
5. Minssen T, Vayena E, Cohen IG. The challenges for regulating medical use of ChatGPT and other large language models. JAMA 2023;330(4):315-316. [doi: 10.1001/jama.2023.9651] [Medline: 37410482]

6. Kanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA 2023;330(1):78-80 [FREE Full text] [doi: 10.1001/jama.2023.8288] [Medline: 37318797]
7. Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A comprehensive overview of large language models. ArXiv Preprint posted online on July 12, 2023. [doi: 10.48550/arXiv.2307.06435]
8. Ayoub NF, Lee YJ, Grimm D, Balakrishnan K. Comparison between ChatGPT and Google search as sources of postoperative patient instructions. JAMA Otolaryngol Head Neck Surg 2023;149(6):556-558 [FREE Full text] [doi: 10.1001/jamaoto.2023.0704] [Medline: 37103921]
9. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. ArXiv Preprint posted online on March 31, 2023. [doi: 10.48550/arXiv.2303.18223]
10. Chang K. Natural language processing: recent development and applications. Appl Sci 2023;13(20):11395. [doi: 10.3390/app132011395]
11. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digital Health 2023;5(4):e179-e181 [FREE Full text] [doi: 10.1016/S2589-7500(23)00048-1] [Medline: 36894409]
12. Garcia MB. Using AI tools in writing peer review reports: should academic journals embrace the use of ChatGPT? Ann Biomed Eng 2024;52(2):139-140. [doi: 10.1007/s10439-023-03299-7] [Medline: 37368125]
13. Azerbayev Z, Schoelkopf H, Paster K, Santos MD, McAleer S, Jiang AQ, et al. Llemma: an open language model for mathematics. ArXiv Preprint posted online on October 16, 2023. [doi: 10.48550/arXiv.2310.10631]
14. Xiong W, Guo Y, Chen H. The program testing ability of large language models for code. ArXiv Preprint posted online on October 09, 2023. [doi: 10.48550/arXiv.2310.05727]
15. Yuan A, Coenen A, Reif E, Ippolito D, Assoc CM. Wordcraft: story writing with large language models. 2022 Presented at: Wordcraft Story Writing With Large Language Models. 27th Annual International Conference on Intelligent User Interfaces (ACM IUI); 2022 2022 Mar ; Univ Helsinki, ELECTR NETWORK; 2022; March 22, 2022; Helsinki, Finland p. 841-852. [doi: 10.1145/3490099.3511105]
16. Chu Y, Liu P. Public aversion against ChatGPT in creative fields? Innovation (Camb) 2023;4(4):100449 [FREE Full text] [doi: 10.1016/j.xinn.2023.100449] [Medline: 37485080]
17. Voelker R. The promise and pitfalls of AI in the complex world of diagnosis, treatment, and disease management. JAMA 2023;330(15):1416-1419. [doi: 10.1001/jama.2023.19180] [Medline: 37755919]
18. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
19. He S, Yang F, Zuo JP, Lin ZM. ChatGPT for scientific paper writing-promises and perils. Innovation (Camb) 2023;4(6):100524 [FREE Full text] [doi: 10.1016/j.xinn.2023.100524] [Medline: 38028132]
20. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
21. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ 2023;9:e48291 [FREE Full text] [doi: 10.2196/48291] [Medline: 37261894]
22. Cooper A, Rodman A. AI and medical education—a 21st-century pandora's box. N Engl J Med 2023;389(5):385-387. [doi: 10.1056/NEJMp2304993] [Medline: 37522417]
23. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. JAMA Ophthalmol 2023;141(6):589-597 [FREE Full text] [doi: 10.1001/jamaophthalmol.2023.1144] [Medline: 37103928]
24. Mehrabanian M, Zariat Y. ChatGPT passes anatomy exam. Br Dent J 2023;235(5):295. [doi: 10.1038/s41415-023-6286-7] [Medline: 37684439]
25. Baker S. AI could be an opportunity for research managers. Nature 2023. [doi: 10.1038/d41586-023-03277-y] [Medline: 37857880]
26. Ray PP. Broadening the horizon: a call for extensive exploration of ChatGPT's potential in obstetrics and gynecology. Am J Obstet Gynecol 2023;229(6):706. [doi: 10.1016/j.ajog.2023.07.016] [Medline: 37454961]
27. Kovoor JG, Gupta AK, Bacchi S. ChatGPT: effective writing is succinct. BMJ 2023;381:1125. [doi: 10.1136/bmj.p1125] [Medline: 37220943]
28. Decker H, Trang K, Ramirez J, Colley A, Pierce L, Coleman M, et al. Large language model-based chatbot vs surgeon-generated informed consent documentation for common procedures. JAMA Netw Open 2023;6(10):e2336997 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36997] [Medline: 37812419]
29. Madani A, Krause B, Greene ER, Subramanian S, Mohr BP, Holton JM, et al. Large language models generate functional protein sequences across diverse families. Nat Biotechnol 2023;41(8):1099-1106 [FREE Full text] [doi: 10.1038/s41587-022-01618-2] [Medline: 36702895]
30. Askr H, Elgeldawi E, Ella HA, Elshaier YAMM, Gomaa MM, Hassanien AE. Deep learning in drug discovery: an integrative review and future challenges. Artif Intell Rev 2023;56(7):5975-6037 [FREE Full text] [doi: 10.1007/s10462-022-10306-1] [Medline: 36415536]

31. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ Digital Med 2022;5(1):194 [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
32. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P. Language models are few-shot learners. 2020 Presented at: Proceedings of the 34th International Conference on Neural Information Processing Systems; 2020 December 06; Red Hook, NY, United States p. 1877-1901.
33. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with clip latents. ArXiv Preprint posted online on October 13, 2022. [doi: 10.48550/arXiv.2204.06125]
34. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M. Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Technol 2020;21(1):5485-5551 [FREE Full text]
35. Carpenter KA, Altman RB. Using GPT-3 to build a lexicon of drugs of abuse synonyms for social media pharmacovigilance. Biomolecules 2023;13(2):387 [FREE Full text] [doi: 10.3390/biom13020387] [Medline: 36830756]
36. Ali R, Tang OY, Connolly ID, Fridley JS, Shin JH, Sullivan PLZ, et al. Performance of ChatGPT, GPT-4, and Google bard on a neurosurgery oral boards preparation question bank. Neurosurgery 2023;93(5):1090-1098. [doi: 10.1227/neu.0000000000002551] [Medline: 37306460]
37. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A. Attention is all you need. In: Advances in Neural Information Processing Systems. United States: Curran Associates, Inc; 2017.
38. Manolitsis I, Feretzakis G, Tzelves L, Kalles D, Katsimperis S, Angelopoulos P, et al. Training ChatGPT models in assisting urologists in daily practice. Stud Health Technol Inform 2023;305:576-579. [doi: 10.3233/SHTI230562] [Medline: 37387096]
39. An J, Ding W, Lin C. ChatGPT: tackle the growing carbon footprint of generative AI. Nature 2023;615(7953):586. [doi: 10.1038/d41586-023-00843-2] [Medline: 36944743]
40. Sharir O, Peleg B, Shoham Y. The cost of training nlp models: a concise overview. ArXiv Preprint posted online on April 19, 2020. [doi: 10.48550/arXiv.2004.08900]
41. Srinivasan V, Gandhi D, Thakker U, Prabhakar R. Training large language models efficiently with sparsity and dataflow. ArXiv Preprint posted online on April 11, 2023. [doi: 10.48550/arXiv.2304.05511]
42. Zhou Z, Wang X, Li X, Liao L. Is ChatGPT an evidence-based doctor? Eur Urol 2023;84(3):355-356. [doi: 10.1016/j.eururo.2023.03.037] [Medline: 37061445]
43. Blum J, Menta AK, Zhao X, Yang VB, Gouda MA, Subbiah V. Pearls and pitfalls of ChatGPT in medical oncology. Trends Cancer 2023;9(10):788-790. [doi: 10.1016/j.trecan.2023.06.007] [Medline: 37407364]
44. Duffourc M, Gerke S. Generative AI in health care and liability risks for physicians and safety concerns for patients. JAMA 2023;330(4):313-314. [doi: 10.1001/jama.2023.9630] [Medline: 37410497]
45. Ji J, Qiu T, Chen B, Zhang B, Lou H, Wang K, et al. AI alignment: a comprehensive survey. ArXiv Preprint posted online on October 30, 2023. [doi: 10.48550/arXiv.2310.19852]
46. Ward E, Gross C. Evolving methods to assess chatbot performance in health sciences research. JAMA Intern Med 2023;183(9):1030-1031. [doi: 10.1001/jamainternmed.2023.2567] [Medline: 37459112]
47. Butte AJ. Artificial intelligence-from starting pilots to scalable privilege. JAMA Oncol 2023;9(10):1341-1342. [doi: 10.1001/jamaoncol.2023.2867] [Medline: 37615950]
48. Hu ZY, Han FJ, Yu L, Jiang Y, Cai G. AI-link omnipotent pathological robot: bridging medical meta-universe to real-world diagnosis and therapy. Innovation (Camb) 2023;4(5):100494 [FREE Full text] [doi: 10.1016/j.xinn.2023.100494] [Medline: 37663933]
49. Thirunavukarasu AJ. Large language models will not replace healthcare professionals: curbing popular fears and hype. J R Soc Med 2023;116(5):181-182 [FREE Full text] [doi: 10.1177/01410768231173123] [Medline: 37199678]
50. The Lancet Regional Health-Europe. Embracing generative AI in health care. Lancet Reg Health Eur 2023;30:100677 [FREE Full text] [doi: 10.1016/j.lanepe.2023.100677] [Medline: 37465322]
51. NA. Will ChatGPT transform healthcare? Nat Med 2023;29(3):505-506. [doi: 10.1038/s41591-023-02289-5] [Medline: 36918736]
52. Jiang LY, Liu XC, Nejatian NP, Nasir-Moin M, Wang D, Abidin A, et al. Health system-scale language models are all-purpose prediction engines. Nature 2023;619(7969):357-362 [FREE Full text] [doi: 10.1038/s41586-023-06160-y] [Medline: 37286606]
53. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. JAMA 2023;330(9):866-869. [doi: 10.1001/jama.2023.14217] [Medline: 37548965]
54. Kluger N. Potential applications of ChatGPT in dermatology. J Eur Acad Dermatol Venereol 2023;37(7):e941-e942. [doi: 10.1111/jdv.19152] [Medline: 37102458]
55. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? Lancet Infect Dis 2023;23(4):405-406. [doi: 10.1016/S1473-3099(23)00113-5] [Medline: 36822213]
56. Kulkarni PA, Singh H. Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. JAMA 2023;330(4):317-318. [doi: 10.1001/jama.2023.11440] [Medline: 37410477]
57. Jiang K, Zhu M, Bernard G. Few-shot learning for identification of COVID-19 symptoms using generative pre-trained transformer language models. In: Koprinska I, editor. Machine Learning and Principles and Practice of Knowledge Discovery

in Databases. ECML PKDD 2022. Communications in Computer and Information Science, vol 1753. Switzerland: Springer, Cham; 2023.

58. Miller K, Gunn E, Cochran A, Burstein H, Friedberg JW, Wheeler S, et al. Use of large language models and artificial intelligence tools in works submitted to journal of clinical oncology. J Clin Oncol 2023;41(19):3480-3481. [doi: 10.1200/JCO.23.00819] [Medline: 37200609]
59. Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
60. Hua HU, Kaakour AH, Rachitskaya A, Srivastava S, Sharma S, Mammo DA. Evaluation and comparison of ophthalmic scientific abstracts and references by current artificial intelligence chatbots. JAMA Ophthalmol 2023;141(9):819-824. [doi: 10.1001/jamaophthalmol.2023.3119] [Medline: 37498609]
61. Chen S, Kann BH, Foote MB, Aerts HJWL, Savova GK, Mak RH, et al. Use of artificial intelligence chatbots for cancer treatment information. JAMA Oncol 2023;9(10):1459-1462 [FREE Full text] [doi: 10.1001/jamaoncol.2023.2954] [Medline: 37615976]
62. Marchandot B, Matsushita K, Carmona A, Trimaille A, Morel O. ChatGPT: the next frontier in academic writing for cardiologists or a pandora's box of ethical dilemmas. Eur Heart J Open 2023;3(2):oead007 [FREE Full text] [doi: 10.1093/ehjopen/oead007] [Medline: 36915398]
63. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(25):2400. [doi: 10.1056/NEJMc2305286] [Medline: 37342941]
64. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
65. Kataoka Y, So R. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(25):2399. [doi: 10.1056/NEJMc2305286] [Medline: 37342939]
66. Fernandes AC, Souto MEVC. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(25):2399-2400. [doi: 10.1056/NEJMc2305286] [Medline: 37342940]
67. Gottlieb S, Silvis L. How to safely integrate large language models into health care. JAMA Health Forum 2023;4(9):e233909 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.3909] [Medline: 37733359]
68. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. JAMA Oncol 2023;9(10):1437-1440. [doi: 10.1001/jamaoncol.2023.2947] [Medline: 37615960]
69. Bernstein IA, Zhang YV, Govil D, Majid I, Chang RT, Sun Y, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. JAMA Netw Open 2023;6(8):e2330320 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.30320] [Medline: 37606922]
70. Kiros R, Salakhutdinov R, Zemel R. Multimodal neural language models. 2014 Presented at: Proceedings of the 31st International Conference on International Conference on Machine Learning—Volume 32; June 21, 2014; Beijing China URL: https://proceedings.mlr.press/v32/kiros14.html
71. Driess D, Xia F, Sajjadi MSM, Lynch C, Chowdhery A, Ichter B, et al. Palm-e: an embodied multimodal language model. ArXiv Preprint posted online on March 06, 2023. [doi: 10.48550/arXiv.2303.03378]
72. Zhang T, Liu N, Xu J, Liu Z, Zhou Y, Yang Y, et al. Flexible electronics for cardiovascular healthcare monitoring. Innovation (Camb) 2023;4(5):100485 [FREE Full text] [doi: 10.1016/j.xinn.2023.100485] [Medline: 37609559]
73. Volpe NJ, Mirza RG. Chatbots, artificial intelligence, and the future of scientific reporting. JAMA Ophthalmol 2023;141(9):824-825. [doi: 10.1001/jamaophthalmol.2023.3344] [Medline: 37498621]
74. Li S. ChatGPT has made the field of surgery full of opportunities and challenges. Int J Surg 2023;109(8):2537-2538 [FREE Full text] [doi: 10.1097/JS9.0000000000000454] [Medline: 37195807]
75. O'Hern K, Yang E, Vidal NY. ChatGPT underperforms in triaging appropriate use of Mohs surgery for cutaneous neoplasms. JAAD Int 2023;12:168-170 [FREE Full text] [doi: 10.1016/j.jdin.2023.06.002] [Medline: 37404248]
76. Xu HL, Gong TT, Liu FH, Chen HY, Xiao Q, Hou Y, et al. Artificial intelligence performance in image-based ovarian cancer identification: a systematic review and meta-analysis. EClinicalMedicine 2022;53:101662 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101662] [Medline: 36147628]
77. Zhu M, Chen Z, Yuan Y. DSI-Net: deep synergistic interaction network for joint classification and segmentation with endoscope images. IEEE Trans Med Imaging 2021;40(12):3315-3325. [doi: 10.1109/TMI.2021.3083586] [Medline: 34033538]
78. Cheng K, Wu C, Gu S, Lu Y, Wu H, Li C. WHO declares the end of the COVID-19 global health emergency: lessons and recommendations from the perspective of ChatGPT/GPT-4. Int J Surg 2023;109(9):2859-2862 [FREE Full text] [doi: 10.1097/JS9.0000000000000521] [Medline: 37246993]
79. Madden MG, McNicholas BA, Laffey JG. Assessing the usefulness of a large language model to query and summarize unstructured medical notes in intensive care. Intensive Care Med 2023;49(8):1018-1020. [doi: 10.1007/s00134-023-07128-2] [Medline: 37338549]

80. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. Lancet Digital Health 2023;5(6):e333-e335 [FREE Full text] [doi: 10.1016/S2589-7500(23)00083-3] [Medline: 37120418]
81. van Heerden AC, Pozuelo JR, Kohrt BA. Global mental health services and the impact of artificial intelligence-powered large language models. JAMA Psychiatry 2023;80(7):662-664. [doi: 10.1001/jamapsychiatry.2023.1253] [Medline: 37195694]
82. Kwok K, Wei W, Tsoi M, Tang A, Chan M, Ip M, et al. How can we transform travel medicine by leveraging on AI-powered search engines? J Travel Med 2023;30(4):taad058. [doi: 10.1093/jtm/taad058] [Medline: 37098161]
83. NA. AI-powered structure-based drug design inspired by the lock-and-key model. Nat Comput Sci 2023;3(10):827-828. [doi: 10.1038/s43588-023-00552-w] [Medline: 38177772]
84. NA. Upswing in AI drug-discovery deals. Nat Biotechnol 2023;41(10):1361. [doi: 10.1038/s41587-023-02002-4] [Medline: 37828287]
85. Xu Z, editor. Using large pre-trained language model to assist FDA in premarket medical device classification. In: SoutheastCon. Orlando, FL, USA: IEEE; 2023.
86. Wu C, Lei J, Zheng Q, Zhao W, Lin W, Zhang X. Can GPT-4V (ision) Serve Medical Applications? Case Studies on GPT-4V for Multimodal Medical Diagnosis. ArXiv Preprint posted online on October 15, 2023. [doi: 10.48550/arXiv.2310.09909]
87. He K, Mao R, Lin Q, Ruan Y, Lan X, Feng M. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. ArXiv Preprint posted online on October 09, 2023. [doi: 10.48550/arXiv.2310.05694]
88. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med 2023;388(13):1201-1208. [doi: 10.1056/NEJMra2302038] [Medline: 36988595]
89. Marks M, Haupt CE. AI chatbots, health privacy, and challenges to HIPAA compliance. JAMA 2023;330(4):309-310. [doi: 10.1001/jama.2023.9458] [Medline: 37410450]
90. Mello MM, Guha N. ChatGPT and physicians' malpractice risk. JAMA Health Forum 2023;4(5):e231938 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.1938] [Medline: 37200013]
91. Kanter GP, Packel EA. Health care privacy risks of AI chatbots. JAMA 2023;330(4):311-312. [doi: 10.1001/jama.2023.9618] [Medline: 37410449]
92. NA. ChatGPT is a black box: how AI research can break it open. Nature 2023;619(7971):671-672. [doi: 10.1038/d41586-023-02366-2] [Medline: 37491394]
93. Tang YD, Dong ED, Gao W. LLMs in medicine: the need for advanced evaluation systems for disruptive technologies. Innovation (Camb) 2024;5(3):100622 [FREE Full text] [doi: 10.1016/j.xinn.2024.100622] [Medline: 38699776]
94. Grigorian A, Shipley J, Nahmias J, Nguyen N, Schwed AC, Petrie BA, et al. Implications of using chatbots for future surgical education. JAMA Surg 2023;158(11):1220-1222. [doi: 10.1001/jamasurg.2023.3875] [Medline: 37728910]
95. Kozlov M, Biever C. AI 'breakthrough': neural net has human-like ability to generalize language. Nature 2023;623(7985):16-17. [doi: 10.1038/d41586-023-03272-3] [Medline: 37880533]
96. Ferryman K, Mackintosh M, Ghassemi M. Considering biased data as informative artifacts in AI-assisted health care. N Engl J Med 2023;389(9):833-838. [doi: 10.1056/NEJMra2214964] [Medline: 37646680]
97. Tan TF, Teo ZL, Ting DSW. Artificial intelligence bias and ethics in retinal imaging. JAMA Ophthalmol 2023;141(6):552-553. [doi: 10.1001/jamaophthalmol.2023.1490] [Medline: 37140916]
98. Harris E. Large language models answer medical questions accurately, but can't match clinicians' knowledge. JAMA 2023;330(9):792-794. [doi: 10.1001/jama.2023.14311] [Medline: 37548971]
99. Kim J, Cai ZR, Chen ML, Simard JF, Linos E. Assessing biases in medical decisions via clinician and AI chatbot responses to patient vignettes. JAMA Netw Open 2023;6(10):e2338050 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.38050] [Medline: 37847506]
100. Piao J, Liu J, Zhang F, Su J, Li Y. Human–AI adaptive dynamics drives the emergence of information cocoons. Nat Mach Intell 2023;5(11):1214-1224. [doi: 10.1038/s42256-023-00731-4]
101. Nordling L. How ChatGPT is transforming the postdoc experience. Nature 2023;622(7983):655-657. [doi: 10.1038/d41586-023-03235-8] [Medline: 37845528]
102. The Lancet. AI in medicine: creating a safe and equitable future. Lancet 2023;402(10401):503. [doi: 10.1016/S0140-6736(23)01668-9] [Medline: 37573071]
103. The Lancet. AI in medicine: creating a safe and equitable future. Lancet 2023;402(10401):503. [doi: 10.1016/S0140-6736(23)01668-9] [Medline: 37573071]
104. Topol EJ. Machines and empathy in medicine. Lancet 2023;402(10411):1411. [doi: 10.1016/S0140-6736(23)02292-4] [Medline: 37865458]
105. Hagendorff T, Fabi S, Kosinski M. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. Nat Comput Sci 2023;3(10):833-838 [FREE Full text] [doi: 10.1038/s43588-023-00527-x] [Medline: 38177754]

106. Yang R, Tan TF, Lu W, Thirunavukarasu AJ, Ting DSW, Liu N. Large language models in health care: development, applications, and challenges. Health Care Sci 2023;2(4):255-263 [FREE Full text] [doi: 10.1002/hcs2.61] [Medline: 38939520]
107. Wu Y, Si Z, Gong H, Zhu SC. Learning active basis model for object detection and recognition. Int J Comput Vis 2009;90(2):198-235. [doi: 10.1007/s11263-009-0287-0]
108. Horgan J. The consciousness conundrum. IEEE Spectr 2008;45(6):36-41. [doi: 10.1109/mspec.2008.4531459]
109. Lam K. ChatGPT for low- and middle-income countries: a Greek gift? Lancet Reg Health West Pac 2023;41:100906 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100906] [Medline: 37745974]
110. Wang X, Sanders HM, Liu Y, Seang K, Tran BX, Atanasov AG, et al. ChatGPT: promise and challenges for deployment in low- and middle-income countries. Lancet Reg Health West Pac 2023;41:100905 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100905] [Medline: 37731897]
111. Hswen Y, Voelker R. New AI tools must have health equity in their DNA. JAMA 2023;330(17):1604-1607. [doi: 10.1001/jama.2023.19293] [Medline: 37819644]
112. Seghier ML. ChatGPT: not all languages are equal. Nature 2023;615(7951):216. [doi: 10.1038/d41586-023-00680-3] [Medline: 36882613]
113. Tan B, Zhu Y, Liu L, Wang H, Zhuang Y, Chen J. RedCoast: a lightweight tool to automate distributed training of LLMs on any GPU/TPUs. 2024 Presented at: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: System Demonstrations); June 01, 2024; Mexico City, Mexico p. 137-147. [doi: 10.18653/v1/2024.naacl-demo.14]
114. Narayanan D, Shoeybi M, Casper J, LeGresley P, Patwary M, Korthikanti V, et al. Efficient large-scale language model training on GPU clusters using megatron-LM. 2021 Presented at: SC21: International Conference for High Performance Computing, Networking, Storage and Analysis; November 13, 2021; New York, NY, United States p. 1-15. [doi: 10.1145/3458817.3476209]
115. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
116. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: a scoping review. iScience 2024;27(5):109713 [FREE Full text] [doi: 10.1016/j.isci.2024.109713] [Medline: 38746668]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations from Transformers
GPU: graphics processing unit
LLM: large language model
MRI: magnetic resonance imaging
NLP: natural language processing
TPU: tensor processing unit

Edited by A Schwartz; submitted 01.04.24; peer-reviewed by Y Khan, K Singh, S Mao, D Patel, D Ghosh, P Sarmadi; comments to author 13.08.24; revised version received 26.08.24; accepted 10.09.24; published 07.01.25.

Please cite as:
Zhang K, Meng X, Yan X, Ji J, Liu J, Xu H, Zhang H, Liu D, Wang J, Wang X, Gao J, Wang YGS, Shao C, Wang W, Li J, Zheng MQ, Yang Y, Tang YD
Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine
J Med Internet Res 2025;27:e59069
URL: https://www.jmir.org/2025/1/e59069
doi:10.2196/59069
PMID:

Original Paper

Association Between Digital Front Doors and Social Care Use for Community-Dwelling Adults in England: Cross-Sectional Study

Jinbao Zhang[1], PhD; Jonathan E Prunty[2], PhD; Alison C Charles[3], PhD; Julien Forder[1†], PhD

[1]Personal Social Services Research Unit, University of Kent, Canterbury, United Kingdom

[2]Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, United Kingdom

[3]Centre for Health Services Studies, University of Kent, Canterbury, United Kingdom

[†]deceased

Corresponding Author:
Jinbao Zhang, PhD
Personal Social Services Research Unit
University of Kent
Cornwallis Central
Giles Ln
Canterbury, CT2 7NF
United Kingdom
Phone: 44 01227 823862
Email: j.zhang@kent.ac.uk

Abstract

Background: Requests for public social care support can be made through an online portal. These digital “front doors” can help people navigate complex social care systems and access services. These systems can be set up in different ways, but there is little evidence about the impact of alternative arrangements. Digital front-door systems should help people better access services, particularly low-intensity services (high-intensity care is likely to require a full in-person assessment).

Objective: This study aimed to investigate the association between 2 primary digital front door arrangements, easy-read information, and self-assessment tools provided on official websites, and the type of social care support that is offered: ongoing low-level support (OLLS), short-term care (STC) and long-term care (LTC).

Methods: Information on front door arrangements was collected from the official websites of 152 English local authorities in 2021. We conducted a cross-sectional analysis using aggregated service use data from official government returns at the local authority level. The independent variables were derived from the policy information collected, specifically focusing on the availability of online digital easy-read information and self-assessment tools for adults and caregivers through official websites. The dependent variables were the rates of using social care support, including OLLS, STC, and LTC, across different age groups: the adult population (aged 18 and older), younger population (aged between 18 and 64 years), and older population (aged 65 and older). Multivariate regression analysis was used to examine the association between digital front door arrangements and access to social care support, controlling for population size, dependency level, and financial need factors.

Results: Less than 20% (27/147) of local authorities provided an integrated digital easy-read format as part of their digital front door system with about 25% (37/147) adopting digital self-assessment within their system. We found that local authorities that offered an integrated digital easy-read information format showed higher rates of using OLLS (β coefficient=0.54; *P*=.03; but no statistically significant association with LTC and STC). The provision of an online self-assessment system was not associated with service use in the 1-year (2021) cross-sectional estimate, but when 2 years (2020 and 2021) of service-use data were analyzed, a significant positive association was found on OLLS rates (β coefficient=0.41; *P*=.21). Notably, these findings were consistent across different age groups.

Conclusions: These findings are consistent with our hypothesis that digital systems with built-in easy-read and self-assessment may make access to (low-intensity) services easier for people. Adoption of these arrangements could potentially help increase the uptake of support among those who are eligible, with expected benefits for their care-related well-being. Given the limited adoption of the digital front door by local authorities in England, expanding their use could improve care-related outcomes and save social care costs.

(J Med Internet Res 2025;27:e53205) doi:10.2196/53205

KEYWORDS

social care support; long term care; access; front door; easy-read; self-assessment; system navigation; digital system; digital technology; internet

Introduction

Background

The “front door” to the adult social care system refers to the channels through which people collect service information, apply for needs assessment, and access social care support. Social care systems vary in how they configure their front door and information may not always be well organized or easily accessible, and different services have specific eligibility requirements based on factors such as wealth, income, and physical care needs. Navigating the front door can be demanding and time consuming for community-dwelling people with functional impairments [1]. When requesting services, people need to proactively acquire service information and approach care professionals to complete face-to-face needs assessment [2]. Without accessible information and streamlined application procedures, individuals may be less able to access social care support, causing unmet needs to escalate [3,4]. Therefore, developing front door arrangements is crucial to ensure that people can access available and timely services, potentially preventing or delaying the escalation of care needs.

Researchers and policy makers have recently become increasingly attentive to deploying digital front door arrangements to facilitate service access. While a consensus on the definition of the digital front door is lacking, we define it as the channels through which individuals request and access services using digital platforms or technologies [5]. Digital technologies may facilitate service access in three primary ways: (1) enabling initial contact with services, (2) substituting traditional face-to-face interactions with remote services, and (3) facilitating access to professional support through innovative technologies [6]. For instance, an online information and referral tool, BenefitsCheckUp, increased the take-up rate of Medicaid among low-income older Americans [7]. In England, the Care Act 2014 has stipulated that local authorities are responsible for providing digital channels (eg, websites and social media) to help individuals make informed choices, which can also reduce public expenditure at the front door [8,9]. Despite the growing emphasis on digital front door arrangements, few studies have examined whether digital channels can effectively promote service access [6].

To fill the gap, this study explores the association between the digital front door and the use of social care support in England. Specifically, we define the digital front door as whether local authorities provide online digital easy-read information and self-assessment tools for adults and caregivers through official websites. In addition, we investigate whether access to different types of social care support is correlated with local authorities’ digital front door arrangements. This study contributes to understanding the relationship between the digital front door and service access, providing insights into how to organize these systems to facilitate access to social care support. The findings could inform policy makers about developing digital front door arrangements to promote service access, eventually enhancing people’s well-being.

Digital Front Doors and Access to Social Care Support

People needing social care have generally relied on social networks (eg, peers and family members) and health and social care professionals as primary resources for information and assistance [1,4,10]. Since the COVID-19 pandemic, digital channels have become an increasingly viable alternative that could save costs and facilitate service access [6,8,11]. The digital front door can do this by alleviating 2 main obstacles to service use: information asymmetry and administrative burdens [2,4,10,12].

Providing online information can address information asymmetry and facilitate service access by potentially simplifying applications and allowing immediate completion. They eliminate travel costs as well as removing waiting times for printed versions of documents [13]. Previous studies have suggested that online information should be presented from trusted sources and in easy-read information formats to meet individual preferences [14,15]. Specifically, people trust information from official websites and medical professionals. Also, easy-read information is generally characterized by plain language, simple layout, large font size, and the adoption of images. Despite individual preferences for digital easy-read information from government websites, few studies have explored the relationship between such information and access to social care support.

In addition to easy-read information online information, online self-assessment tools mitigate the administrative burden associated with applying for social care support. Self-assessment aims to reduce direct professional involvement in the assessment process. Savings can arise from reduced paperwork and waiting, travel, and face-to-face time associated with professionally led needs assessment [16,17]. Self-assessment approaches could encourage access for people with low-level needs when perceived barriers to application may outweigh the potential benefits [16,18]. Although previous studies mainly highlighted the positive association between service access and self-assessment tools [16,17], little work has focused on their relationship in an online context. Potential arguments against self-assessment include a potential lack of identification of person-specific needs and idiosyncrasies [19].

Social Care and Front Doors in England

Local authorities in England are responsible for delivering adult social care for people with care needs. In order to access care, individuals need to navigate the front door of social care systems, including gathering information on services and completing needs assessment [20]. Information is available in different formats, such as printed leaflets or brochures, conversations with professionals, and the internet (especially official websites) [21]. Though the Care Act 2014 stipulated local authorities’ responsibilities for providing multiple information formats, the proportion of older people who found it easy to obtain information had fallen from 75.2% in 2016 to 65.6% in 2021 [22]. This decline indicates the need for understanding how to provide information and alleviate information asymmetry to promote service access.

Traditionally, assessment has been performed by professionals through telephone and face-to-face communications. As of August 2022, there were half a million people waiting for a care assessment, for their care and support to begin, or for a review of their care plan [23]. By March 2023, the number of people who had been waiting for over 6 months for their care assessment rose to 82,087. Every year, local authorities can together expect around 2 million requests for care, with an average of 5420 requests for support received each day [24]. Within this context, many local authorities introduced self-assessment as a complementary tool for professionally led assessments [11,16].

People generally access publicly funded services following a hospital discharge or are referred from the community. In 2021, most (79.1%) of care requests originated from the community, while 18.7% were discharged from the hospital (with 2.2% from other routes) [24]. While the number of requests has grown during the past 7 years, the pattern of requests by route of access has largely remained unchanged.

Following the care request, people eligible for publicly funded services can receive 3 main types of support: short-term care (STC), long-term care (LTC), and other services, including end-of-life care and ongoing low-level support (OLLS) that targets community-dwelling people with minimal care needs and offers them ongoing services (eg, telecare, minicom live and community alarm).

Study Aims

Despite the development of the digital front door in England, the relationship between these arrangements and service access remains unclear. Understanding their associations could guide policy makers and practitioners to improve front door arrangements and facilitate service use, thus improving people's well-being. Accordingly, this study investigates the prevalence of providing easy-read information online information and self-assessment tools through official websites and the association between such digital channels and social care use.

We focus on people requesting care from the community because they are most likely to use digital front door arrangements. By contrast, for people being discharged from the hospital, there are generally different and more specific arrangements to access social care. Furthermore, people discharged from hospital are likely to have higher levels of need. Access from the community accounts for approximately 80% of all care requests, people also tend to wait longer for assessment and support (60 days) than those discharged from the hospital (38 days) [25].

Methods

Data

This study used upper-tier local-authority-level data from local authorities' websites, the short and long-term (SALT) collection on care access, use, and expenditure [24], and Stat-Xplore for benefits data (Department for Work and Pensions). To identify local authorities' digital front door arrangements, we gathered policy information from 152 local authorities' websites between December 2022 and April 2023 and coded these documents based on established criteria. Specifically, we defined local authorities that used the "easy-read information" keyword in their official websites to introduce adult social care systems and application procedures as providing easy-read information. Those that did not have such a keyword, or did not permit immediate access through official websites, were coded as not having easy-read information. Our choice of focusing on "easy-read information" allows us to use a straightforward criterion to identify intentional effort by local authorities to provide online digital easy-read information. Likewise, local authorities that provided online self-assessment forms (excluding contact forms) were coded as having online self-assessment tools for adults and caregivers, respectively. To ensure the consistency of policy texts, 2 researchers (JZ and AC) independently coded each local authority's digital front door arrangements. Disagreements were addressed through discussions between the 2 investigators (JZ and AC).

We also obtained information about how clients accessed care services (through the community or hospital route) and their subsequent care destination (STC, LTC, and other services) from the SALT data. This data set has been published annually since 2016 and contains information about clients' journeys through the social care system in England, including the number of requests for social care, the access route for people requesting support, and their care sequel (what happened next, eg, community and residential care).

In addition, this study collected local area characteristics, such as population estimates and pension credit, from the Stat-Xplore website. Stat-Xplore website provides aggregated benefit data administered by the Department for Work and Pensions, including pension credit and Carer's Allowance. When combining digital front door information with data from SALT and Stat-Xplore, we only included the latest wave (2021) to ensure that our results would be representative of current policy arrangements. Given that data was not available from 5 local authorities (Hackney, City of London, Isles of Scilly, North Northamptonshire, and West Northamptonshire), data from 147 local authorities only were used in our study. The data are available in Multimedia Appendix 1.

Measurement

Dependent Variables

The outcome of interest was the rate of using social care support, including LTC, STC, and OLLS. Specifically, reablement services, an important type of STC that supports individuals to regain independence after an illness or hospital discharge, were included in our analysis. We also identified 3 types of LTC:

community care, residential care, and nursing services. For each service type, we calculated rates using the number of people receiving the service as the numerator and the population aged 18 years or 65 years and older as the denominator.

Independent variables

We assessed local authorities' digital front arrangements with 3 dichotomous variables, whether local authorities provided easy-read information online information regarding adult social care and whether they provided online self-assessment for both cared-for adults and for caregivers (no=0, yes=1).

Covariates

Population size, dependency level, and financial need factors were selected as covariates [26]. Population size was measured using the number of people (100,000) in 3 age groups: 18 years and older, between 18 years and 64 years, and 65 years or older. Proxy variables for care needs included the proportion of the older population aged 80 years and older, and the proportion of older people who received Attendance Allowance (the primary universal benefit for older people with social care needs). Financial need was assessed by the number of recipients receiving pension credit divided by the older population, and the proportion of the population receiving carer's allowance (cash benefit for caregivers who provide care at least 35 hours per week and earn less than £139 (US $177) per week in 2023).

Analysis

Multivariate regression analysis was performed. The main analysis used a cross-sectional ordinary least squares (OLS) estimation. The OLS approach was deemed appropriate for this analysis because our dependent variables are continuous. To correct the right-skewed distribution of the dependent variables, a natural log transformation was applied before analysis. This transformation helps ensure the normality assumption of OLS regression is satisfied, a method commonly used in previous studies [27,28]. We were limited to using SALT data from 2021 as the data for 2022 were not yet published at the time of analysis. As such, there is a small mismatch in timings between our categorization of front door arrangements and the care use data. Nonetheless, given that the rate of change for these variables is relatively slow, we argue that this is an acceptable limitation.

Listwise deletion was used to handle missing values on 2 dependent variables (ie, LTC for older people and OLLS). In both cases, less than 0.7% (1/147) of the cases were missing. Given the potential problem of multicollinearity, we conducted collinearity diagnostics. The mean variance inflation factors ranged from 1.28 to 2.49 and thus did not exceed the suggested threshold (variance inflation factors >10), indicating no evidence of multicollinearity [29].

Ethical Considerations

This study used publicly available data aggregated at the local authority level. The data did not involve the collection of any personally identifiable information, nor did it involve direct interaction with human participants. As a result, this research did not require ethical approval from an ethics committee.

Results

Descriptive Statistics

Table 1 summarizes relevant characteristics of the 147 local authorities in 2021 by age group, that is, all adults aged 18 years and older, younger people aged between 18 years and 64 years, and older people aged 65 years and older. The average rate of using LTC, STC, and OLLS per 100,000 population for all adults was 4.73, 7.04, and 11.53, respectively. Approximately 20% (27/147) of local authorities provided digital, web-based easy-read information online information about adult social care systems. The proportion of local authorities that provided online digital self-assessment for adults and caregivers was 25% (37/147) and 27% (40/147), respectively. The average adult population (aged 18 years and older) across local authorities was 294,000, while the average older population (aged 65 years and older) was 63,000.

In addition, we used chi-square tests to identify whether the provision of the digital front door varied among the 9 larger administrative areas in England. The results are presented in Table S1 in Multimedia Appendix 2, which suggests significant differences across regions. Among the 9 regions, the East of England had the highest proportion of local authorities that offered easy-read information online information, and the Southeast had the highest proportion of local authorities providing online self-assessment for adults and caregivers. In contrast, local authorities in the Northeast and the Northwest did not offer easy-read information online information, and the Southwest had the lowest proportion of local authorities providing online self-assessments for adults and caregivers.

Table 1. Characteristics of local authorities (n = 147) in England in 2021 by age group.

Variable	Age 18 years and older, mean (SD)	Age 18 years and 64 years, mean (SD)	Age 65 years and older, mean (SD)
Dependent variables			
Long-term care rate	4.73 (3.30)	3.55 (2.57)	8.92 (5.60)
Short-term care rate	7.04 (7.36)	5.53 (6.34)	12.43 (11.39)
Ongoing low-level support rate	11.53 (11.92)	8.90 (9.16)	20.68 (21.33)
Independent variables			
Easy-read information (=1)	0.18 (0.39)	0.18 (0.39)	0.18 (0.39)
Self-assessment for adults (=1)	0.25 (0.44)	0.25 (0.44)	0.25 (0.44)
Self-assessment for caregivers (=1)	0.27 (0.45)	0.27 (0.45)	0.27 (0.45)
Covariates			
Population size (100,000)	2.94 (2.22)	2.24 (1.63)	0.70 (0.63)
Proportion of older people aged 80 years and older	0.27 (0.02)	—[a]	0.27 (0.02)
Proportion of older population receiving Attendance Allowance	0.50 (0.07)	—	0.50 (0.07)
Proportion of older population receiving pension credit	0.54 (0.24)	—	0.54 (0.24)
Proportion of population receiving caregiver's allowance	0.10 (0.04)	0.11 (0.04)	0.10 (0.04)

[a] Not applicable.

Digital Front Doors and Social Care Support

Table 2 shows the association between the digital front door and the rate of using social care support. Panels 1, 2, and 3 show the estimates for LTC, STC, and OLLS, respectively. For each outcome, we investigated the association with the digital front door by three age groups: (1) all adults aged 18 years and older, (2) younger people aged between 18 years and 64 years, and (3) older people aged 65 years and above. The rate of using OLLS for all adults (β coefficient=0.54; *P*=.03), younger people (β coefficient=0.48; *P*=.04), and older people (β coefficient=0.53; *P*=.04) was positively associated with providing easy-read information online information. Providing online self-assessment tools for adults and caregivers was not significantly related to the rate of using OLLS. Providing easy-read information online information and online self-assessments for adults and caregivers were not significantly associated with LTC and STC.

Given the small sample size in our main analysis, we conducted a robustness check by also including care use data for 2020, assuming the same configuration of easy-read information and self-assessment as for 2021. The results are presented in Table S3 in Multimedia Appendix 2. We found that providing easy-read information online information was still positively associated with the rate of using OLLS, regardless of age group. However, providing online self-assessment for adults was now positively associated with the rate of using OLLS for all age groups using this larger sample.

Given that LTC and STC are general categories that encompass many types of services, we investigated the association between the digital front door and specific subdivisions of these services including restorative services, community care, residential care, and nursing care. The results of this analysis are summarized in Table S4 in Multimedia Appendix 2, together with the relationship between the digital front door and funded social care support. All associations between these types of services and providing easy-read information and online self-assessment were non-significant for both adults and caregivers.

Table 2. Results of multivariate regression analysis examining the association between the digital front door and social care support in England in 2021

	Age 18 years and older		Age 18 years and 64 years		Age 65 years and older	
	Values	*P* value	Values	*P* value	Values	*P* value
Panel 1: Long-term care						
Easy-read information, β coefficient (SE)	–0.15 (0.15)	.32	–0.14 (0.15)	.34	–0.03 (0.10)	.78
Self-assessment for adults, β coefficient (SE)	0.10 (0.12)	.42	0.07 (0.13)	.56	0.08 (0.12)	.48
Self-assessment for caregivers, β coefficient (SE)	–0.04 (0.11)	.72	–0.01 (0.13)	.91	–0.06 (0.11)	.59
Covariates[a]	Yes	—[b]	Yes	—	Yes	—
Local authorities, n	147	—	146	—	146	—
F test (*df*)	3.66 (8)	<.001	3.63 (5)	.004	1.67 (8)	.11
R^2	0.22	—	0.13	—	0.10	—
Panel 2: Short-term care						
Easy-read information, β coefficient (SE)	–0.32 (0.24)	.19	–0.35 (0.26)	.17	–0.30 (0.22)	.18
Self-assessment for adults, β coefficient (SE)	–0.21 (0.25)	.39	–0.25 (0.25)	.33	–0.19 (0.25)	.44
Self-assessment for caregivers, β coefficient (SE)	0.22 (0.25)	.38	0.25 (0.25)	.33	0.22 (0.26)	.39
Covariates	Yes	—	Yes	—	Yes	—
Local authorities, n	147	—	147	—	147	—
F test (*df*)	1.73 (8)	.10	2.27 (5)	.05	0.98 (8)	.11
R^2	0.11	—	0.08	—	0.06	—
Panel 3: Ongoing low-level support						
Easy-read information, β coefficient (SE)	0.54 (0.24)	.03	0.48 (0.24)	.04	0.53 (0.25)	.04
Self-assessment for adults, β coefficient (SE)	0.37 (0.27)	.17	0.34 (0.28)	.23	0.35 (0.27)	.19
Self-assessment for caregivers, β coefficient (SE)	–0.04 (0.28)	.88	–0.03 (0.29)	.91	–0.05 (0.28)	.85
Covariates	Yes	—	Yes	—	Yes	—
Local authorities, n	146	—	146	—	146	—
F test (*df*)	6.86 (8)	<.001	8.16 (5)	<.001	4.42 (8)	<.001
R^2	0.23	—	0.20	—	0.19	—

[a] We presented estimators of all covariates in Table S2 in Multimedia Appendix 2.

[b] Not applicable.

Discussion

Principal Findings

To our knowledge, this is the first study that provides information regarding the prevalence of digital easy-read information and self-assessment tools from official websites in England and investigates the association between these tools and access to social care support. Using data at the local authority level, we found that only 20% (27/147) of local authorities provided online, digital easy-read information, and approximately 25% (37/147) used digital self-assessment approaches to promote service access. This is important as we found local authorities that provided easy-read information had a higher rate of using OLLS than those without such a front door arrangement.

Comparison With Previous Work

Providing easy-read information has been shown to reduce learning costs and enhance comprehension [30,31]. Our study adds to this literature by demonstrating the association between easy-read information online information and service access. The availability of digital easy-read information was positively associated with the use of OLLS, but we found no significant association with rates of LTC or STC use. One possible interpretation as to why easy-read information online information only facilitated access for those eligible for low-level ongoing care is that people using web-based digital tools to access care services may have less physical and cognitive impairment, that is lower needs, than their counterparts [32,33]. Conversely, people with more severe needs may have a higher likelihood of being referred directly to a full professional assessment, bypassing the need for online access or assessment. People with lower needs are also more likely to be assessed as needing OLLS rather than long-term and STC [34].

We did not find a significant association between the availability of online self-assessment tools for adults and caregivers, and access to social care support (although alternative analysis using 2 years of data did find a positive association). In practice, the provision of an online self-assessment facility does not necessarily imply that such a system is fully used. One pilot program in England showed that about two-thirds of people requested a face-to-face professionally led assessment rather than a self-assessment, and this was especially true for those with urgent and complex needs [35]. This lower rate of online self-assessment use may explain the nonsignificant association between online self-assessment and service use. In contrast, people with less intensive needs typically prefer using a self-assessment, especially when professional assistance is readily available [16]. Therefore, policy makers could develop both professionally led assessments and online self-assessments to meet the individual preferences of service users. Also, online self-assessment could be augmented with online support from care professionals (eg, through a chat function) to further assist users.

Limitations

This study has several limitations to consider. First, the digital front door is a broad concept with multiple policy components, which may not be fully measured. To address this challenge, we investigated commonly discussed barriers to service access and digital front door arrangements while acknowledging that some digital channels are not operationalized in this study. For example, individuals may use information and assistance from nonprofit organizations' websites, such as Age UK and Carers UK. Third-sector websites are an essential aspect of the digital front door, and future research could examine their impact on service access. In addition, although local authorities that provide easy-read information typically incorporate the term "easy-read information," we acknowledge that some websites offer such content without explicitly using this exact keyword.

Second, this study provides an up-to-date snapshot of the digital front door in England by offering a cross-sectional analysis of current policy practices. Information regarding the digital front door between late 2022 and early 2023 was combined with social care data from 2021 to 2022 as these were the most up-to-date statistics available at the time of analysis. As noted above, given that general changes in service use are relatively slow, we believe this limitation to be minimal. Nonetheless, we do accept the possibility that, in some cases, the 2022 and 2023 social care data may better reflect the association with current digital front door policies, leading to smaller estimated effect sizes in the current analysis.

Finally, the cross-sectional design of this study limits our ability to establish causal relationships between the digital front door and social care use. While we assumed that these digital arrangements promote access to social care, the possibility of reverse causality cannot be excluded. For example, local authorities facing high demand for OLLS may introduce online digital easy-read information to reduce service inquiries, resulting in a potentially spurious correlation between these digital systems and service use. Therefore, our findings should be interpreted as indicative associations rather than causal relationships. Future research could use longitudinal data and causal inference methods to more rigorously examine the impact of the digital front door on access to social care support.

Strengths and Implications

Despite these limitations, this study is among the first to empirically examine the association between providing easy-read information and self-assessment tools from official websites and service access. The findings presented here could inform policy makers interested in developing digital channels for service access. They are also relevant for discussions about increasing service use, reducing unmet needs, and enhancing the well-being of service users.

Our findings have implications for policy makers and practitioners who aim to promote service use. Though digital channels have the potential to save costs, we found relatively few local authorities provided online digital easy-read information and online self-assessment tools, as noted above. Although we do not have information about the costs of implementing such a facility, our findings suggest a positive association with (low-level) service use. This association could potentially lead to lower levels of unmet need and better care-related outcomes, which may, in turn, lead to cost-savings downstream. Given that digital channels might facilitate service access, there is a case for a wider roll-out of digital easy-read information facilities, particularly where the costs of implementation are minimal.

Acknowledgments

This research was funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Kent, Surrey, Sussex (grant NIHR 200179). The views are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health and Social Care.

Data Availability

All data generated or analyzed during this study are included in this published article and its Multimedia Appendix files.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data.
[XLSX File (Microsoft Excel File), 209 KB - jmir_v27i1e53205_app1.xlsx]

Multimedia Appendix 2
Supplementary results.
[DOCX File , 49 KB - jmir_v27i1e53205_app2.docx]

References

1. Bieber A, Nguyen N, Meyer G, Stephan A. Influences on the access to and use of formal community care by people with dementia and their informal caregivers: a scoping review. BMC Health Serv Res 2019 Feb 01;19(1):88 [FREE Full text] [doi: 10.1186/s12913-018-3825-z] [Medline: 30709345]
2. Felix HC, Ali M, Bird TM, Cottoms N, Stewart MK. Are community health workers more effective in identifying persons in need of home and community-based long-term services than standard-passive approaches. Home Health Care Serv Q 2019;38(3):194-208. [doi: 10.1080/01621424.2019.1604461] [Medline: 31060448]
3. Penkunas MJ, Eom KY, Chan AWM. Classification trees for identifying non-use of community-based long-term care services among older adults. Health Policy 2017 Oct;121(10):1093-1099. [doi: 10.1016/j.healthpol.2017.05.008] [Medline: 28942091]
4. Xie Y, Hamilton M, Peisah C, Anstey KJ, Sinclair C. Navigating community-based aged care services from the consumer perspective: a scoping review. Gerontologist 2024 Feb 01;64(2):gnad017. [doi: 10.1093/geront/gnad017] [Medline: 37120292]
5. Rastogi N. Healthcare’s new frontier: the digital front door. BMJ Innov 2022 Mar 15;8(2):129-132. [doi: 10.1136/bmjinnov-2021-000874]
6. Kunonga TP, Spiers GF, Beyer FR, Hanratty B, Boulton E, Hall A, et al. Effects of digital technologies on older people's access to health and social care: umbrella review. J Med Internet Res 2021 Nov 24;23(11):e25887 [FREE Full text] [doi: 10.2196/25887] [Medline: 34821564]
7. Napier A, Rhodes B, Khavjou O, Knowles M, Popham L, Walsh EG, et al. Impact of a web-based information and referral tool on access to federal and state programs for older adults. J Aging Soc Policy 2023 May 04;35(3):360-373. [doi: 10.1080/08959420.2021.1926206] [Medline: 34016014]
8. Bolton J. Six Steps to Managing Demand in Adult Social Care: A Performance Management Approach. Oxford, UK: Oxford Brookes University; 2017.
9. Manthorpe J, Moriarty J, Damant J, Brimblecombe N, Marczak J, Knapp M, et al. Local online information for carers in England: content and complexity. Practice 2023;35(7):1-25. [doi: 10.1080/09503153.2023.2232567]
10. Baxter K, Wilberforce M, Birks Y. What skills do older self-funders in England need to arrange and manage social care? Findings from a scoping review of the literature. Brit J Soc Work 2021;51(7):2703-2721. [doi: 10.1093/bjsw/bcaa102]
11. Adult Social Care Front Door: Lambeth. Local Government Association. 2019. URL: https://www.local.gov.uk/case-studies/adult-social-care-front-door-lambeth [accessed 2023-05-11]
12. Quigley R, Foster M, Harvey D, Ehrlich C. Entering into a system of care: a qualitative study of carers of older community-dwelling Australians. Health Soc Care Community 2022;30(1):319-329. [doi: 10.1111/hsc.13405] [Medline: 33955616]
13. Turnpenny A, Beadle-Brown J. Use of quality information in decision-making about health and social care services--a systematic review. Health Soc Care Community 2015;23(4):349-361. [doi: 10.1111/hsc.12133] [Medline: 25243842]
14. Laragy C, David C, Moran N. A framework for providing information in individualised funding programmes. Qual Soc Work 2015;15(2):190-208. [doi: 10.1177/1473325015589402]
15. Baxter K, Glendinning C. Making choices about support services: disabled adults' and older people's use of information. Health Soc Care Community 2011;19(3):272-279. [doi: 10.1111/j.1365-2524.2010.00979.x] [Medline: 21158999]
16. Abendstern M, Hughes J, Clarkson P, Tucker S, Challis D. Exploring the contribution of self-assessment to preventative services in social care. British Journal of Social Work 2012;44(3):729-746. [doi: 10.1093/bjsw/bcs144]
17. Clarkson P, Hughes J, Challis D, Thorley L, Kilshaw C. Targeting, care management and preventative services for older people: the cost-effectiveness of a pilot self-assessment approach in one local authority. Brit J Soc Work 2010;40(7):2255-2273. [doi: 10.1093/bjsw/bcq042]
18. Herd P. Improving older adults' health by reducing administrative burden. Milbank Q 2023;101(S1):507-531 [FREE Full text] [doi: 10.1111/1468-0009.12629] [Medline: 37096624]
19. Griffiths P. Self-assessment of health and social care needs by older people: a review. Br J Community Nurs 2005;10(11):520, 522-520, 527. [doi: 10.12968/bjcn.2005.10.11.19963] [Medline: 16301927]
20. Accessing adult social care in England. Office for National Statistics. 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/socialcare/methodologies/accessingadultsocialcareinengland [accessed 2023-05-24]

21. Care and Support Statutory Guidance. Department of Health. 2014. URL: https://www.gov.uk/government/publications/care-act-statutory-guidance/care-and-support-statutory-guidance#chapter-3 [accessed 2023-05-24]
22. Measures from the adult social care outcomes framework. NHS Digital. 2023. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-social-care-outcomes-framework-ascof [accessed 2024-12-10]
23. ADASS spring survey 2023: care waiting lists down but needs increasing. ADASS. 2023. URL: https://www.adass.org.uk/adass-spring-survey-2023-care-waiting-lists-down-but-needs-increasing#:~:text=A [accessed 2023-09-21]
24. Adult social care activity and finance report, England, 2021-22. NHS England. 2022. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-social-care-activity-and-finance-report/2021-22 [accessed 2022-12-12]
25. What the public want from social care. Healthwatch England. 2018. URL: https://www.healthwatch.co.uk/sites/healthwatch.co.uk/files/20180801 [accessed 2024-12-10]
26. Fernandez JL, Forder J. Local variability in long-term care services: local autonomy, exogenous influences and policy spillovers. Health Econ 2015;24 Suppl 1:146-157 [FREE Full text] [doi: 10.1002/hec.3151] [Medline: 25760589]
27. Gaínza-Lein M, Sánchez Fernández I, Jackson M, Abend NS, Arya R, Brenton JN, Pediatric Status Epilepticus Research Group. Association of time to treatment with short-term outcomes for pediatric patients with refractory convulsive status epilepticus. JAMA Neurol 2018;75(4):410-418 [FREE Full text] [doi: 10.1001/jamaneurol.2017.4382] [Medline: 29356811]
28. He Y, Yang Q, Alish Y, Ma S, Qiu Z, Chen J, et al. Relationship between product features and the prices of e-cigarette devices sold in web-based vape shops: comparison study using a linear regression model. JMIR Form Res 2024;8:e49276 [FREE Full text] [doi: 10.2196/49276] [Medline: 38723251]
29. Alin A. Multicollinearity. WIREs Computational Stats 2010;2(3):370-374. [doi: 10.1002/wics.84]
30. Chinn D, Homeyard C. Easy read and accessible information for people with intellectual disabilities: is it worth it? A meta-narrative literature review. Health Expect 2017;20(6):1189-1200 [FREE Full text] [doi: 10.1111/hex.12520] [Medline: 27862757]
31. Sutherland RJ, Isherwood T. The evidence for easy ‐ read for people with intellectual disabilities: a systematic literature review. Policy Practice Intel Disabi 2016;13(4):297-310. [doi: 10.1111/jppi.12201]
32. Wen W, Zhang Y, Shi W, Li J. Association between internet use and physical health, mental health, and subjective health in middle-aged and older adults: nationally representative cross-sectional survey in China. J Med Internet Res 2023;25:e40956 [FREE Full text] [doi: 10.2196/40956] [Medline: 36943368]
33. Heponiemi T, Kainiemi E, Virtanen L, Saukkonen P, Sainio P, Koponen P, et al. Predicting internet use and digital competence among older adults using performance tests of visual, physical, and cognitive functioning: longitudinal population-based study. J Med Internet Res 2023;25:e42287 [FREE Full text] [doi: 10.2196/42287] [Medline: 37145836]
34. Short and Long Term (SALT) Data Return 2021-22 guidance. NHS Digital. 2021. URL: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/social-care-collection-materials-2022/salt-data-return-2021-2022-guidance [accessed 2024-12-10]
35. Clarkson P, Brand C, Hughes J, Challis D, Tucker S, Abendstern M. Cost effectiveness of pilot self-assessment sites in community care services in England. Aust Health Rev 2013;37(5):666-674. [doi: 10.1071/AH13056] [Medline: 24160485]

Abbreviations

LTC: long-term care
NIHR: National Institute for Health and Care Research
OLLS: ongoing low-level support
OLS: ordinary least squares
SALT: short and long term
STC: short-term care

Edited by A Coristine; submitted 05.02.24; peer-reviewed by Y Xie, F Longo; comments to author 31.05.24; revised version received 25.07.24; accepted 10.09.24; published 02.01.25.

Please cite as:
Zhang J, Prunty JE, Charles AC, Forder J
Association Between Digital Front Doors and Social Care Use for Community-Dwelling Adults in England: Cross-Sectional Study
J Med Internet Res 2025;27:e53205
URL: https://www.jmir.org/2025/1/e53205
doi:10.2196/53205
PMID:

Original Paper

Efficacy and Acceptability of a Mobile App for Monitoring the Clinical Status of Patients With Chronic Obstructive Pulmonary Disease Receiving Home Oxygen Therapy: Randomized Controlled Trial

Anisbed Naranjo-Rojas[1,2*]; Luis Ángel Perula-de Torres[3*]; Freiser Eccehomo Cruz-Mosquera[4*]; Guillermo Molina-Recio[5*]

[1]Health and Education Research Group (GINEYSA), Faculty of Health, Universidad Santiago de Cali, Santiago de Cali, Colombia

[2]Biomedicine Doctoral Program, University of Córdoba, Córdoba, Spain

[3]Research Network on Chronicity Primary Care and Health Promotion (RICAPPS), Cooperative Research Networks Oriented to Health Results (RICORS) Carlos III Health Institute, Madrid, Spain

[4]Integral Health Research Group (GISI), Faculty of Health, Universidad Santiago de Cali, Santiago de Cali, Colombia

[5]Nursing Pharmacology and Physiotherapy Department University of Córdoba, Lifestyles Innovation and Health (GA-16) Maimonides Biomedical Research Institute of Córdoba (IMIBIC) Spain, University of Córdoba, Córdoba, Spain

[*]all authors contributed equally

Corresponding Author:
Anisbed Naranjo-Rojas
Health and Education Research Group (GINEYSA)
Faculty of Health
Universidad Santiago de Cali
Calle 5 # 62 esquina
Santiago de Cali, 763531
Colombia
Phone: 57 602 5183000 ext 6515
Email: anisbednaranjo24@gmail.com

Abstract

Background: Chronic obstructive pulmonary disease (COPD) primarily originates from exposure to tobacco smoke, although factors, such as air pollution and exposure to chemicals, also play a role. One of the primary treatments for COPD is oxygen therapy, which helps manage dyspnea and improve survival rates. Mobile health (mHealth) technologies have demonstrated significant potential in monitoring patients with chronic diseases, offering new avenues for enhancing patient care and disease management.

Objective: The purpose of this study was to evaluate the efficacy and acceptability of a mobile app designed for the clinical monitoring of patients with COPD and home oxygen (HO) therapy, compared with conventional monitoring in real-world community settings.

Methods: A parallel-group, nonblinded, multicenter randomized controlled trial was conducted with 45 participants; the intervention group (IG), which used the mobile app in addition to conventional monitoring (n=23) and the control group, which received only conventional monitoring (n=22), administered by therapists over a duration of 3 months. The primary outcomes included the chronic obstructive pulmonary disease assessment test (CAT) score, the level of dyspnea measured by the Borg scale, and oxygen saturation percentage, assessed at both the beginning and end of the trial. Secondary outcomes included the frequency of app use, the number of hospitalizations, and survival rates. In addition, a satisfaction survey and an interview were conducted with the IG.

Results: The median use of the mobile app was 21 (IQR 16-28) days. At the end of the follow-up, the Borg dyspnea scale was significantly lower in patients who used the mobile app for HO therapy monitoring (mean 0.6, SD 0.8 vs mean 4.1, SD 1.4; *P*=.001). Regarding the impact of COPD on quality of life, as measured by the CAT, no differences were found in the scores between baseline and end-of-follow-up within the control group. However, a significant decrease was observed in the IG (baseline median CAT 27, IQR 23-31 vs final median CAT 22, IQR 14-28; *P*<.001). In addition, the CAT score was significantly higher in patients receiving conventional monitoring compared with those monitored with the mobile app (median 30, IQR 23-32 vs median 22, IQR 14-28; *P*=.02).

Conclusions: The use of the mobile app, AppO2 (SINCO), designed for the clinical monitoring of patients with COPD and HO therapy, is associated with improved quality of life. In addition, the app is highly accepted by users, promotes self-care, and fosters patient confidence in managing their own condition.

Trial Registration: ClinicalTrials NCT04820790; https://clinicaltrials.gov/study/NCT04820790

International Registered Report Identifier (IRRID): RR2-https://doi.org/10.1186/s12875-021-01450-8

(J Med Internet Res 2025;27:e65888) doi:10.2196/65888

KEYWORDS

m-Health; chronic obstructive pulmonary disease; quality of life; mobile health applications; home oxygen therapy; mobile phone

Introduction

The main cause of chronic obstructive pulmonary disease (COPD) is exposure to tobacco smoke; however, there are also other associated factors such as air pollution, exposure to chemicals, and a history of respiratory diseases [1]. The World Health Organization considers it the third cause of death worldwide, making it a public health problem [2]. It is estimated that approximately 8% of the Colombian population—equivalent to about 3.6 million people—suffers from COPD [3,4]. Oxygen therapy is one of the main treatments for this disease [5], aiding in the control of symptoms such as dyspnea and improving survival [5]. Mesquita et al [6] found patients with COPD who did not adhere to oxygen therapy experienced a long-term decrease in quality of life compared to those who followed their prescribed treatment [7]. Similarly, several authors [6,8,9] have identified factors such as age, communication barriers between doctor and patient, clinical follow-up, and disease knowledge as significant challenges to adhering to oxygen therapy [6]. Low adherence can lead to exacerbations [10], which deteriorate lung function, negatively affect patients' quality of life, and increase mortality [11].

Furthermore, these exacerbations contribute to substantial health care costs. In Colombia, each exacerbation costs an average of US $98 [12], with costs rising to nearly US $700 if hospitalization is required [12]. In addition, this high cost can be attributed to the fact that once a patient is hospitalized, there is a high risk of rehospitalization or death. For example, Niewoehner [13] found that of 1016 patients hospitalized for COPD, some were rehospitalized and 33% (n=335) died within 6 months of discharge.

Conversely, mobile health (mHealth) technologies hold significant promise for monitoring patients with chronic diseases [14]. These apps offer benefits such as remote monitoring, direct communication with health care professionals, recording of clinical signs, and the development of self-care skills [14]. A Cochrane review [15] suggested that self-monitoring of symptoms through mobile apps positively impacts the development of self-care and self-management skills in patients with chronic diseases [9]. Furthermore, studies, such as the one conducted by Knox et al [16], have focused on designing mobile apps for patients with COPD, concluding that reporting clinical information through these apps can aid in disease management [16].

Some authors emphasize the importance of designing these mHealth apps with a user-centered approach [17,18]. Understanding the perceptions and needs of the end user as well as adapting to their sociodemographic characteristics and level of technological literacy are indispensable strategies for developing mHealth apps that increase adherence to use and, consequently, to treatment and self-care [17-19]. Therefore, evaluating the usability [20] of mHealth apps before their implementation is crucial. Conducting usability tests to assess user performance on specific tasks can identify areas for design improvements and enhance user satisfaction [4,21].

Based on the above, the authors of this study designed and developed a mobile app called AppO2 (SINCO) before conducting the research. This app was created using a user-centered design methodology and underwent usability tests [4,19]. AppO2 facilitates the monitoring of the clinical status of patients with COPD receiving home oxygen (HO) therapy, with 2 user profiles—patients (or their caregivers) and health care professionals [4,19]. The functions of the patient profile are centered on self-care skill development and quality of life improvement. These functions include tutorials on measuring and recording vital signs, accessing information related to the prescription of HO treatment, and communicating and interacting with health care professionals [4,19]. Conversely, the functions of the professional profile are designed to monitor patients' clinical status on a daily basis, record and review clinical changes to inform decision-making, and control the dosage of HO therapy [4,19].

Finally, the functionalities of this mobile app were designed to promote self-care in patients with COPD receiving HO therapy. It also sought to address the primary challenges these patients face in improving adherence to HO treatment, preventing or detecting exacerbations early, and enhancing their quality of life [4,19].

In this context, this study aimed at evaluating the efficacy and acceptability of AppO2 in real community settings compared with conventional home care monitoring for patients with COPD-prescribed HO.

Methods

Study Design and Participants

A 3-month, open-label, 2-arm, parallel-arm, multicenter, randomized controlled, nonblinded, clinical trial was conducted to determine the efficacy of a mobile app in monitoring the clinical status of patients with COPD receiving HO therapy. The control group (CG) comprised patients receiving HO and monitored using conventional methods, such as weekly home visits by a health care professional. The intervention group (IG) comprised patients who were also monitored through weekly home visits in addition to using the AppO2 mobile app. The clinical trial was conducted in Cali, Colombia, with the participation of 3 home care institutions—TodoMed, Amanecer Médico, and Cuidarte en Casa. The clinical trial protocol was registered in ClinicalTrials.gov (ID NCT04820790) [4].

Patients were recruited from May to December 2023 based on the following inclusion criteria: (1) being ≥18 years old, (2) having a medical prescription for HO for more than 1 year, (3) having a caregiver, (4) owning a smartphone, and (5) consenting to audio recording. The exclusion criterion was patients undergoing mechanical ventilation.

In total, 2 health care professionals were included based on the following inclusion criteria: (1) being ≥18 years old, (2) being respiratory therapists and physiotherapists affiliated with a home care company, (3) having more than 6 months of experience in home care, and (4) having a smartphone. The exclusion criterion was established as professionals with less than 6 months of experience in managing patients receiving HO.

Sample Size

To achieve a 1:1 ratio between the IG and CG, with 80% power (β error=20%) and 95% CI (α error=5%), and estimating that the AppO2 app in the IG would result in a 3% increase self-management of dyspnea, with SD 6 [22], a minimum sample of 32 participants (16 in each group) was estimated. Finally, to minimize the effects of potential patient losses that would reduce statistical power, the sample was expanded to a final size of 45 participants (23 in the IG and 22 in the CG) [4].

Randomization

Overview

Participants were equally and randomly assigned to either the IG or CG using a simple randomization method with Epidat 3.1 (Consellería de Sanidade de la Xunta de Galicia, in collaboration with the Pan American Health Organization [PAHO]). Blinding was not possible for participants or health care professionals owing to the intervention method used.

Sample Selection

Consecutive sampling was used to recruit patients and all participants who met the selection criteria and attended consultations at participating institutions were included in the study. Sampling was done until the predetermined sample size was reached.

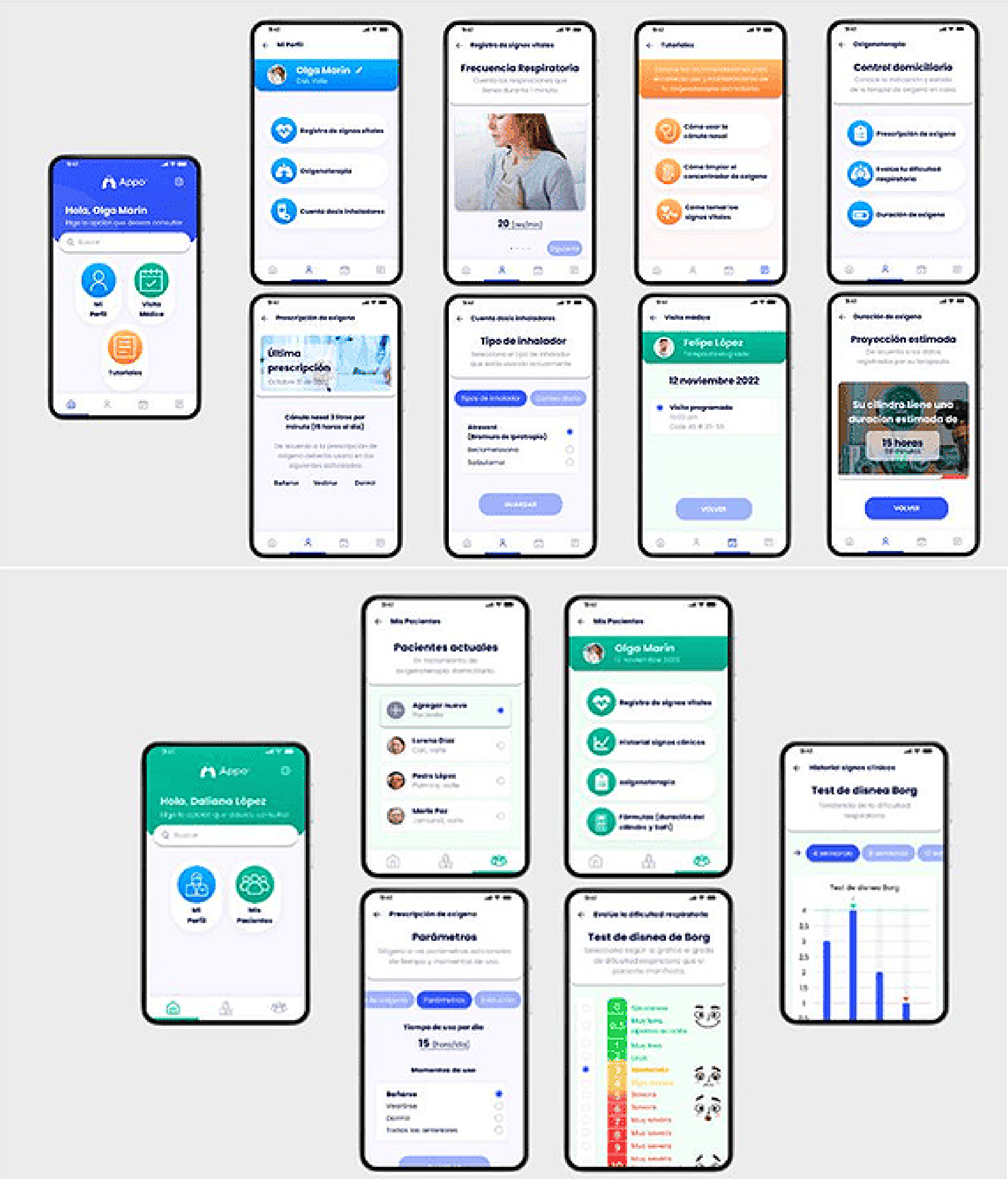
Intervention

After selecting the trial participants, the IG downloaded the AppO2 app on their mobile phones. The objectives of the investigation and the operation of the app were explained to them. In addition, it was recommended to use the app at least once a week. All recorded information was stored anonymously in a cloud-based database, identifiable by a code only known to the principal investigator. The database was accessible to researchers through a web app for review and analysis.

The AppO2 mobile app consists of 2 user profiles, one for patients and caregivers and another for professionals. There are 3 sections in the patient and caregiver profile, “My Profile,” “Medical Visit,” and “Tutorials.” In the “My Profile” section, patients can record their vital signs and the degree of respiratory difficulty, view the medical prescription for oxygen therapy including the duration of the oxygen therapy regimen, and record the number of daily inhalations if they have a prescription for medication. When recording their vital signs, patients from their user profile receive notifications informing them if vital signs are outside the predetermined ranges. These alerts suggest that they tell their health care professional to adjust treatment as necessary or recommend that they carefully follow medical indications if a slight alteration in vital signs is detected. However, these notifications are not sent to medical staff in real-time. Instead, professionals can view patients’ clinical history as they see fit through their profile [4,19]. This functionality allows them to monitor the behavior of vital signs and make the appropriate adjustments during scheduled home visits, accessible through the “Medical Visit” section. This approach allows for more flexible management and avoids an overload of unnecessary alerts, which could generate “false positives” by prioritizing intervention based on a comprehensive patient assessment. The “Medical Visit” section allowed users to view the schedule of upcoming medical visits and the history of past visits. In addition, the “Tutorials” section featured videos that provided education and instructions on taking vital signs and using oxygen systems [4,19].

Similarly, the professionals’ profile has 2 sections, “My profile,” where they can register their personal data and update their account, if necessary, and “My patients,” where they can record the assessment, prescribe oxygen, and monitor the clinical status of patients. They can also view graphs that display the monthly trends of each vital sign. Through the professionals’ profile, information on HO prescription parameters, visit schedules, and data on the estimated duration of oxygen cylinders are sent to the IG [4] (Figure 1).

Figure 1. AppO2 user profiles. (A) Patient and caregiver profile. (B) Professional profile.



Measures

Independent variables, including participants' sociodemographic characteristics (eg, age, sex, marital status, educational level, and place of residence) and other aspects of lifestyle and treatment (such as smoking, biomass exposure, and duration of HO treatment), were obtained from the databases and record systems of the 3 health institutions to which they belonged.

During the home visits, therapists collected information was collected on variables related to vital signs, such as oxygen saturation through pulse oximetry, respiratory rate, central heart rate, number of exacerbations, number of hospitalizations, number of emergency room admissions, survival, inhaler use, and changes in oxygen prescription. Cylinder duration time was calculated considering the conversion factor, oxygen system flow, residual pressure, and cylinder pressure [23]. This calculation was performed manually in the CG, whereas the mobile app was used in the IG. In addition, therapists reminded patients and caregivers to use the app during each visit. Vital signs and the relationship between oxygen saturation and fraction of inspired oxygen [24] were recorded once a week from the date of randomization until week 12 or until the time of death from any cause, whichever occurred first.

The degree of respiratory distress was determined using the Borg Dyspnea Scale [25], a standardized, validated analog scale in Spanish, widely used since the 1970s. The scale ranges from 0 to 10 and is designed to quickly and easily assess patients' perception of dyspnea. It includes a graph associated with each quantitative value, which helps patients identify their level of respiratory distress, with 10 indicating the highest level of perceived dyspnea [26]. Assessment using the Borg test was conducted from the beginning of the study until week 12 or until the date of death from any cause, whichever occurred first. In addition, the number of days the app was used, access to each screen, and connection time within the mobile app were tracked in the IG.

Questionnaires

The chronic obstructive pulmonary disease assessment test (CAT) questionnaire was used to assess the impact of COPD on patients' quality of life [27]. This questionnaire was administered to both the IG and CG at the beginning and at the end of the intervention. The acceptance of AppO2 was evaluated using the technology acceptance model (TAM) [28], which assessed the perceived usefulness and ease by participants in relation to AppO2 [28]. This questionnaire, along with an interview to gauge the perception of AppO2, was administered to health care professionals and the IG at the end of the follow-up period.

Chronic Obstructive Pulmonary Disease Assessment Test

This is a validated [29,30] and publicly accessible questionnaire consisting of 8 questions, each evaluated on a scale of 0-5 points. The minimum score is 0, and the maximum score is 40 [29,30], reflecting the impact COPD has on a patient's quality of life. The scores are classified as (1) low impact (1-10 points), where most days are "good days"; (2) medium impact (11-20 points), with few "good days"; (3) high impact (21-30 points), with no "good days"; and (4) very high impact (31-40 points), where the disease's limitation is at its maximum. A progressive increase in CAT scores indicates an increase in the impact of COPD on the patient's quality of life [27,29].

AppO2 Acceptance Questionnaire and Perception Interviews

The AppO2 mobile app acceptance questionnaire was administered to patients and professionals. It was based on the TAM [30] designed by Davis [31,32], aimed at evaluating people's perception of usefulness and ease regarding acceptance in the use of devices or software in digital environments. Although several variations have been made by different authors [33-35], it is essential to evaluate the following dimensions: (1) perception of usefulness and (2) perception of ease. Based on these dimensions, 12 items translated and validated into Spanish [36,37] were applied, with minor adaptations in their content for better understanding. Responses were assessed using a Likert scale, ranging from "totally agree" to "totally disagree" (Multimedia Appendix 1). To determine the acceptance of the AppO2 mobile app, 1 point was assigned to the item with which participants were moderately or totally in agreement, and 0 was assigned to items with which they were moderately or totally in disagreement. Cutoff points for the level of acceptance were obtained by measuring quartiles. Low acceptance was considered if it was in quartile 1 (1-3 points), moderate if it was between quartiles 2 and 3 (4-9 points), and high if it was in quartile 4 (10-11 points). Furthermore, to complete the evaluation of participants' perceptions regarding the AppO2 app, an interview was conducted with both the IG and health care professionals. The objective of the questions was to inquire about what they liked or disliked most about the app, the perceived benefits, and their preferences. These interviews were conducted at the end of the intervention (Multimedia Appendix 2).

Primary and Secondary Outcomes

Primary outcomes consisted of the CAT score, the degree of dyspnea measured using the Borg test, and the percentage of oxygen saturation at the beginning and end of the clinical trial. Secondary outcomes were the frequency of use of the app, number of hospitalizations, and survival.

Statistical Analysis

Data analysis was performed using SPSS Statistics software, version 25.0 (IBM Corp). Initially, a descriptive analysis of the study variables was conducted, characterizing the patients and generating graphs and tables of absolute and relative frequencies for qualitative variables. For quantitative variables, measures of central tendency and position (mean and median) with their respective measures of dispersion (SD and IQR) were calculated.

In the bivariate analysis, the Student *t* test or its nonparametric equivalent (Mann-Whitney *U* test) was used to compare 2 means. For categorical variables, the chi-square test was used. To analyze 3 or more mean values, repeated measures analysis of variance or the Friedman test was used. Crude odds ratios (ORs) were initially calculated to determine variables independently associated (AppO2 use and sex) with the impact of COPD on quality of life, as measured by the CAT. In addition, an analysis of covariance (ANCOVA) was used, which allowed for adjustment of differences in the final Borg dyspnea scores based on the initial baseline score. This ensured that observed changes were attributed to the intervention with the app rather than initial variations in dyspnea.

Finally, a binary logistic regression model was applied, including variables with a *P* value <.20 in the initial analysis presented or those related to the outcome by biological plausibility. This allowed for estimating the standardized β coefficients, an adjusted coefficient of determination, and the residual values. The binary logistic regression analysis dichotomized the CAT outcome into "low" (patients with a low or medium CAT score) and "high" (patients with a high or very high CAT score). The model was adjusted for possible confounding variables (time of oxygen use and years of cigarette consumption). An α error of ≤0.05 was considered statistically significant, and 95% CIs were calculated.

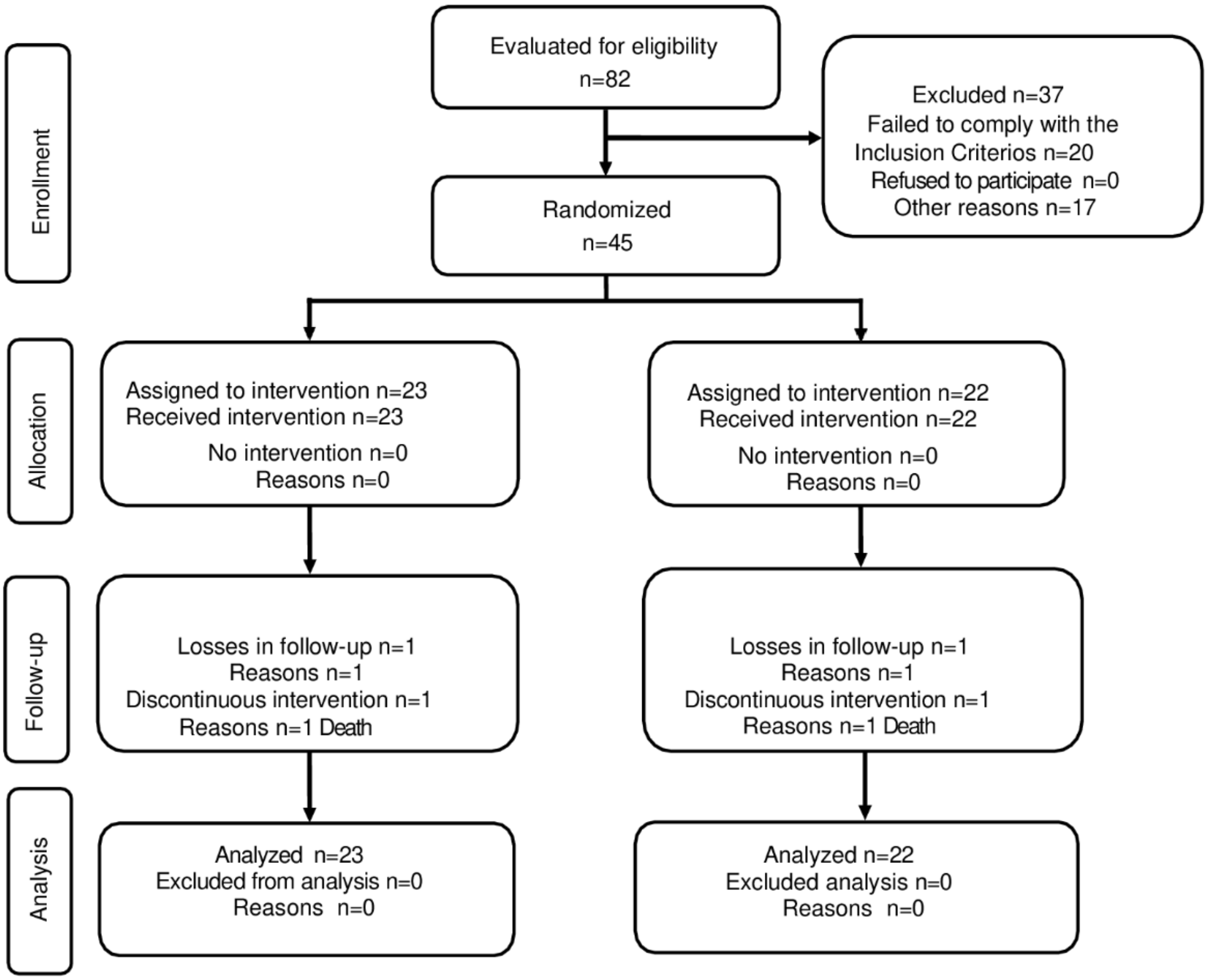
Ethical Considerations

The clinical trial protocol was approved by the Ethics Committee of the Department of Health of Universidad Santiago de Cali, Colombia, Act No. 02, as well as by the 3 home care institutions in the city of Cali, Colombia. The investigation complies with international regulations, such as the Declaration of Helsinki. All participants gave their written informed consent before inclusion in the study. Informed consent was obtained, and the participants were informed of the possibility of opting out or withdrawing at the time of their choice, if applicable. The data were anonymized so that participants could not be identified, and no amount of compensation was offered or awarded.

Results

Participants were equally and randomly assigned to either the IG or CG using a simple randomization method with Epidat 3.1 (Consellería de Sanidade de la Xunta de Galicia, in collaboration with the Pan American Health Organization [PAHO]). Blinding was not possible for participants or health care professionals owing to the intervention method used (Figure 2). The CONSORT Checklist is included as Multimedia Appendix 3.

Figure 2. Modified CONSORT (Consolidated Standards of Reporting Trials) flow diagram for individual randomized controlled trials of nonpharmacological treatments.



A total of 45 participants were included (23 from the IG and 22 from the CG). The cohort was predominantly female (n=31, 69%) with a mean age of 75.4 (SD 15.9) years (95% CI 70.7-80.0 years). Furthermore, 73% (n=33) of the participants had been exposed to biomass at some point in their lives, and 93% (n=42) had a history of smoking, with a time of consumption that exceeded 15 years in most cases (n=30, 67%). When comparing baseline characteristics between the groups, no statistically significant differences were observed in sociodemographic variables and background except for the “time of oxygen use for more than 15 years,” which was greater in the IG (n=17, 74% vs n=9, 41%, *P*=.02; Table 1).

Table 1. Baseline characteristics of patients in the control and intervention groups enrolled in the study (n=45).

Variable	Intervention (n=23)	Control (n=22)	*P* value
Sex, n (%)			.50[a]
Male	8 (35)	6 (27)	
Female	15 (65)	16 (73)	
Education level, n (%)			.50[a]
Primary	18 (78)	16 (73)	
Secondary	5 (22)	5 (23)	
University	0 (0)	1 (4.5)	
AppO2 usage time, n (%)			<.001[a]
More than 15 hours	17 (74)	9 (41)	
Less than 15 hours	6 (26)	13 (59)	
Activities using AppO2, n (%)			.20[a]
Bathing	0 (0)	0 (0)	
Dressing	2 (9)	0 (0)	
Sleeping	7 (30.4)	4 (18)	
Moving around	7 (30.4)	7 (32)	
All of the above	7 (30.4)	11 (50)	
Biomass exposure, n (%)			.90[a]
Yes	17 (74)	16 (73)	
No	6 (26)	6 (27)	
Smoking history, n (%)			.08[b]
Yes	20 (87)	22 (100)	
No	3 (13)	0 (0)	
Years of smoking, n (%)			.30[a]
15 years or less	9 (39)	6 (27)	
15 years or more	14 (61)	16 (73)	
Institution, n (%)			<.001[a]
TodoMed	10 (43.5)	11 (50)	
Amanecer médico	3 (13)	9 (41)	
Cuidarte en casa	10 (43.5)	2 (9)	
Age (years), mean (SD)	72.1 (18.2)	78.9 (13)	.15[c]
Baseline Borg score, mean (SD)	1.2 (0.91)	4.3 (1.4)	<.001[c]
Baseline COPD[d] assessment test score, mean (SD)	26.3 (8.2)	28.9 (6.9)	.28[c]

[a] Chi-square.

[b] Continuity correction.

[c] Mann-Whitney *U* test.

[d] COPD: chronic obstructive pulmonary disease.

Regarding the behavior of vital signs during the 3-month follow-up for patients with conventional follow-up, no significant differences were observed in respiratory rate, heart rate, temperature, saturation, and Borg Dyspnea Scale scores. Likewise, in the IG, no statistically significant differences were found in vital signs, except in the Borg Dyspnea Scale score, which decreased over time (month 1: mean 1.2, SD 0.9; month 2: mean 1.0, SD 0.9; month 3: mean 0.6, SD 0.8; *P*=.01) (Figure 3).

When comparing the clinical variables at the end of follow-up between the groups, it was found that the Borg Dyspnea Scale score was significantly lower in the IG than in the CG (mean 0.6, SD 0.8 vs mean 4.1, SD 1.4; *P*=.001; Table 2).

After adjustment for baseline values, significant differences were maintained in both Borg Dyspnea Scale and CAT final scores, favoring the intervention group, with lower adjusted means (1.3 vs 2.7; *P*=.02 and 22.33 vs 26.69; *P*=.002, respectively; Table 3).

Figure 3. Clinical behavior during the follow-up of patients included in the study (n=45).

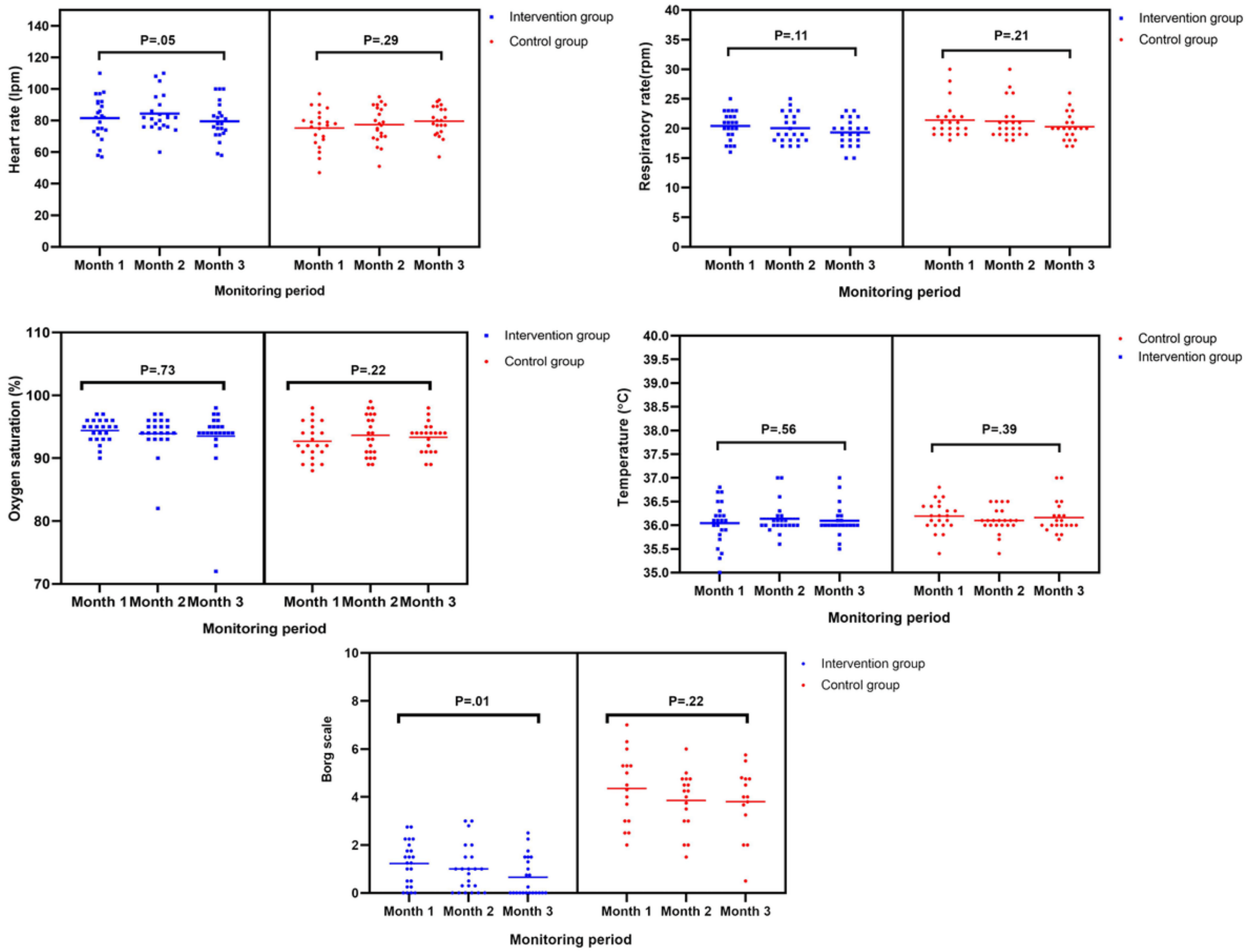


Table 2. Comparison of clinical variables of patients in the intervention and control groups (n=45).

Variable	Intervention (n=23), mean (SD)	Control (n=22), mean (SD)	95% CI	*P* value	Intervention (n=23), mean difference[a] (SD)	Control (n=22), mean difference[a] (SD)	*P* value[b]
Heart rate	79.5 (11.8)	79 (9.1)	–5.6 to 6.6	.87	–3.1 (2.19)	4 (2.80)	.10
Respiratory frequency	19 (2.3)	20 (2.3)	–2.3 to 0.3	.15	–1 (0.71)	–1 (0.71)	.99
Oxygen saturation	93 (5)	93 (2.3)	–2.2 to 2.2	.95	–1 (0.71)	–1 (0.71)	.99
Temperature	36.0 (0.3)	36.1 (0.3)	–027 to 0.07	.30	0 (0)	–0.1 (–0.07)	__[c]
Borg score	0.6 (0.8)	4.1 (1.4)	–4.17 to –2.83	.001	–0.6 (0.42)	–0.8 (0.56)	.72

[a] Mean difference: mean at the end−mean at the beginning of follow-up.

[b] Student *t* test.

[c] Not applicable.

Table 3. Analysis of covariance of the impact of the use of AppO2 on Borg Dyspnea Scale and chronic obstructive pulmonary disease assessment test scores at the end of follow-up, monitoring for the baseline levels of dyspnea.

	Mean (SD)	95% CI	*P* value[a]	Mean (SD)	95% CI	*P* value[b]
Borg Dyspnea Scale[c]			<.001			.02
Intervention group (n=23)	0.65 (0.8)	0.29-1.01		1.3 (1.2)	0.77-1.84	
Control group (n=14)	3.8 (1.4)	2.96-4.64		2.7 (1.6)	1.83-3.60	
COPD[d] Assessment Test (CAT)			.01			.002
Intervention group (n=23)	21.41 (9.01)	17.41-25.41		22.33 (4.27)	20.49-24.17	
Control group (n=14)	27.67 (7.17)	24.4-30.93		26.69 (4.2)	24.80-28.50	

[a] Student *t* test.

[b] Analysis of covariance (ANCOVA).

[c] R^2 adjusted for baseline Borg score: 0.766.

[d] COPD: chronic obstructive pulmonary disease.

Regarding the impact of COPD on quality of life measured with the CAT questionnaire, although no differences were found in the score between the baseline measurement and at the end of follow-up in the control group, a significant decrease was observed in the intervention group (median baseline CAT 27, IQR 23-31 vs median final CAT 22, IQR 14-28; *P*<.001; Figure 4). In addition, the CAT score was significantly higher in patients undergoing conventional follow-up compared with those followed with the mobile app (median 30, IQR 23-32 vs median 22 IQR 14-28; *P*=.02).

When evaluating the record of HO therapy prescriptions in each group, it was found that the proportion of individuals who performed the HO therapy prescriptions correctly was higher in the IG throughout each month of follow-up. The frequency of accessing tutorials or educational records was significantly higher in the IG (95.7% vs 59%; *P*=.003) in the first month of follow-up (Table 4). Furthermore, during the study period, 50% (n=11) of participants in the IG used the mobile app for more than 21 days (median 21, IQR 16-28). The median was greater than 14 days for most variables requiring frequent recording or entry: vital signs recording (median 19, IQR 15-27), Borg scale recording (median 16, IQR 14-22), and oxygen prescription recording (median 14, IQR 11-18; Multimedia Appendix 4).

Satisfaction was mostly observed among users of the AppO2 app, including patients, caregivers, and therapists. Details regarding the measurement of satisfaction with the AppO2 app are presented in Multimedia Appendix 5.

Similarly, in interviews assessing the perception and acceptance of technology using the TAM, the IG reported that the use of the mobile app fostered confidence and improved communication with health care professionals. The group emphasized the value of notifications regarding their vital signs, noting that having access to this information made them more responsible in managing and caring for their disease. Multimedia Appendix 6 presents some expressions reflecting the perceptions and acceptance of AppO2.

Figure 4. Comparison of chronic obstructive pulmonary disease assessment test scores at baseline and at the end of follow-up for control and intervention group patients included in the study. CAT: chronic obstructive pulmonary disease assessment test.

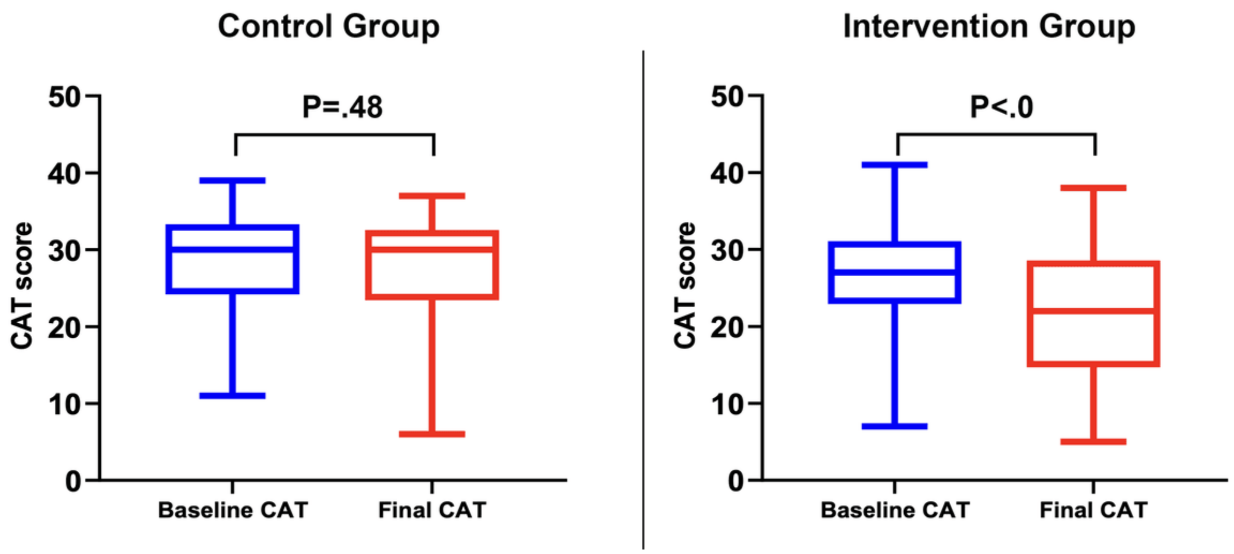


Table 4. Comparison of oxygen therapy recording, education, and oxygen cylinder duration in the control group versus the intervention group.

Variable	Intervention (n=23), n (%)	Control (n=22), n (%)	*P* value
Records oxygen prescription			
Month 1	22 (96)	14 (64)	.006
Month 2	22 (96)	11 (50)	.002
Month 3	22 (96)	10 (48)	.001
Records education or reviews tutorials			
Month 1	22 (96)	13 (59)	.003
Month 2	14 (61)	9 (41)	.18
Month 3	4 (17)	14 (67)	.002
Records duration of oxygen cylinder			
Month 1	6 (26)	2 (9)	.13
Month 2	4 (17)	3 (14)	.72
Month 3	5 (17)	7 (33.)	.44

Finally, the binary logistic regression model was statistically significant (chi-square value=5.9, *P*<.001) and demonstrated a good fit (Hosmer-Lemeshow test=3.92, *P*=.80). The use of the mobile app was the only variable independently associated with the impact on the quality of life of patients with COPD, as measured using the CAT questionnaire (adjusted OR 0.15; 95% CI 0.02-0.82; Table 5).

Table 5. Use of AppO2 and its association with decreased impact of chronic obstructive pulmonary disease on quality of life measured by chronic obstructive pulmonary disease assessment test.

Variable	Crude OR[a] (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
App usage				
Yes	0.15 (0.02-0.80)	.02	0.15 (0.02-0.82)	.02
No	6.57 (1.21-35.5)	.02	6.57 (1.21-35.52)	.02
Sex				
Male	0.79 (0.18-3.30)	.75	1.09 (0.18-6.38)	.92
Female	1.25 (0.29-5.20)	.75	0.91 (0.15-5.34)	.92
Oxygen use duration				
More than 15 hours	0.73 (0.17-3.00)	.66	0.58 (0.10-3.28)	.53
Less than 15 hours	1.36 (0.33-5.50)	.66	1.72 (0.30-9.70)	.53
Smoking duration				
More than 15 years	1.6 (0.35-7.20)	.54	—[b]	—
Less than 15 years	0.62 (0.13-2.82)	.54	—	—
Smoking and biomass exposure				
Yes	2.1 (0.51-8.70)	.29	—	—
No	0.46 (0.11-1.90)	.29	—	—

[a] OR: odds ratio.

[b] Not applicable.

Discussion

Overview

A 2-arm randomized clinical trial was conducted to determine the effectiveness and acceptability of the AppO2 mobile app for managing HO therapy in patients with COPD in Cali, Colombia.

The use of health technology tools that offer clinical interventions is emerging as a promising strategy in the health care setting [38,39]. In recent years, self-management interventions supported by smartphone apps have been shown to improve disease management in chronic patients, thus decreasing hospitalizations and emergency room visits and improving quality of life [40].

Quality of Life of Patients With COPD Receiving Home Oxygen When Using AppO2

The main finding of this investigation is that the use of the AppO2 app is associated with a decrease in the score of the impact of COPD on quality of life. This reduction in the CAT score suggests a lower negative impact of the disease on the quality of life of patients and an improvement in their general well-being. We believe that this finding is related to the user-centered design approach of AppO2 [19], which addressed specific needs expressed by patients and their caregivers during the initial stages of the project. Therefore, this app is not only limited to being a clinical monitoring tool, but it also promotes the development of self-care skills and recognition of clinical signs. Wang et al [41], who also used the CAT to assess the impact on patient's quality of life, observed significant improvements in patients who used a mobile app compared with those who received conventional medical care for COPD.

Another important aspect is the improvement in the dyspnea score according to the Borg scale observed in the IG. Participants mentioned that the information provided by AppO2 through images and notifications when determining the degree of dyspnea allowed them to establish a direct relationship with their perception of effort and respiratory difficulty. They also stated that this facilitated communication with therapists through the mobile app and during home visits. These findings are consistent with those reported by Kayyali et al [42], who demonstrated that patients who received clinical information through technological tools, such as mobile apps, could establish correlations with their symptoms and determine when to contact health care professionals.

A Cochrane review [43] suggested that digital tools in health care might improve patients' quality of life, although long-term effects remain unclear. Prolonged use of these digital interventions could improve symptoms such as dyspnea. Our findings align with studies [44,45] that indicate mHealth apps enhance patients' quality of life.

Health Perception and Self-Care of Patients With COPD Receiving Home Oxygen When Interacting With AppO2

Our study found no significant differences between the groups regarding the recording of the clinical signs such as oxygen saturation, respiratory rate, and heart rate. This might be due to the chronic condition [38,40] of the patients, which did not show notable clinical changes during follow-up. However, users in the IG reported that AppO2 improved their health and confidence in recognizing and interpreting vital signs. This suggests that the usefulness of the mobile app extends beyond clinical sign monitoring, positively impacting patients' self-care and their perception of the app's use, as supported by other research on mHealth technologies [42,46,47].

Another point worth mentioning is the interaction of the patients with the oxygen prescription section. The few patients in the IG functionality found it helpful for managing their treatment autonomously by viewing their oxygen prescription at any time, facilitating the use of the HO. Despite previous positive feedback on this feature [4,19], its low usage in this study might warrant its reconsideration or removal. Conversely, the CG only had access to this information during home visits from health care professionals.

Furthermore, patients who used AppO2 mentioned feeling more responsible and confident in managing their disease, emphasizing that access to clinical information aided them in adhering to guidelines for monitoring their clinical signs. This finding aligns with previous investigations [43,47], which indicate that patients with chronic diseases consider the use of digital tools for self-care as a good strategy to access clinical information and optimize disease management.

Acceptability and Connection Time With AppO2

According to the results of the level of acceptability measured using the TAM tool, most patients expressed being completely satisfied with the usefulness and ease of use of AppO2. In addition, patients in the IG expressed feeling safe when using the app. These results are consistent with those of the investigation conducted by Knox et al [16], who designed an app for patients with COPD also using the TAM. They observed that patients demonstrated high motivation and a positive attitude toward using the app, especially in making health-related decisions [16].

On the other hand, while the average use of the app for 21 days over a 3-month period may seem small, it is important to consider that its design does not require continuous daily use to be effective. The app focuses on building patient confidence and promoting self-awareness of their symptoms. Through intermittent but strategic use, patients acquire key tools to monitor and manage their condition more autonomously, potentially contributing to a better quality of life. On the other hand, patients mentioned in interviews that they felt more secure and confident knowing that therapists could access their records at any time, assess their clinical condition, and adjust treatment if necessary. This sense of constant accompaniment helped them to feel "less alone" in the process of their treatment, enhancing their mental well-being.

Likewise, some authors have indicated that certain factors may influence the frequency of AppO2 use. These factors highlight the acquisition of skills and knowledge for self-reporting and managing their disease, which decreases the need to consult tutorials or informative sections. As patients become more familiar with these skills, they experience greater autonomy, leading to more selective use of the app's functions, focusing on those they consider most relevant to their daily routine [48,49].

In this sense, the trend of AppO2 use is reflected in the maximum period of disconnection of 2 weeks observed among the participants, demonstrating a significant engagement with the app. Patients expressed greater confidence and education in recognizing their vital signs. Thus, effective use of AppO2 is not only measured by frequency of access but also by its ability to foster self-management and clinical monitoring.

In previous studies [50-52], it has been documented that adherence to mHealth apps tends to decline over time, especially after an initial phase of intensive use. Perski et al [51] highlight that this decline in user engagement may influence clinical outcomes. However, they also identify factors that may prolong use, such as personalization and regular reminders. Similarly, Byambasuren et al [52] noted that the "novelty" of mobile apps often wears off after the first few months. However, they suggest that intuitive design [53] and ongoing support from health professionals may improve long-term user retention.

In our study, the median number of app logins was 21 days during the 3-month follow-up period. In addition, we observed that the maximum time offline was 10 days. These data reflect a pattern of use that aligns with the app's design and does not require continuous daily access to be effective. The design's simplicity [53], vital sign notifications, and ease of use have been key factors in maintaining patient engagement.

As mentioned earlier, access to the tutorials in the IG decreased by the third month. The authors attribute this finding to the fact that the content of the videos did not vary over time, which may have led patients to view the education provided as adequate and integrated into their activities. This allowed them to be more selective in choosing the sections that best suited their needs and daily routine. In this sense, it is crucial to explore reinforcement strategies in future studies to ensure continued app use and maximize its clinical impact. This emphasis on continuous improvement in app usage is essential for the long-term success of such interventions [54,55].

Finally, the IG showed a maximum disconnection period of 2 weeks, indicating that the patients continuously interacted with the app. This adherence may be attributed to the high usability of AppO2, achieved through the end-user-centered design methodology applied in this research [4,18,19]. Conversely, the viewing of tutorials decreased in the final month of follow-up in the IG, which could be attributed to patients becoming more adept at self-monitoring and recognizing their vital signs. Similarly, Ding et al [56] reported that the use of the mobile app was more frequent during the first week of launch but decreased over time. Similarly, previous research [48,49] observed that users stop consulting some functions of mobile apps over time. Some factors that may influence this are the acquisition of skills and knowledge for self-registration and management of their disease, which reduces the need to consult tutorials or information sections. The continuous use of the app can also lead to a more selective use of its functions, focusing on those that are essential in their daily routine [48,49].

Principal Results

The findings of this study indicate that the use of the "AppO2" mobile app is associated with a significant improvement in the quality of life for patients with COPD and HO therapy. The high acceptability of the app is evidenced by a notable degree of satisfaction and adherence among users. Furthermore, "AppO2" not only facilitates self-care by enabling patients to manage their treatment more autonomously but also enhances their confidence in self-managing their condition. This contributes to better adaptation to treatment and greater self-care capabilities.

Limitations

The sample size comprised 45 participants. To avoid a lack of statistical power, the study included 3 home care institutions, which helped increase participant representativeness and the generalizability of the results for home health programs.

A limitation of the study is the 3-month duration. Although this period was sufficient to record favorable changes in the CAT score, which measures the impact of COPD on patients' quality of life, as well as their self-care and health perceptions, a longer follow-up could provide insights into the long-term effects.

The ANCOVA results indicated that the use of the app reduced dyspnea, as measured by the Borg scale, at month 3. However, the statistical significance of the initial Borg score (Borg month 1) suggests that the initial levels of dyspnea influenced the final results. The influence of the covariate suggests that other unmeasured variables, such as disease severity or comorbidities, may have affected the results, although this factor was controlled for in the analysis. Similarly, another possible limitation is the Hawthorne effect or observation bias, which occurs when study participants can systematically alter their behavior when they know they are being observed or monitored [50]. This, along with the inability to blind participants, may have influenced participant behavior. Future research should consider more discreet observation methods to minimize these biases.

Conclusions

The use of AppO2 is associated with a better quality of life in patients with COPD receiving HO therapy. The acceptability results for AppO2 show a high degree of satisfaction and adherence to its use. In addition, this mobile app promotes self-care and allows patients to develop confidence in managing their disease.

Acknowledgments

This research has been funded by the Dirección General de Investigaciones of Universidad Santiago de Cali (call 01-2024). This research has also been supported by the Universidad Santiago de Cali (research project 442-621121-3169). We would also like to thank all the patients and their families, as well as the professionals and institutions that participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adaptation of the AppO2 acceptance questionnaire (dimension and item).

[PDF File (Adobe PDF File), 36 KB - jmir_v27i1e65888_app1.pdf]

Multimedia Appendix 2
Questions on the perception of the AppO2 mobile app.
[PDF File (Adobe PDF File), 37 KB - jmir_v27i1e65888_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1170 KB - jmir_v27i1e65888_app3.pdf]

Multimedia Appendix 4
Use of the mobile app by patients in the intervention group (n=23).
[PDF File (Adobe PDF File), 30 KB - jmir_v27i1e65888_app4.pdf]

Multimedia Appendix 5
Satisfaction of patients, caregivers, and therapists using the mobile app for home oxygen therapy monitoring (n=24). (A) Perceived usefulness. (B) Perceived ease of use. Question 1: Using this app helps me complete my tasks more quickly. Question 2: Using this app enhances my performance. Question 3: Using this app increases my productivity. Question 4: Using this app enhances the effectiveness of my work or self-care. Question 5: Using this app makes it easier for me to perform my work or self-care. Question 6: I find this app very useful for my work or self-care. Question 7: Learning to use this app was easy for me. Question 8: I find this app easy to do what I need to do. Question 9: My interaction with this app was clear and understandable. Question 10: I find this app flexible to interact with. Question 11: It will be easy for me to become an expert in using the app. Question 12: I find this app easy to use.
[PNG File , 83 KB - jmir_v27i1e65888_app5.png]

Multimedia Appendix 6
Technology Acceptance Model and AppO2 perception interviews.
[PDF File (Adobe PDF File), 86 KB - jmir_v27i1e65888_app6.pdf]

References

1. Donaldson GC, Seemungal TAR, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 2002;57(10):847-852 [FREE Full text] [doi: 10.1136/thorax.57.10.847] [Medline: 12324669]
2. Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL, Held LS, et al. Chronic obstructive pulmonary disease: current burden and future projections. Eur Respir J 2006;27(2):397-412 [FREE Full text] [doi: 10.1183/09031936.06.00025805] [Medline: 16452599]
3. Caballero A, Torres-Duque CA, Jaramillo C, Bolívar F, Sanabria F, Osorio P, et al. Prevalence of COPD in five Colombian cities situated at low, medium, and high altitude (PREPOCOL study). Chest 2008;133(2):343-349. [doi: 10.1378/chest.07-1361] [Medline: 17951621]
4. Naranjo-Rojas A, Perula-de-Torres L, Cruz-Mosquera FE, Molina-Recio G. Mobile application for monitoring patients under home oxygen therapy: a protocol for a randomized controlled trial. BMC Fam Pract 2021;22(1):104 [FREE Full text] [doi: 10.1186/s12875-021-01450-8] [Medline: 34039277]
5. Spielmanns M, Boeselt T, Huber S, Kaur Bollinger P, Ulm B, Peckaka-Egli A, et al. Impact of a smartphone application (KAIA COPD app) in combination with activity monitoring as a maintenance program following pulmonary rehabilitation in COPD: the protocol for the AMOPUR Study, an international, multicenter, parallel group, randomized, controlled study. Trials 2020;21(1):636 [FREE Full text] [doi: 10.1186/s13063-020-04538-1] [Medline: 32653025]
6. Mesquita CB, Knaut C, Caram LMDO, Ferrari R, Bazan SGZ, Godoy I, et al. Impact of adherence to long-term oxygen therapy on patients with COPD and exertional hypoxemia followed for one year. J Bras Pneumol 2018;44(5):390-397 [FREE Full text] [doi: 10.1590/S1806-37562017000000019] [Medline: 30517340]
7. Jones I, Kirby A, Ormiston P, Loomba Y, Chan K, Rout J, et al. The needs of patients dying of chronic obstructive pulmonary disease in the community. Fam Pract 2004;21(3):310-313. [doi: 10.1093/fampra/cmh317] [Medline: 15128695]
8. Marcondes VK, Kuwazuru TS, Silva LPCE, Cezare TJ, Franco EAT, Prudente R, et al. Evaluation of the association of adherence to long-term home oxygen therapy and clinical markers and five-year mortality in patients with Chronic obstructive pulmonary disease. J Bras Pneumol 2020;46(6):e20190158 [FREE Full text] [doi: 10.36416/1806-3756/e20190158] [Medline: 32844890]

9. Gauthier A, Bernard S, Bernard E, Simard S, Maltais F, Lacasse Y. Adherence to long-term oxygen therapy in patients with chronic obstructive pulmonary disease. Chron Respir Dis 2019;16:1479972318767724 [FREE Full text] [doi: 10.1177/1479972318767724] [Medline: 29631423]
10. Wedzicha JA, Donaldson GC. Exacerbations of chronic obstructive pulmonary disease. Respir Care 2003;48(12):1204-13; discussion 1213 [FREE Full text] [Medline: 14651761]
11. Casanova C, Cote C, Marin JM, Pinto-Plata V, de Torres JP, Aguirre-Jaíme A, et al. Distance and oxygen desaturation during the 6-min walk test as predictors of long-term mortality in patients with COPD. Chest 2008;134(4):746-752. [doi: 10.1378/chest.08-0520] [Medline: 18625667]
12. Botero Mesa S, Restrepo DA. Conceptos esenciales de la EPOC, prevalencia e impacto en América Latina. Medicina UPB. 2015. URL: https://revistas.upb.edu.co/index.php/medicina/article/view/1379 [accessed 2024-11-20]
13. Niewoehner DE. The impact of severe exacerbations on quality of life and the clinical course of chronic obstructive pulmonary disease. Am J Med 2006;119(10 Suppl 1):38-45. [doi: 10.1016/j.amjmed.2006.08.006] [Medline: 16996898]
14. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012;12(12):CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
15. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;2013(3):CD008776 [FREE Full text] [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
16. Knox L, Gemine R, Rees S, Bowen S, Groom P, Taylor D, et al. Using the technology acceptance model to conceptualise experiences of the usability and acceptability of a self-management app (COPD.Pal®) for chronic obstructive pulmonary disease. Health Technol (Berl) 2021;11(1):111-117 [FREE Full text] [doi: 10.1007/s12553-020-00494-7] [Medline: 33262925]
17. Turchioe MR, Heitkemper EM, Lor M, Burgermaster M, Mamykina L. Designing for engagement with self-monitoring: a user-centered approach with low-income, latino adults with type 2 diabetes. Int J Med Inform 2019;130:103941 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.08.001] [Medline: 31437618]
18. Molina-Recio G, Molina-Luque R, Romero-Saldaña M. The importance of knowing and listening to all those involved in the design and use of nutrition mobile apps. Getting to know the great GApp. Nutr Hosp 2021;38(3):555-562 [FREE Full text] [doi: 10.20960/nh.03385] [Medline: 33813833]
19. Naranjo-Rojas A, Perula-de-Torres L, Molina-Recio G. Patients, caregivers, and healthcare professionals' needs when designing the content of a mobile application for the clinical monitoring of patients with chronic obstructive pulmonary disease and home oxygen therapy: a user-centered design. Internet Interv 2022;29:100552 [FREE Full text] [doi: 10.1016/j.invent.2022.100552] [Medline: 35756891]
20. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. New York, New York, USA: ACM Press; 1993 Presented at: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; 1993 May 01; New York, NY, United States p. 206-213.
21. Kruis AL, Smidt N, Assendelft WJJ, Gussekloo J, Boland MRS, Rutten-van Mölken M, et al. Integrated disease management interventions for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2013(10):CD009437. [doi: 10.1002/14651858.CD009437.pub2] [Medline: 24108523]
22. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
23. Noseda A. Dyspnoea and perception of airway obstruction. Rev Mal Respir 2003;20(3 Pt 1):364-372. [Medline: 12910111]
24. Salas JJR. Correlation of PaO2/FiO2 and SpO2/FiO2 indices in the postoperative period of cardiac surgery in a cardiovascular postsurgical therapy unit. Critical Medicine 2013;27(2):71-76 [FREE Full text]
25. Kendrick KR, Baxi SC, Smith RM. Usefulness of the modified 0-10 Borg scale in assessing the degree of dyspnea in patients with COPD and asthma. J Emerg Nurs 2000;26(3):216-222. [doi: 10.1016/s0099-1767(00)90093-x] [Medline: 10839848]
26. de Freitas APVM, Belo LF, Martinez L, Hernandes NA, Pitta F. Use of the borg dyspnea scale to identify dynamic hyperinflation during the 6-minute walking test in individuals with moderate-severe COPD: A pilot study. Pulmonology 2023;29(4):335-337 [FREE Full text] [doi: 10.1016/j.pulmoe.2023.02.003] [Medline: 36907815]
27. Jimenez-Ruiz CA, Pascual Lledó JF, Cícero Guerrero A, Cristóbal Fernández M, Mayayo Ulibarri M, Villar Laguna C. Analysis of quality of life in patients with chronic obstructive pulmonary disorder (COPD) who give up smoking. Semergen 2018;44(5):310-315. [doi: 10.1016/j.semerg.2017.08.003] [Medline: 28918179]
28. Rocha V, Jácome C, Martins V, Marques A. Are in person and telephone interviews equivalent modes of administrating the CAT, the FACIT-FS and the SGRQ in people with COPD? Front Rehabil Sci 2021;2:729190 [FREE Full text] [doi: 10.3389/fresc.2021.729190] [Medline: 36188865]
29. Weldam SW, Schuurmans MJ, Liu R, Lammers JJ. Evaluation of quality of life instruments for use in COPD care and research: a systematic review. Int J Nurs Stud 2013;50(5):688-707. [doi: 10.1016/j.ijnurstu.2012.07.017] [Medline: 22921317]

30. Jones PW, Harding G, Berry P, Wiklund I, Chen W, Kline Leidy N. Development and first validation of the COPD assessment test. Eur Respir J 2009;34(3):648-654 [FREE Full text] [doi: 10.1183/09031936.00102509] [Medline: 19720809]
31. Gaveikaite V, Fischer C, Schonenberg H, Pauws S, Kitsiou S, Chouvarda I, et al. Telehealth for patients with chronic obstructive pulmonary disease (COPD): a systematic review and meta-analysis protocol. BMJ Open 2018;8(9):e021865 [FREE Full text] [doi: 10.1136/bmjopen-2018-021865] [Medline: 30232108]
32. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
33. Alghamdi SM, Rajah AMA, Aldabayan YS, Aldhahir AM, Alqahtani JS, Alzahrani AA. Chronic obstructive pulmonary disease patients' acceptance in E-Health clinical trials. Int J Environ Res Public Health 2021;18(10):5230 [FREE Full text] [doi: 10.3390/ijerph18105230] [Medline: 34069028]
34. Brown W, Yen PY, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
35. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. Med Care Res Rev 2007;64(6):650-672. [doi: 10.1177/1077558707305942] [Medline: 17717378]
36. Lobos K, Cobo-Rendón R, Guzmán E, Bruna C. Adaptación y validación de dos cuestionarios sobre implementación de la tecnología en la docencia universitaria. Form. Univ 2022;15(5):1-14 [FREE Full text] [doi: 10.4067/S0718-50062022000500001]
37. Bravo LG. Possibilities for using TAM and technology frames models to assess the acceptance of new technologies in the chilean higher education quality assurance. Rev. Electrón. Educ 2015;19(2):181-196.
38. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
39. Wilkinson TMA, Donaldson GC, Hurst JR, Seemungal TAR, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]
40. Rixon L, Hirani SP, Cartwright M, Beynon M, Doll H, Steventon A, et al. A RCT of telehealth for COPD patient's quality of life: the whole system demonstrator evaluation. Clin Respir J 2017;11(4):459-469. [doi: 10.1111/crj.12359] [Medline: 26260325]
41. Wang L, Guo Y, Wang M, Zhao Y. A mobile health application to support self-management in patients with chronic obstructive pulmonary disease: a randomised controlled trial. Clin Rehabil 2021;35(1):90-101. [doi: 10.1177/0269215520946931] [Medline: 32907384]
42. Kayyali R, Savickas V, Spruit MA, Kaimakamis E, Siva R, Costello RW, et al. Qualitative investigation into a wearable system for chronic obstructive pulmonary disease: the stakeholders' perspective. BMJ Open 2016;6(8):e011657 [FREE Full text] [doi: 10.1136/bmjopen-2016-011657] [Medline: 27580831]
43. Janjua S, Banchoff E, Threapleton CJ, Prigmore S, Fletcher J, Disler RT. Digital interventions for the management of chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2021;4(4):CD013246 [FREE Full text] [doi: 10.1002/14651858.CD013246.pub2] [Medline: 33871065]
44. Dunn J, Coravos A, Fanarjian M, Ginsburg GS, Steinhubl SR. Remote digital health technologies for improving the care of people with respiratory disorders. Lancet Digit Health 2024;6(4):e291-e298 [FREE Full text] [doi: 10.1016/S2589-7500(23)00248-0] [Medline: 38402128]
45. Cox NS, Dal Corso S, Hansen H, McDonald CF, Hill CJ, Zanaboni P, et al. Telerehabilitation for chronic respiratory disease. Cochrane Database Syst Rev 2021;1(1):CD013040 [FREE Full text] [doi: 10.1002/14651858.CD013040.pub2] [Medline: 33511633]
46. Middlemass JB, Vos J, Siriwardena AN. Perceptions on use of home telemonitoring in patients with long term conditions - concordance with the health information technology acceptance model: a qualitative collective case study. BMC Med Inform Decis Mak 2017;17(1):89 [FREE Full text] [doi: 10.1186/s12911-017-0486-5] [Medline: 28651588]
47. Lundell S, Holmner, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir Med 2015;109(1):11-26 [FREE Full text] [doi: 10.1016/j.rmed.2014.10.008] [Medline: 25464906]
48. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Perceptions of factors influencing engagement with health and well-being apps in the United Kingdom: qualitative interview study. JMIR Mhealth Uhealth 2021;9(12):e29098 [FREE Full text] [doi: 10.2196/29098] [Medline: 34927597]
49. 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;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
50. Donevant SB, Messias DKH, Estrada RD. Utilization of mobile applications in collaborative patient-provider monitoring of chronic health conditions: an examination of three theoretical frameworks to guide practice. J Inform Nurs 2018;3(2):6-11 [FREE Full text] [Medline: 31179417]

51. 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;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
52. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mHealth apps in Australian general practice: qualitative study. JMIR Mhealth Uhealth 2020;8(7):e17447 [FREE Full text] [doi: 10.2196/17447] [Medline: 32729839]
53. 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;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
54. de Veer AJE, Peeters JM, Brabers AEM, Schellevis FG, Rademakers JJDJM, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
55. Tsertsidis A, Kolkowska E, Hedström K. Factors influencing seniors' acceptance of technology for ageing in place in the post-implementation stage: a literature review. Int J Med Inform 2019;129:324-333 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.06.027] [Medline: 31445274]
56. Ding H, Fatehi F, Maiorana A, Bashi N, Hu W, Edwards I. Digital health for COPD care: the current state of play. J Thorac Dis 2019;11(Suppl 17):S2210-S2220 [FREE Full text] [doi: 10.21037/jtd.2019.10.17] [Medline: 31737348]

Abbreviations

ANCOVA: analysis of covariance
CAT: chronic obstructive pulmonary disease assessment test
CG: control group
COPD: chronic obstructive pulmonary disease
HO: home oxygen
IG: intervention group
mHealth: mobile health
OR: odds ratio
TAM: technology acceptance model

Edited by YH Lin; submitted 28.08.24; peer-reviewed by B Chen, PJ Honkoop; comments to author 26.09.24; revised version received 14.10.24; accepted 14.10.24; published 06.01.25.

Please cite as:
Naranjo-Rojas A, Perula-de Torres LÁ, Cruz-Mosquera FE, Molina-Recio G
Efficacy and Acceptability of a Mobile App for Monitoring the Clinical Status of Patients With Chronic Obstructive Pulmonary Disease Receiving Home Oxygen Therapy: Randomized Controlled Trial
J Med Internet Res 2025;27:e65888
URL: https://www.jmir.org/2025/1/e65888
doi:10.2196/65888
PMID:

Original Paper

The Association Between Internet Addiction and the Risk of Suicide Attempts in Chinese Adolescents Aged 11-17 Years: Prospective Cohort Study

Sihong Li[1], PhD; Xingyue Jin[1], PhD; Lintong Song[1], PhD; Tianqing Fan[1], PhD; Yanmei Shen[1*], PhD; Jiansong Zhou[1*], PhD

Department of Psychiatry, National Clinical Research Center for Mental Disorders, and National Center for Mental Disorders, The Second Xiangya Hospital of Central South University, Changsha, China

*these authors contributed equally

Corresponding Author:
Yanmei Shen, PhD
Department of Psychiatry, National Clinical Research Center for Mental Disorders, and National Center for Mental Disorders
The Second Xiangya Hospital of Central South University
No.139, Renmin Road Central Changsha-Hunan-China
Changsha, 410011
China
Phone: 86 15116269263
Email: ymshen@csu.edu.cn

Abstract

Background: Suicide is a critical public health issue in adolescents worldwide. Internet addiction may play a role in the increased rate of suicide attempts in this population. However, few studies have explored the relationship between pathological internet use and suicide attempts among adolescents.

Objective: This study aimed to conduct a prospective cohort study to examine whether higher severity of pathological internet use was associated with an increased risk of suicide attempts among Chinese adolescents.

Methods: A total of 782 adolescents were recruited from a middle school from November 2020 to December 2020 and followed up for 6 months. An online self-reported questionnaire was used to collect the participants' demographic data and assess their mental health. The Depression, Anxiety, and Stress Scale–21 items (DASS-21) was used to evaluate depression, anxiety, and stress. The Chen Internet Addiction Scale–Revised (CIAS-R) was used to assess the symptoms and severity of pathological internet use. χ^2 test and ANOVA were used for intergroup comparison, and logistic regression analysis was used to examine the relationship between the severity of pathological internet use and suicide attempts. We also used a restricted cubic splines model to investigate the pattern of the association.

Results: The participants had an average age of 12.59 (SD 0.64) years, with the majority being of Han ethnicity (743/782, 95.01%) and more than half being male (426/782, 54.48%). Most participants had no previous history of depression (541/782, 69.18%), anxiety (415/782, 53.07%), or stress (618/782, 79.03%). The rate of newly reported suicide attempts was 4.6% (36/782). A significant positive association was observed between internet addiction and suicide attempts (odds ratio 3.88, 95% CI 1.70-8.82), which remained significant after adjusting for age, sex, ethnicity, anxiety, depression, and stress (odds ratio 2.65, 95% CI 1.07-6.55). In addition, this association exhibited a linear pattern in the restricted cubic spline regression model.

Conclusions: This study suggested that internet addiction, rather than internet overuse, was associated with a higher likelihood of suicide attempts, which highlighted the importance of addressing internet addiction symptoms among Chinese adolescents for suicide prevention.

(J Med Internet Res 2025;27:e52083) doi:10.2196/52083

KEYWORDS

adolescents; pathological internet use; internet addiction; suicide attempts; risk factors; cohort study

Introduction

Suicide has emerged as a major public health concern globally recently, as it has brought significant challenges to human beings with severe consequences [1]. Over the past 5 decades, suicide remains the only contributor to rising mortality [2]. Globally, suicide ranked as the fourth leading cause of mortality among adolescents aged 15 to 19 years, as cited in a 2021 report by the World Health Organization (WHO) [3]. In the United States, it was reported that there were 2744 suicides among adolescents aged 10-19 years in 2019, positioning suicide as the second most common cause of death [3]. Therefore, reducing suicide - related deaths has become a national priority in the United States, with more than US $22 million spent per year on suicide prevention. However, the suicide rate in the United States has not significantly declined [4]. China also faces the same challenge. Although the overall suicide rate in China has declined, the suicide rate among young people has remained high [5]. It is broadly recognized that suicide attempts may be the most significant predictor for completed suicide [6-9]. According to the data released by the National Trauma Data Bank, the mortality rate was 13% among adolescents aged younger than 20 years who attempted suicide [10]. A previous study indicated that teenagers who engaged in suicide attempts were 30 times more likely to die by suicide [11]. In addition, compared with those engaging in other forms of injury that lead to hospitalization, individuals with suicide attempts often have higher severities of injury, higher demands for health care services, and higher mortality [12]. Thus, examining the factors contributing to suicide attempts is crucial for the prevention of suicide among adolescents.

The internet has become increasingly accessible and necessary in the daily lives of adolescents. In China, internet access was available in 98.4% of elementary and middle schools in 2019 [13]. Meanwhile, in 2020, approximately 13.5% of all internet users were adolescents (aged 10-19 years) [14]. Proper internet use can help improve social skills, promote self-learning, and diversify recreation [15,16]. However, due to the highly time-consuming nature of the internet, an increasing number of adolescents have internet overuse, spending even more time using the internet than sleeping or studying [17]. In severe cases, excessive internet use may lead to clinically significant dysfunction and distress among adolescents [18,19]. This new form of behavioral addiction is listed as a tentative disorder named internet addiction. Internet addiction has also been increasing in Mainland China, with a reported prevalence of 10.4% to 26.5% among adolescents [20,21]. Earlier studies suggested a link between internet addiction and an increased incidence of suicide attempts [22,23]. A meta-analysis demonstrated that even after adjusting for demographic factors and depressive symptoms, those who experienced internet addiction continued to show a higher frequency of suicide attempts (pooled adjusted odds ratio [OR] 1.559) [23]. Some cross-sectional studies also indicated a link between internet addiction and suicidal behavior. For example, a study showed that adolescents with internet addiction had a greater likelihood of attempting suicide than those without this problem [24,25]. In addition, a 1-year follow-up study revealed that internet addiction was predictive of self-destructive behaviors among adolescents [26]. However, most previous studies are cross-sectional, which precluded the inference of causality. Furthermore, limited studies have investigated the relationship between the severity of pathological internet use and suicide attempts. Thus, in this prospective cohort study, we aimed to examine whether higher severity of pathological Internet use was associated with an increased risk of suicide attempts among Chinese adolescents [23].

Methods

Study Population

In this prospective cohort study, a 2-stage cluster sampling design was used to enhance the feasibility and representativeness. The sampling frame consisted of all public secondary schools across all districts in Changsha, as provided by the local education bureau. In the first stage, 1 public secondary school was randomly selected from the sampling frame. In the second stage, all students in grades 7 and 8 from the selected school were included in the study, resulting in a total sample of 1162 students (609 in grade 7 and 553 in grade 8). The selected school is comparable with most public secondary schools in the area in terms of size and demographic factors, such as gender ratio and socioeconomic background. This makes the school representative of the student population in Changsha and similar cities.

Recruitment occurred from November 2020 to December 2020, and follow-up data were collected from May 2021 to June 2021. Adolescents completed a standardized, self-administered questionnaire distributed through WeChat, one of the largest social media platforms in China. Before the commencement of the survey, we provided comprehensive training to schoolteachers on the procedures for distributing the questionnaire, offering guidance, and assisting students with difficulties. Data were collected from self-reported questionnaires administered through WeChat. All variables were collected through WeChat with the support of school personnel.

With the assistance of the school personnel, we obtained informed consent from both the adolescents and their parents. The participants accessed the survey using their parents' mobile phones and provided electronic informed consent on the first page of the survey. Upon completion, we thoroughly checked the data to ensure the data quality. Questionnaires were considered unqualified if the completion time was excessively short, if there were evident logical inconsistencies in responses, or if critical answers were missing.

Inclusion criteria of participants for the baseline and follow-up surveys were as follows: (1) middle school students aged 11-17 years, (2) with no severe medical conditions that prevented them from completing the survey, (3) capable of comprehending and finishing the questionnaires, and (4) provided a formal agreement to participate in the research. Participants who reported suicide attempts at baseline were excluded from the follow-up study. At baseline, 1162 grade 7-8 students were recruited, among whom 78 refused to participate. Therefore, 1084 students completed the survey at baseline. For the follow-up survey, 135 individuals were excluded due to suicide attempts at baseline, and 167 were excluded for refusal to participate; thus, 782 participants were included in the follow-up survey.

Assessment

Confounders

Sociodemographic Characteristics

Demographic variables, including sex, age, and ethnicity, were collected from the participants.

Depression, Anxiety, and Stress

The Depression, Anxiety, and Stress Scale–21 items (DASS-21) was used to assess participants' levels of depression, anxiety, and stress during the previous week [27]. The assessment includes 3 subscales: Depression, Anxiety, and Stress. Each of these subscales comprises 7 items that are evaluated using a 4-point Likert scale, which ranges from 0 (not at all) to 3 (most of the time) [28]. The thresholds were set at ≥13 for depression, ≥7 for anxiety, and ≥10 for stress [27]. The DASS-21 has shown high internal consistency (Cronbach α=0.92), indicating its good cross-cultural validity [29].

Exposures: Pathological Internet Use

Based on the conceptual framework, we used the Chen Internet Addiction Scale–Revised (CIAS-R) to assess the severity of pathological internet use [30]. The CIAS-R is a widely used instrument for evaluating pathological internet use, particularly in the Chinese context [31]. This self-reported scale includes 19 items, where each item is evaluated using a 4-point Likert scale that ranges from 1 (complete noncompliance) to 4 (complete conformity). The Chen Internet Addiction Scale (CIAS) total score varies between 0 and 76, with higher scores indicating more severe pathological internet use [32]. According to the total score, participants were further divided into 3 groups: the normal internet use group (<46), the internet overuse group (46-52), and the internet addiction group (≥53) [32]. The CIAS has proven to be a reliable and valid measure of pathological internet use among adolescents [30,33,34].

Outcomes: Suicide Attempts

Suicide attempts refer to the act of self-injuring that a person engages in with the intention of taking their own life. In this study, suicide attempts were defined as "someone trying to hurt themselves on purpose with the intention of ending their life, but it did not succeed" to tailor to junior high school students' comprehension levels. The question "Have you ever tried to commit suicide?" [22] was used at baseline to measure suicide attempts, and the question "Have you tried to commit suicide in the past 6 months?" was used to measure suicide attempts during the 6-month follow-up.

Statistical Analysis

In this study, the χ^2 test was used to compare categorical variables across groups. Continuous data are shown using mean and SDs, and the comparison of continuous variables between groups was conducted using ANOVA. The relationship between pathological internet use and suicide attempts was examined using logistic regression analysis using 3 models. Model 1 was used to examine the unadjusted OR; Model 2 was used to examine the adjusted OR after controlling for age, sex, and ethnicity; and Model 3 was used to examine the adjusted OR after adjusting for age, sex, ethnicity, anxiety, depression, and stress. The restricted cubic splines model was used to investigate the pattern of the association between the CIAS and suicide attempts after adjusting for age, sex, ethnicity, anxiety, depression, and stress. Finally, a post hoc power analysis was performed using G*Power 3.1 (Heinrich Heine University Düsseldorf) to evaluate the statistical power of the logistic regression analysis. The modeling of restricted cubic splines was conducted with R (version 4.2.3; R Foundation for Statistical Computing), while SPSS (version 26.0; IBM) was used for all additional statistical analyses. A 2-tailed $P<.05$ indicated statistical significance.

Ethical Considerations

This study protocol was examined and endorsed by the Ethics Committee of Central South University's Second Xiangya Hospital (K009). With the assistance of the school personnel, we obtained informed consent from both the adolescents and their parents on the first page of the questionnaire. The personal details of the participants were kept confidential, and a distinct study identification number was allocated to every participant for data entry, management, and analysis. The participants received free psychological assessments and medical help when needed as compensation for their participation.

Results

The analysis included 782 participants, whose average age was 12.59 (SD 0.64) years. Among these participants, more than half were male (426/782, 54.48%), and the majority were of Han ethnicity (743/782, 95.01%). Most of the participants reported no previous history of depression (541/782, 69.18%), anxiety (415/782, 53.07%), or stress (618/782, 79.03%). The rate of newly reported suicide attempts was 4.6% (36/782). Table 1 presents the comparison of demographic and clinical characteristics of participants among the normal internet use group, the internet overuse group, and the internet addiction group. The 3 groups showed significant differences in ethnicity ($P=.03$), depression ($P<.001$), anxiety ($P<.001$), stress ($P<.001$), and suicide attempts ($P=.003$). Compared with the normal internet use group, the internet overuse group and the internet addiction group were more likely to have stress, anxiety, depression, and suicide attempts after 6 months ($P<.001$ for stress, $P<.001$ for anxiety, $P<.001$ for depression, and $P=.003$ for suicide attempts).

Table 1. Demographic and clinical characteristics of students by pathological internet use.

Variables	NIU[a] (n=609)	IO[b] (n=99)	IA[c] (n=74)	*F* test or chi-sqaure test (*df*)	*P* value
Age (years), mean (SD)	12.59 (0.62)	12.60 (0.62)	12.52 (0.78)	0.38 (2779)	.68
Sex, n (%)				0.18 (2)	.92
Male	330 (54.19)	54 (54.55)	42 (56.76)		
Female	279 (45.81)	45 (45.45)	32 (43.24)		
Ethnicity, n (%)				7.20 (2)	.03
Han	580 (95.24)	97 (97.98)	66 (89.19)		
others	29 (4.76)	2 (2.02)	8 (10.81)		
Depression, n (%)				52.14 (2)	<.001
No	460 (75.53)	47 (47.47)	34 (45.95)		
Yes	149 (24.47)	52 (52.53)	40 (54.05)		
Anxiety, n (%)				38.22 (2)	<.001
No	359 (58.95)	32 (32.32)	24 (32.43)		
Yes	250 (41.05)	67 (67.68)	50 (67.57)		
Stress, n (%)				54.97 (2)	<.001
No	516 (84.73)	61 (61.62)	41 (55.41)		
Yes	93 (15.27)	38 (38.38)	33 (44.59)		
SA[d] in 6 months, n (%)				11.96 (2)	.003
No	588 (96.55)	93 (93.94)	65 (87.84)		
Yes	21 (3.45)	6 (6.06)	9 (12.16)		

[a] NIU: normal internet use.

[b] IO: internet overuse.

[c] IA: internet addiction.

[d] SA: suicide attempts.

The associations between pathological internet use and suicide attempts across various models are presented in Table 2. Compared with the normal internet use group, the internet addiction group consistently showed a higher risk of suicidal attempts in Model 1 (crude OR 3.88 95% CI 1.70-8.82), Model 2 (adjusted OR 3.88, 95% CI 1.65-9.11), and Model 3 (adjusted OR 2.65, 95% CI 1.07-6.55). The restricted cubic spline regression model showed a linear association between CIAS and suicide attempts (Figure 1). Furthermore, internet overuse was also linked to an increased risk of suicidal attempts, but the association lacked statistical significance. The post hoc power analysis indicated that the statistical power of this study was 82.60% based on a sample size of 782 participants, an α level of .05, and the event probability under the null hypothesis of 3.45%.

Figure 1. Cubic model of the association between the CIAS score and the risk of suicide attempts after adjusting for age, sex, ethnicity, anxiety, depression, and stress. CIAS: Chen Internet Addiction Scale; OR: odds ratio.

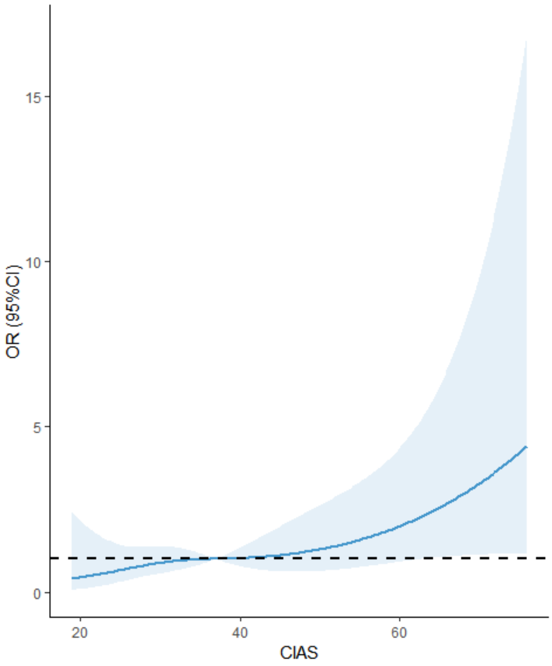


Table 2. The associations between pathological internet use and suicide attempts.

Group	Suicidal attempts, n/N (%)	Model 1[a]		Model 2[b]		Model 3[c]	
		OR[d] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
NIU[e] (ref[f])	21/609 (3.45)	1.00	—[g]	1.00	—	1.00	—
IO[h]	6/99 (6.06)	1.81 (0.71-4.59)	.21	1.90 (0.74-4.90)	.18	1.38 (0.52-3.68)	.52
IA[i]	9/74 (12.16)	3.88 (1.70-8.82)	.001	3.88 (1.65-9.11)	.002	2.65(1.07-6.55)	.04
P value for trend	—	—	.01	—	.01	—	.11

[a] Model 1: crude odds ratio.

[b] Model 2: odds ratio adjusted for age, sex, and ethnicity.

[c] Model 3: odds ratio adjusted for age, sex, ethnicity, anxiety, depression, and stress.

[d] OR: odds ratio.

[e] NIU: normal internet use.

[f] Ref: Reference.

[g] Not available.

[h] IO: internet overuse.

[i] IA: internet addiction.

Discussion

Principal Findings

To our knowledge, our research represents the initial investigation into the association between varying degrees of pathological internet use and suicide attempts among Chinese adolescents, using a prospective cohort design. Our findings suggested that the risk of suicide attempts was higher in individuals with internet addiction, even when controlling for variables such as age, sex, ethnicity, anxiety, depression, and stress. Furthermore, the restricted cubic splines model supported the linearity of the association between pathological internet use and the risk of suicide attempts.

This study found a notable association between internet addiction and suicide attempts through the 6 months of follow-up, which is in line with previous findings [35]. A study involving 8098 Chinese college students showed that those with internet addiction were at a higher risk for suicidal behaviors, with a prevalence of 21.4% [22]. Furthermore, a 1-year longitudinal study revealed that internet addiction was predictive of self-destructive behaviors, including self-harm and suicidal behaviors [26]. Furthermore, 2 meta-analyses also indicated that internet addiction was positively associated with suicidal behaviors [23,36]. Some reasons might explain this association. First, individuals with internet addiction have a greater chance of exposure to online content involving suicidal thoughts and behaviors, which may result in imitation of suicidal behaviors [37], especially for individuals with mental health problems [38]. Second, risk factors for both suicidal attempts and internet addiction, such as depression, anxiety, and low socioeconomic status, seem to overlap [39,40]. Third, alternative explanations for the link between internet addiction and suicide attempts may involve brain and cognitive dysfunctions.

For instance, internet addiction is often comorbid with some mental health disorders involving impairment of impulse control, such as attention-deficit/hyperactivity disorder and substance use disorders, suggesting the impulsive nature associated with internet addiction [41,42]. Furthermore, a magnetic resonance imaging study showed that individuals with internet addiction had impaired prefrontal lobe function, which might lead to decreased ability of impulse control or even cognitive deficits [43]. Therefore, adolescents with internet addiction may exhibit a propensity for risky decision-making, such as suicide attempts [44].

After adjusting for age, sex, ethnicity, stress, anxiety, and depression, the OR of suicidal attempts decreased from 3.88 (95% CI 1.70-8.82) to 2.64 (95% CI 1.07-6.55). Still, the link between internet addiction and suicide attempts continued to be significant. A study conducted in Korea revealed a significant positive association between internet addiction, depression, and suicidal thoughts among adolescents [45]. A study involving Chinese middle school students also pointed out that internet addiction was linked to suicide attempts among adolescents, and the association might be mediated by depression and bullying behaviors [46]. This association may also be related to the proneness of comorbidity of internet addiction and other mental health problems such as depression, decreased self-esteem, lower tolerance to frustration, and emotional distress [47,48], which also increases the risk of suicidal attempts. Some studies suggested that the connection between internet addiction and suicide attempts might be mediated by these comorbidities, which, therefore, increased the risk of suicide attempts [49].

Interestingly, this study found that internet overuse did not have a significant association with suicide attempts, indicating that suicide attempts might only be related to internet addiction, which is more severe. This aligns with an earlier study on university students, which showed that most of the mental health problems in participants with mild internet addiction were comparable with or slightly above the average. However, the rates of mental health problems sharply increased in those with moderate to severe internet addiction [50]. Some other studies suggested that engaging with the internet at levels below addiction may increase the feeling of being accepted and alleviate feelings of loneliness and shame instead of leading to severe consequences [51].

Limitations

Some limitations need to be addressed. First, this study did not collect the specific dates of suicide attempts during the follow-up. This study only used a fixed 6-month follow-up period, which may introduce potential bias to our study results. Future studies should consider adding the specific dates of suicide attempts to get a more robust conclusion. Second, although we controlled for key risk factors such as anxiety, depression, and stress, which are known to influence suicidal behaviors, other important factors, such as academic burden and parent-child relationship, were not included in our analysis. Future research needs to include a broader range of risk factors and intermediate processes (eg, suicidal ideation) to get a more comprehensive understanding of the risk factors of suicide attempts. Third, the participants' internet addiction and clinical correlates were collected using self-report questionnaires and scales, which might have resulted in bias such as recall bias and social desirability bias. Future studies should consider using diagnostic interviews to get more accurate assessments. Fourth, the sample was recruited from 1 middle school in Hunan Province, China, and may not represent students from other schools in other areas. Future multicenter studies across various schools and regions are needed to get a more representative sample and improve the generalizability of our results. Finally, suicide attempts were assessed based on a single self-reported question instead of a standardized scale, hindering the cross-comparison with other studies using standard measurement tools [52]. For instance, Yang et al [53] developed a standardized and reliable assessment tool for suicide attempts, which may yield more accurate and consistent data across studies. Future studies should consider using more standard tools to capture the complexity of suicidal behaviors fully.

Conclusion

This study found that adolescents experiencing internet addiction and overuse were at a higher risk for stress, anxiety, and depression. Furthermore, internet addiction was associated with an increased risk of suicide attempts among adolescents in China. Given the considerable impact of suicide on society, the results of this study underscore the importance of addressing internet addiction among Chinese adolescents as part of suicide prevention and management efforts.

Acknowledgments

We would like to express our gratitude to all the participants in this project, as well as the school personnel, who have offered invaluable support.

This work was supported by the Scientific Research Launch Project for New Employees of the Second Xiangya Hospital of Central South University, National Natural Science Foundation of China (82201703), and Hunan Provincial Natural Science Foundation of China (2023JJ60085). The authors would also like to acnowledge the following funding sources: The STI2030-Major Projects-2021ZD0200700； The National Natural Science Foundation of China (82071543, 82471568); and The Key Research and Development Program of Hunan Province (2023SK2028). No generative artificial intelligence was used in the manuscript writing.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JZ, YS, SL, XJ, and TF designed the study. SL, LS, JZ, and YS were responsible for the integrity of the data and the accuracy of data analysis. JZ, YS, and SL prepared the manuscript. SL, XJ, and TF edited and revised the manuscript. JZ, YS, and SL were responsible for the critical revision of the manuscript. All authors have contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Abio A, Owusu PN, Posti JP, Bärnighausen T, Shaikh MA, Shankar V, et al. Cross-national examination of adolescent suicidal behavior: a pooled and multi-level analysis of 193,484 students from 53 LMIC countries. Soc Psychiatry Psychiatr Epidemiol 2022;57(8):1603-1613 [FREE Full text] [doi: 10.1007/s00127-022-02287-x] [Medline: 35445842]
2. Clayton MG, Pollak OH, Owens SA, Miller AB, Prinstein MJ. Advances in research on adolescent suicide and a high priority agenda for future research. J Res Adolesc 2021;31(4):1068-1096. [doi: 10.1111/jora.12614] [Medline: 34820949]
3. World Health Organization. Suicide in the world: global health estimates. 2019. URL: https://iris.who.int/bitstream/handle/10665/326948/WHO-MSD-MER-19.3-eng.pdf?sequence=1&isAllowed=y [accessed 2024-12-05]
4. Office of the Surgeon General (US), National Action Alliance for Suicide Prevention (US). 2012 National Strategy for Suicide Prevention: Goals and Objectives for Action: A Report of the U.S. Surgeon General and of the National Action Alliance for Suicide Prevention. Washington (DC): US Department of Health & Human Services; 2012:184.
5. Wang CW, Chan CLW, Yip PSF. Suicide rates in China from 2002 to 2011: an update. Soc Psychiatry Psychiatr Epidemiol 2014;49(6):929-941. [doi: 10.1007/s00127-013-0789-5] [Medline: 24240568]
6. Zhong BL, Chan SSM, Liu TB, Chiu HFK. Nonfatal suicidal behaviors of Chinese rural-to-urban migrant workers: attitude toward suicide matters. Suicide Life Threat Behav 2019;49(5):1199-1208. [doi: 10.1111/sltb.12519] [Medline: 30315652]
7. Dai J, Zhong B, Xiang Y, Chiu HFK, Chan SSM, Yu X, et al. Internal migration, mental health, and suicidal behaviors in young rural Chinese. Soc Psychiatry Psychiatr Epidemiol 2015;50(4):621-631 [FREE Full text] [doi: 10.1007/s00127-014-0985-y] [Medline: 25403568]
8. Zhong BL, Xie WX, Zhu JH, Lu J, Chen H. Prevalence and correlates of suicide attempt among Chinese individuals receiving methadone maintenance treatment for heroin dependence. Sci Rep 2019;9(1):15859 [FREE Full text] [doi: 10.1038/s41598-019-52440-x] [Medline: 31666637]
9. Xu Y, Wang C, Shi M. Identifying Chinese adolescents with a high suicide attempt risk. Psychiatry Res 2018;269:474-480. [doi: 10.1016/j.psychres.2018.08.085] [Medline: 30195741]
10. Mathews EM, Woodward CJ, Musso MW, Jones GN. Suicide attempts presenting to trauma centers: trends across age groups using the National trauma data bank. Am J Emerg Med 2016;34(8):1620-1624. [doi: 10.1016/j.ajem.2016.06.014] [Medline: 27321943]
11. Shaffer D, Gould MS, Fisher P, Trautman P, Moreau D, Kleinman M, et al. Psychiatric diagnosis in child and adolescent suicide. Arch Gen Psychiatry 1996;53(4):339-348. [doi: 10.1001/archpsyc.1996.01830040075012] [Medline: 8634012]
12. King CA, Grupp-Phelan J, Brent D, Dean JM, Webb M, Bridge JA, et al. Predicting 3-month risk for adolescent suicide attempts among pediatric emergency department patients. J Child Psychol Psychiatry 2019;60(10):1055-1064 [FREE Full text] [doi: 10.1111/jcpp.13087] [Medline: 31328282]
13. The 46th report on Internet development in China. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/202012/P020201201530023411644.pdf [accessed 2024-12-04]
14. CINI: 2019 national survey on internet use of minors. 2020. URL: https://www3.cnnic.cn/NMediaFile/old_attach/P020200513370410784435.pdf
15. Nguyen HV, Huang HC, Wong MK, Yang YH, Huang TL, Teng CI. Moderator roles of optimism and weight control on the impact of playing exergames on happiness: the perspective of social cognitive theory using a randomized controlled trial. Games Health J 2018;7(4):246-252. [doi: 10.1089/g4h.2017.0165] [Medline: 29708775]
16. Gao YX, Wang JY, Dong GH. The prevalence and possible risk factors of internet gaming disorder among adolescents and young adults: Systematic reviews and meta-analyses. J Psychiatr Res 2022;154:35-43. [doi: 10.1016/j.jpsychires.2022.06.049] [Medline: 35926424]
17. Marin MG, Nuñez X, de Almeida RMM. Internet addiction and attention in adolescents: a systematic review. Cyberpsychol Behav Soc Netw 2021;24(4):237-249. [doi: 10.1089/cyber.2019.0698] [Medline: 33121255]

18. Bickham DS. Current research and viewpoints on internet addiction in adolescents. Curr Pediatr Rep 2021;9(1):1-10 [FREE Full text] [doi: 10.1007/s40124-020-00236-3] [Medline: 33457108]
19. Shaw M, Black DW. Internet addiction: definition, assessment, epidemiology and clinical management. CNS Drugs 2008;22(5):353-365. [doi: 10.2165/00023210-200822050-00001] [Medline: 18399706]
20. Chi X, Hong X, Chen X. Profiles and sociodemographic correlates of internet addiction in early adolescents in southern China. Addict Behav 2020;106:106385. [doi: 10.1016/j.addbeh.2020.106385] [Medline: 32187574]
21. Xin M, Xing J, Pengfei W, Houru L, Mengcheng W, Hong Z. Online activities, prevalence of internet addiction and risk factors related to family and school among adolescents in China. Addict Behav Rep 2018;7:14-18 [FREE Full text] [doi: 10.1016/j.abrep.2017.10.003] [Medline: 29450251]
22. Shen Y, Meng F, Xu H, Li X, Zhang Y, Huang C, et al. Internet addiction among college students in a Chinese population: prevalence, correlates, and its relationship with suicide attempts. Depress Anxiety 2020;37(8):812-821. [doi: 10.1002/da.23036] [Medline: 32400075]
23. Cheng YS, Tseng PT, Lin PY, Chen TY, Stubbs B, Carvalho AF, et al. Internet addiction and its relationship with suicidal behaviors: a meta-analysis of multinational observational studies. J Clin Psychiatry 2018;79(4):17r11761. [doi: 10.4088/JCP.17r11761] [Medline: 29877640]
24. Lin I, Ko C, Chang Y, Liu T, Wang P, Lin H, et al. The association between suicidality and internet addiction and activities in Taiwanese adolescents. Compr Psychiatry 2014;55(3):504-510. [doi: 10.1016/j.comppsych.2013.11.012] [Medline: 24457034]
25. Kim K, Ryu E, Chon M, Yeun E, Choi S, Seo J, et al. Internet addiction in Korean adolescents and its relation to depression and suicidal ideation: a questionnaire survey. Int J Nurs Stud 2006;43(2):185-192. [doi: 10.1016/j.ijnurstu.2005.02.005] [Medline: 16427966]
26. Pan PY, Yeh CB. Internet addiction among adolescents may predict self-harm/suicidal behavior: a prospective study. J Pediatr 2018;197:262-267. [doi: 10.1016/j.jpeds.2018.01.046] [Medline: 29550226]
27. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the beck depression and anxiety inventories. Behav Res Ther 1995;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
28. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. Psychol. Assess 1998;10(2):176-181. [doi: 10.1037//1040-3590.10.2.176]
29. Wang K, Shi H, Geng F, Zou L, Tan S, Wang Y, et al. Cross-cultural validation of the depression anxiety stress scale-21 in China. Psychol Assess 2016;28(5):e88-e100. [doi: 10.1037/pas0000207] [Medline: 26619091]
30. Mak K, Lai C, Ko C, Chou C, Kim D, Watanabe H, et al. Psychometric properties of the Revised Chen Internet Addiction Scale (CIAS-R) in Chinese adolescents. J Abnorm Child Psychol 2014;42(7):1237-1245. [doi: 10.1007/s10802-014-9851-3] [Medline: 24585392]
31. Zhang WZ, Bing D. The application evaluation of the Revised Chen Internet Addiction Scale among urban middle school students. Chin. J. Sch. Health 2015;36(10):1489-1491 [FREE Full text]
32. Bai Y, Fan F. A study on the internet dependence of college students: the revising and applying of a measurement. Psychol. Dev. Edu 2005:99-104 [FREE Full text]
33. Ko C, Yen C, Yen C, Chen CC, Yen CN, Chen S. Screening for internet addiction: an empirical study on cut-off points for the chen Internet addiction scale. Kaohsiung J Med Sci 2005;21(12):545-551 [FREE Full text] [doi: 10.1016/S1607-551X(09)70206-2] [Medline: 16670046]
34. Lai CM, Mak KK, Watanabe H, Ang RP, Pang JS, Ho RCM. Psychometric properties of the Internet Addiction Test in Chinese adolescents. J Pediatr Psychol 2013;38(7):794-807. [doi: 10.1093/jpepsy/jst022] [Medline: 23671059]
35. Tang J, Ma Y, Lewis SP, Chen R, Clifford A, Ammerman BA, et al. Association of internet addiction with nonsuicidal self-injury among adolescents in China. JAMA Netw Open 2020;3(6):e206863 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.6863] [Medline: 32496567]
36. Wang J, Hao Q, Tu Y, Peng W, Wang Y, Li H, et al. Assessing the association between internet addiction disorder and health risk behaviors among adolescents and young adults: a systematic review and meta-analysis. Front Public Health 2022;10:809232 [FREE Full text] [doi: 10.3389/fpubh.2022.809232] [Medline: 35433568]
37. Baker D, Fortune S. Understanding self-harm and suicide websites: a qualitative interview study of young adult website users. Crisis 2008;29(3):118-122. [doi: 10.1027/0227-5910.29.3.118] [Medline: 18714907]
38. Ko CH, Yen JY, Yen CF, Chen CS, Chen CC. The association between internet addiction and psychiatric disorder: a review of the literature. Eur Psychiatry 2012;27(1):1-8. [doi: 10.1016/j.eurpsy.2010.04.011] [Medline: 22153731]
39. Sueki H, Yonemoto N, Takeshima T, Inagaki M. The impact of suicidality-related internet use: a prospective large cohort study with young and middle-aged internet users. PLoS ONE 2014;9(4):e94841. [doi: 10.1371/journal.pone.0094841] [Medline: 24740115]
40. Tang CSK, Koh YYW. Online social networking addiction among college students in Singapore: comorbidity with behavioral addiction and affective disorder. Asian J Psychiatr 2017;25:175-178. [doi: 10.1016/j.ajp.2016.10.027] [Medline: 28262144]

41. Zhang Y, Mei S, Li L, Chai J, Li J, Du H. The relationship between impulsivity and internet addiction in Chinese college students: a moderated mediation analysis of meaning in life and self-esteem. PLoS One 2015;10(7):e0131597 [FREE Full text] [doi: 10.1371/journal.pone.0131597] [Medline: 26171860]
42. Weinstein A, Yaacov Y, Manning M, Danon P, Weizman A. Internet addiction and attention deficit hyperactivity disorder among schoolchildren. Isr Med Assoc J 2015;17(12):731-734 [FREE Full text] [Medline: 26897972]
43. Huang Y, Xu L, Kuang L, Wang W, Cao J, Xiao M. Abnormal brain activity in adolescents with internet addiction who attempt suicide: an assessment using functional magnetic resonance imaging. Neural Regen Res 2020;15(8):1554-1559 [FREE Full text] [doi: 10.4103/1673-5374.274346] [Medline: 31997822]
44. Cheng C, Sun P, Mak KK. Internet addiction and psychosocial maladjustment: avoidant coping and coping inflexibility as psychological mechanisms. Cyberpsychol Behav Soc Netw 2015;18(9):539-546. [doi: 10.1089/cyber.2015.0121] [Medline: 26348815]
45. Ryu EJ, Choi KS, Seo JS, Nam BW. The relationships of internet addiction, depression, and suicidal ideation in adolescents. Taehan Kanho Hakhoe Chi 2004;34(1):102-110. [doi: 10.4040/jkan.2004.34.1.102] [Medline: 15314344]
46. Pan S, Wuquan L, Ming L, Guo L, Deng X. Relationship between internet addiction and suicidal behavior among middle school students in Guangzhou. Chinese Journal of School Health. Chinese Journal of School Health 2018 Feb 25:229-231 [FREE Full text] [doi: 10.16835/j.cnki.1000-9817.2018.02.021]
47. Ko C, Yen J, Yen C, Chen C, Wang S. The association between internet addiction and belief of frustration intolerance: the gender difference. Cyberpsychol Behav 2008;11(3):273-278. [doi: 10.1089/cpb.2007.0095] [Medline: 18537496]
48. Wang L, Luo J, Bai Y, Kong J, Luo J, Gao W, et al. Internet addiction of adolescents in China: prevalence, predictors, and association with well-being. Addict. Res. Theory 2012;21(1):62-69. [doi: 10.3109/16066359.2012.690053]
49. Liu HC, Liu SI, Tjung JJ, Sun FJ, Huang HC, Fang CK. Self-harm and its association with internet addiction and internet exposure to suicidal thought in adolescents. J Formos Med Assoc 2017;116(3):153-160 [FREE Full text] [doi: 10.1016/j.jfma.2016.03.010] [Medline: 27146126]
50. Guo W, Tao Y, Li X, Lin X, Meng Y, Yang X, et al. Associations of internet addiction severity with psychopathology, serious mental illness, and suicidality: large-sample cross-sectional study. J Med Internet Res 2020;22(8):e17560 [FREE Full text] [doi: 10.2196/17560] [Medline: 32780029]
51. Duggan JM, Heath NL, Lewis SP, Baxter AL. An examination of the scope and nature of non-suicidal self-injury online activities: implications for school mental health professionals. School Mental Health 2011;4(1):56-67. [doi: 10.1007/s12310-011-9065-6]
52. Cao XL, Zhong BL, Xiang YT, Ungvari GS, Lai KY, Chiu HF, et al. Prevalence of suicidal ideation and suicide attempts in the general population of China: a meta-analysis. Int J Psychiatry Med 2015;49(4):296-308 [FREE Full text] [doi: 10.1177/0091217415589306] [Medline: 26060259]
53. Yang F, Tong J, Zhang S, Zhang J, Zhong B. Prevalence and correlates of suicide attempts in Chinese individuals with borderline personality disorder. Front Psychiatry 2022;13:942782 [FREE Full text] [doi: 10.3389/fpsyt.2022.942782] [Medline: 36104990]

Abbreviations

CIAS: Chen Internet Addiction Scale
CIAS-R: Chen Internet Addiction Scale-Revised
DASS-21: Depression, Anxiety, and Stress Scale–21 items
OR: odds ratio
WHO: World Health Organization

Edited by T de Azevedo Cardoso; submitted 22.08.23; peer-reviewed by M Wu, Y Latifeh, BL Zhong; comments to author 22.07.24; revised version received 21.08.24; accepted 18.11.24; published 03.01.25.

Please cite as:
Li S, Jin X, Song L, Fan T, Shen Y, Zhou J
The Association Between Internet Addiction and the Risk of Suicide Attempts in Chinese Adolescents Aged 11-17 Years: Prospective Cohort Study
J Med Internet Res 2025;27:e52083
URL: https://www.jmir.org/2025/1/e52083
doi:10.2196/52083
PMID:

Original Paper

US State Public Health Agencies' Use of Twitter From 2012 to 2022: Observational Study

Samuel R Mendez[1], MS; Sebastian Munoz-Najar[2], PhD; Karen M Emmons[1], PhD; Kasisomayajula Viswanath[1,3], PhD

[1]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[2]Harvard Graduate School of Education, Cambridge, MA, United States

[3]Dana Farber Cancer Institute, Boston, MA, United States

Corresponding Author:
Samuel R Mendez, MS
Department of Social and Behavioral Sciences
Harvard T.H. Chan School of Public Health
677 Huntington Ave
Boston, MA, 02115
United States
Phone: 1 617 432 1135
Email: smendez@g.harvard.edu

Abstract

Background: Twitter (subsequently rebranded as X) is acknowledged by US health agencies, including the US Centers for Disease Control and Prevention (CDC), as an important public health communication tool. However, there is a lack of data describing its use by state health agencies over time. This knowledge is important amid a changing social media landscape in the wake of the COVID-19 pandemic.

Objective: The study aimed to describe US state health agencies’ use of Twitter from 2012 through 2022. Furthermore, we organized our data collection and analysis around the theoretical framework of the networked public to contribute to the broader literature on health communication beyond a single platform.

Methods: We used Twitter application programming interface data as indicators of state health agencies’ engagement with the 4 key qualities of communication in a networked public: scalability, persistence, replicability, and searchability. To assess scalability, we calculated tweet volume and audience engagement metrics per tweet. To assess persistence, we calculated the portion of tweets that were manual retweets or included an account mention. To assess replicability, we calculated the portion of tweets that were retweets or quote tweets. To assess searchability, we calculated the portion of tweets using at least 1 hashtag.

Results: We observed a COVID-19 pandemic–era shift in state health agency engagement with scalability. The overall volume of tweets increased suddenly from less than 50,000 tweets in 2019 to over 94,000 in 2020, resulting in an average of 5.3 per day. Though mean tweets per day fell in 2021 and 2022, this COVID-19 pandemic–era low was still higher than the pre–COVID-19 pandemic peak. We also observed a more fragmented approach to searchability aligning with the start of the COVID-19 pandemic. More state-specific hashtags were among the top 10 during the COVID-19 pandemic, compared with more general hashtags related to disease outbreaks and natural disasters in years before. We did not observe such a clear COVID-19 pandemic–era shift in engagement with replicability. The portion of tweets mentioning a CDC account gradually rose and fell around a peak of 7.0% in 2018. Similarly, the rate of retweets of a CDC account rose and fell gradually around a peak of 5.4% in 2018. We did not observe a clear COVID-19 pandemic–era shift in persistence. The portion of tweets mentioning any account reached a maximum of 21% in 2013. It oscillated for much of the study period before dropping off in 2021 and reaching a minimum of 10% in 2022. Before 2018, the top 10 mentioned accounts included at least 2 non-CDC or corporate accounts. From 2018 onward, state agencies were much more prominent.

Conclusions: Overall, we observed a more fragmented approach to state health agency communication on Twitter during the pandemic, prioritizing volume over searchability, formally replicating existing messages, and leaving traces of interactions with other accounts.

(J Med Internet Res 2025;27:e59786) doi:10.2196/59786

KEYWORDS

social media; health communication; Twitter; tweet; public health; state government; government agencies; information technology; data science; communication tool; COVID-19 pandemic; data collection; theoretical framework; message; interaction

Introduction

Twitter as a Networked Public

This paper describes US state public health agencies' activity on Twitter (subsequently rebranded as X) from 2012 through 2022. We used the theoretical framework of a networked public, an interactive space that networked technologies make possible, as well as the collection of people that inhabit it [1].

According to this theoretical framework, a networked public's defining features are profiles, contact lists, and public communication tools. This framework outlines 4 emergent qualities of communication these features produce: scalability, persistence, replicability, and searchability. Twitter's features were always open to users with potentially conflicting motivations. Viewed generously, this aligns with notions of Twitter's democratization of health communication [2]. Viewed through a more critical lens, however, this illustrates how Twitter has been a site for both challenging and propagating inequities.

This tension was evident from Twitter's 2006 launch, with dominant narratives of its Silicon Valley origins obscuring its relationship to open-source innovations from activists [3]. The hashtag illustrates such long-term tensions that exist in networked publics, between corporate practices and grassroots discourse, as well as between social networks with opposing goals. Hashtag use began as an informal organizational practice among power users before Twitter formally incorporated it as a feature in 2007 [4]. While activists eventually adopted hashtags for social justice campaigns like #MeToo and #BlackLivesMatter, and corporations adopted hashtags for advertising purposes, hashtags were also tools of political marginalization. For example, the use of the hashtag #ChineseVirus was a marker of anti-Chinese sentiment in March 2020, associated with anti-Asian hashtags more broadly after the US president's use of the term "Chinese Virus" [5]. Hashtags can also be sites of conflict around public health topics, such as the coordinated flooding of the provaccine #DoctorsSpeakUp hashtag with antivaccine messaging in 2020 [6].

Twitter as a State Public Health Tool

Though the use of Twitter in public health has been critiqued as "tweeting to the choir" [7], this type of communication toward professional audiences could reflect a viable platform-specific communication strategy. For example, the US Centers for Disease Control and Prevention (CDC) has recommended that public health communicators and organizational leaders use Twitter to reach reporters with crisis communication [8]. This could help explain why US health agencies' COVID-19 tweets focused more on data than did their Facebook (Meta) posts [9].

Existing observational research reinforces this notion of Twitter as an information-focused public health communication tool. A study of state Medicaid program Twitter use between 2014 and 2019 found a similar informational focus, which garnered little audience engagement [10]. A study of Canadian public health agencies' Twitter use during the first half of 2020 found a similar informational focus, even as more action-oriented tweets received higher audience engagement [11]. A study of health agency use of Twitter across 7 countries during the spring of 2020 found evidence of this trend internationally, with announcements and reporting being the most common tweet theme [12]. At a local level in the US, there have been experiments in using Twitter as a community service tool, such as in the case of the Chicago Department of Public Health's foodborne illness response program [13].

Descriptive studies of US state public health agencies' Twitter use, however, are either not comprehensive or are not up to date. Twitter was found to be an emerging platform for state public health agencies, based on a study covering a 2-month period in 2011 [14]. An early content analysis of state public health agency tweets in 2012 found a focus on the transmission of personal health advice [15] and a much higher Twitter adoption rate than local health departments around the same time [16]. A more recent study analyzed all state health agency tweets from a 4-month period around the emergency use authorization of the first COVID-19 vaccine in the United States [17]. Findings indicated a much higher volume of tweets than earlier studies, as well as a lack of key terms related to vaccination, inequities, and racism, lagging behind on-the-ground efforts to address COVID-19 racial health inequities.

We present an instrumental case study [18] to provide a more complete picture of state public health agency Twitter use long-term. We chose this approach as it aligns with our research interest in producing descriptive knowledge with theoretical implications beyond the individual case of Twitter. Our goal was to provide transferable knowledge to help identify trends and relationships between theoretical constructs in health communication. This knowledge is important amid a changing social media landscape in the wake of the COVID-19 pandemic. We follow STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [19] in reporting this study. The completed STROBE statement checklist is available in Multimedia Appendix 1.

Methods

Overview

We conducted a descriptive analysis of state public health agency tweets from 2012 through 2022, structured as an instrumental case study organized around an issue question [19]: How did US state public health agencies engage with Twitter as a networked public?

Our data came from a download of public health agency tweets using the "academictwitteR" R package to access the Twitter application programming interface (API) through the "Twitter API v2 for Academic Research" on February 3, 2023 [18]. We filtered the data to focus on tweets in English from state public health agencies. The language filter relied on labels available by Twitter's proprietary API data. We identified tweets from the state-level public health agency in each state by leveraging associations between usernames and author IDs recorded during the data download process [20]. These accounts were initially identified through manual searches of state public health agency websites, as well as searches on Google and Twitter where necessary. Refer to Multimedia Appendix 2 for a list of account usernames and the associated state public health agencies. The study period begins on January 1, 2012, which allowed us to build on a previous study taking a random sampling approach to analyze state public health agency tweets in the same year [15]. The study period ends on December 31, 2022, which was the most recent full calendar year at the time of data collection. We did not account for missing data, as the API download method includes an exhaustive record of all published tweets from all specified accounts during the study period. We conducted this analysis using R (version 4.2.3, R Studio version 2023.06.1+524). The scripts enabling this analysis are available in an Open Science Foundation repository [21].

We calculated indicator variables to describe health agencies' engagement with the key qualities of communication in a networked public: scalability, persistence, replicability, and searchability [1].

Scalability

"Scalability" describes social media content's hypothetical ability to reach an entire network. We assessed scalability by tweet volume and audience engagement. We calculated tweets per year, categorized into types: replies, retweets, or quote tweets (from their formal introduction in 2015) [22]. We further categorized replies into self-replies and other replies by comparing the author and in-reply-to user IDs. We used the regular expression "RT @" to identify "manual retweets," a custom metric describing text copies of messages. Here, we pooled manual retweets with formal retweets. Finally, we calculated the mean, median, and IQR of states' mean tweets per day each year, as well as the likes, replies, retweets, and quotes on non-retweets. "Non-retweets" include manual retweets, as they are unique posts rather than pointers to source tweets.

Persistence

"Persistence" describes the digital records of interactions on networked publics. We assessed persistence by manual retweets. We also calculated mentions (a hyperlinked account username) among non-retweets, as their records do not depend on other accounts' activity. We calculated the percentage of tweets that were manual retweets each year. We calculated the percentage of non-retweets with at least 1 mention each year.

Replicability

"Replicability" refers to social media content's infinite duplicability. We assessed replicability by retweets and quote tweets. We calculated the percentage of tweets that were formal retweets and quote tweets each year. We further calculated the percentage of tweets that were formal retweets of a CDC account, as well as the percentage of tweets including a CDC account mention. We used regular expressions to identify usernames starting with "CDC" or "NIOSH," which comprised the majority of usernames listed as official CDC accounts [23]. For other accounts, we relied on exact matches with entire usernames.

Searchability

"Searchability" refers to ways that social media content is findable, which we assessed by hashtag use. We calculated the percentage of tweets with at least 1 hashtag per year. We used regular expressions to extract hashtags and calculate the top 10 most common per year (after converting to lowercase). We further calculated the percentage of tweets per year using #flu or #hiv as hashtags with long-term national relevance.

Ethical Considerations

The study data consisted of publicly available information from government institutional authors. We did not seek an ethics review and have no individual participant protections in place, as they are not relevant to this study design.

Results

Overview

We identified 570,335 tweets from state public health agency accounts in English from 2012 through 2022. All 50 states were represented in this data set. However, all 50 states were only active on Twitter in 2018 and 2019, as states gradually adopted Twitter during the study period, and 1 state published its last tweet in 2019. We included 52 accounts in this data set to account for 1 state creating a replacement public health agency account during the study period, as well as another state communicating through 2 accounts under its public health agency. An exhaustive summary data table is available in Multimedia Appendix 3.

Scalability

In total, 41 states were active at any time in 2012, meaning they posted at least 1 tweet. This number increased each year, up to 50 in 2018. It decreased to 49 in 2020. The total number of tweets across all accounts increased from 25,276 in 2012 to 51,132 in 2018. The total rose to 94,205 in 2020 and decreased to 62,797 in 2022 (Figure 1). Overall, mean tweets per day rose from 1.7 in 2012 to 2.8 in 2018. It reached a maximum of 5.3 in 2020 and remained above the 2018 level. Figure 2 displays the distribution of mean tweets per day among active states each year. Delaware, North Carolina, Florida, Indiana, Pennsylvania, Massachusetts, and West Virginia each had the most tweets in a given year. Delaware, Florida, and West Virginia each had the most for multiple years in a row. Massachusetts was the most active in a single year, averaging >21 tweets per day in 2020.

Figure 1. Stacked bar chart displaying an exhaustive enumeration of tweets from US state public health agency accounts for each year from 2012 through 2022, divided into 5 categories: retweets, self-replies, other-replies, quote tweets, and standalone tweets.

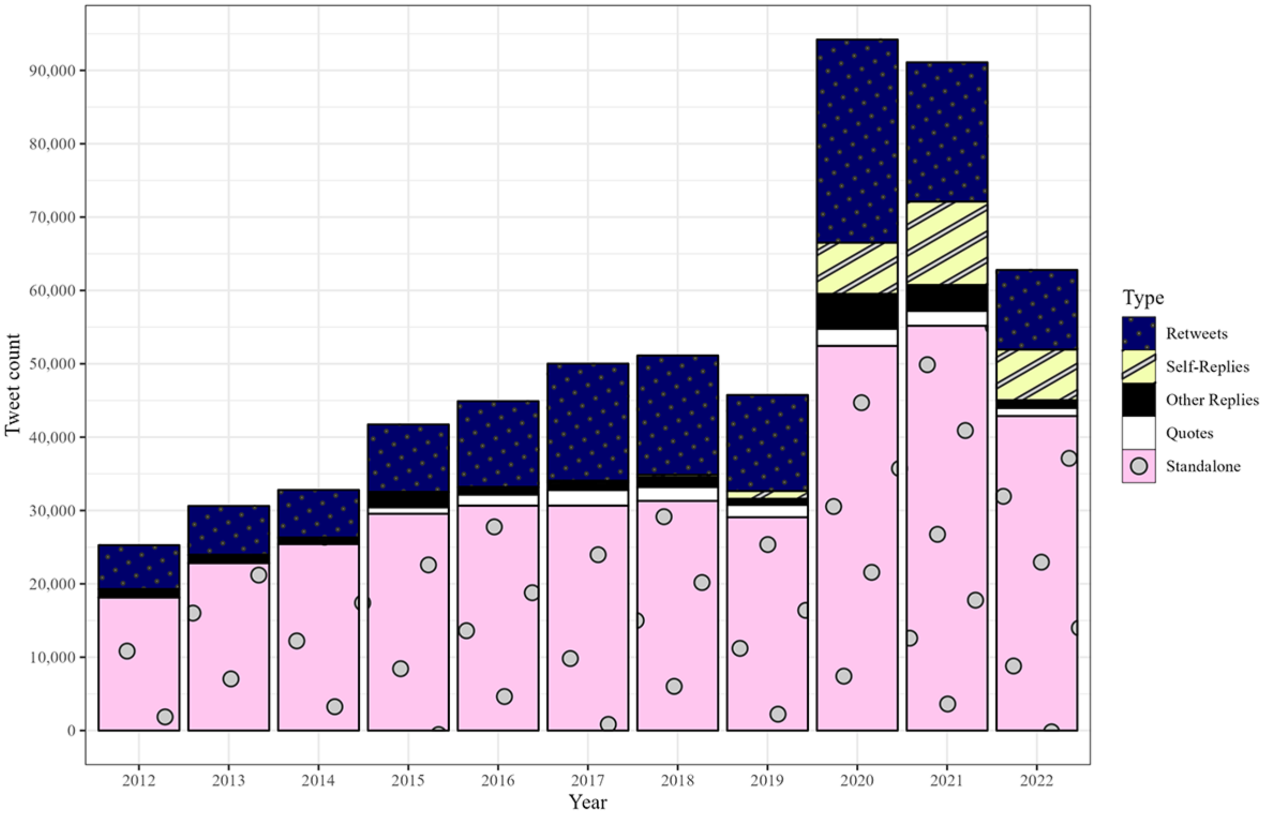
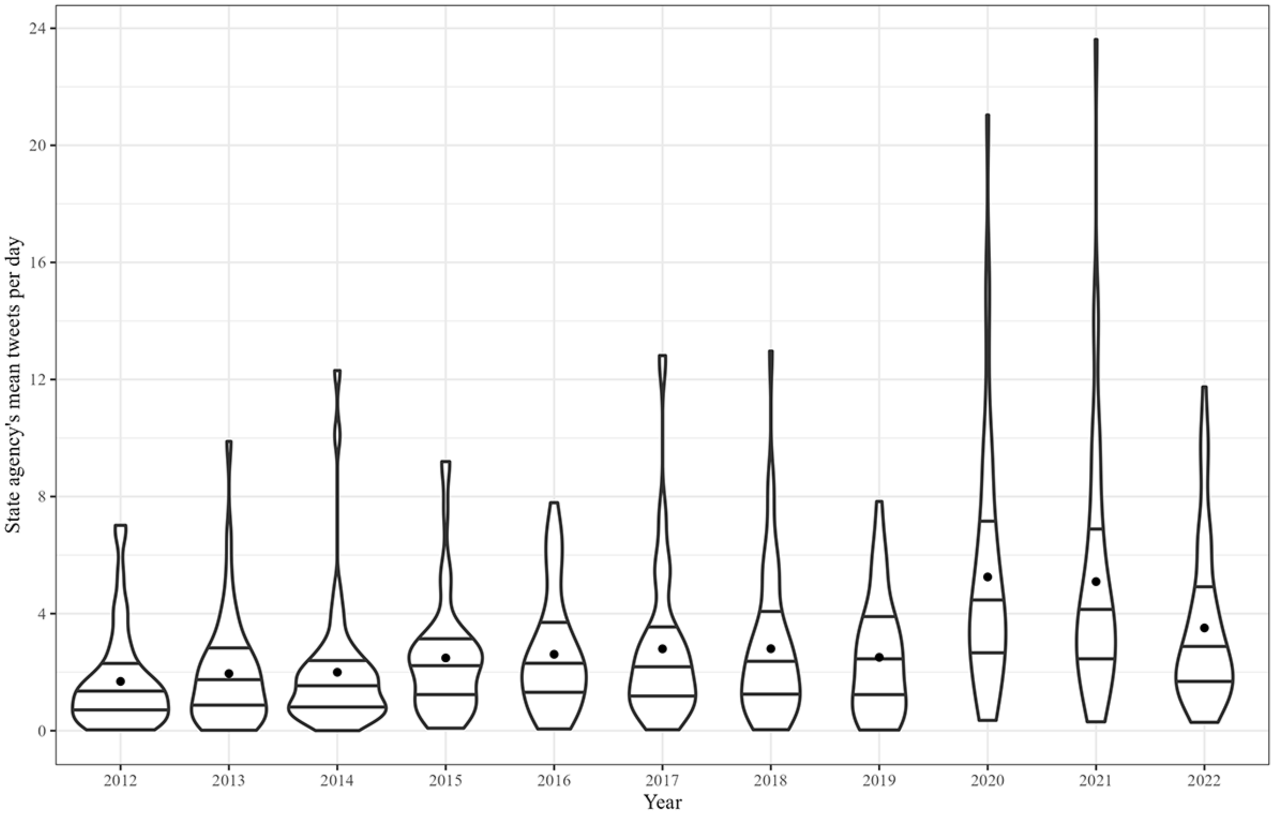


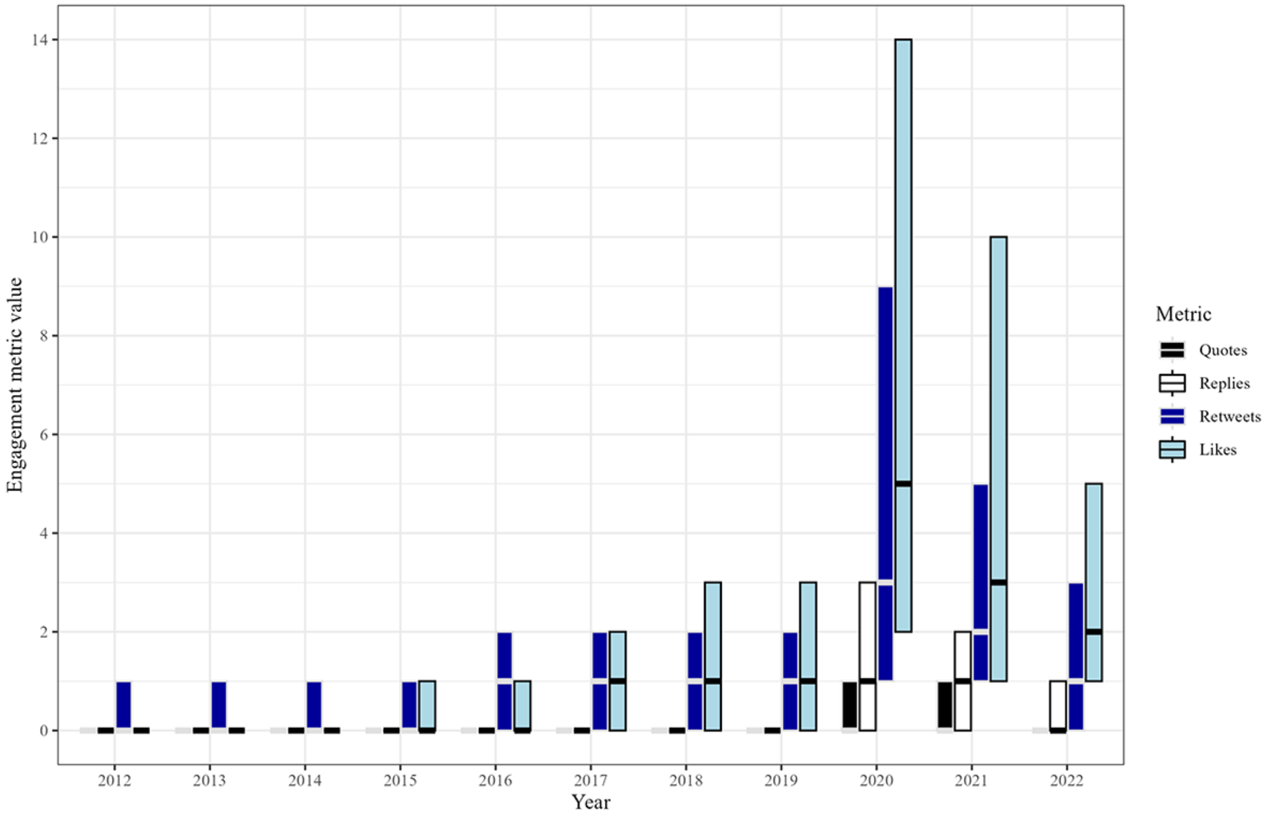
Figure 2. Among an exhaustive enumeration of tweets from US state public health agency accounts, violin plots displaying the distribution of the mean daily tweets per account, each year from 2012 through 2022. IQRs for each year are marked with horizontal lines. The mean of means for each year is marked with a dot.



The percentage of non-retweets with 0 on all audience engagement metrics was 62% (12,965/21,023) in 2012, decreasing to 20% (6465/32,668) in 2019. It reached a minimum of 6% (4123/66,507) in 2020 and 12% (6306/51,950) in 2022.

Figure 3 presents each year's IQR for audience engagement on state agencies' non-retweets (ie, replies, quote tweets, and standalone tweets), summarized in the text below.

Figure 3. Among an exhaustive enumeration of non-retweets from US state public health agency accounts, grouped bar charts showing the IQR of audience engagement metrics, for each year from 2012 through 2022. The engagement metrics include quote tweets, replies, retweets, and likes. Note: medians are marked with a thick horizontal line. The bottom edge of each boxplot marks the 25th percentile. The top edge marks the 75th percentile.



The mean number of likes was <0.6 before increasing to 1.2 in 2016 and then 2.7 in 2019. Its maximum was 17.6 in 2020. It decreased to 5.4 in 2022. The median was 0 before rising to 1 in 2017. Its maximum was 5 in 2020. It decreased to 2 in 2022. The maximum (n=69,798) was on a 2020 tweet about social distancing in Ohio. The second highest (n=12,022) was on a 2020 tweet about face coverings in California.

The mean number of replies was <0.3 before rising to a maximum of 3.5 in 2020. It decreased to 1.8 in 2022. The median was 0 every year except for 2020 and 2021, when it was 1. The maximum (n=2003) was on a 2022 tweet about the COVID-19 pandemic and breastfeeding or chestfeeding from Washington. The second highest (n=1954) was on a 2021 tweet about face coverings in Alabama.

The mean number of retweets was <1 until 2015, rising to 1.8 in 2018. Its maximum was 10.9, in 2020. It decreased to 2.9 in 2022. The median was 0 until rising to 1 in 2016. Its maximum was 3 in 2020. It decreased to 1 in 2022. The maximum (n=39,950) was on the previously mentioned Ohio tweet. The second highest (n=4844) was on a 2020 tweet about COVID-19 cases in Florida.

The mean number of quotes was <0.3 until rising to a maximum of 1.9 in 2020 and decreasing to 0.6 in 2022. The median was 0 for the entire study period. The maximum (n=7041) was on the previously mentioned Ohio tweet. The second highest (n=2552) was on a 2020 tweet about raw meat in Wisconsin.

Persistence

The percentage of non-retweets with at least 1 mention of an account other than the authoring state public health agency decreased from a maximum of 21% (5399/25,267) in 2013 before decreasing to 17% (4437/26,750) in 2014. It then went up and down within that range before dropping off to 13% (9661/72,124) in 2021 and a minimum of 10% (5402/51,950) in 2022.

Among non-retweets, the top-mentioned account each year was @CDCgov. Before 2018, the top 10 accounts typically included at least 2 accounts representing a company or a non-CDC federal government account. In 2018, 9 of the top 10 represented state government agencies or officials, with the top 10 maintaining a similar balance for several years. In 2022, the top 10 included 2 additional CDC accounts and 2 other federal agencies.

The overall percentage of manual retweets decreased from 7% (1729/25,276) in 2012 to 1% (427/32,800) in 2014, then to <0.1% (5/50,024) in 2017 and 0% (0/91,099) in 2021.

Replicability

The percentage of tweets that were formal retweets increased from 17% (4263/25,276) in 2012 to its maximum of 32% (15,955/50,024) in 2017. It decreased to 29% (27,698/94,205) in 2019 and increased slightly in 2020 before decreasing to 17% (10,847/62,797) in 2022.

The percentage of tweets that mentioned a CDC account was 4.2% (1056/25,276) in 2012, increasing to a maximum of 7% (3559/51,132) in 2018 and decreasing to a minimum of 3.5% (2218/62,797) in 2022. The percentage of tweets retweeting a CDC account increased from a minimum of 2.1% (521/25,276) in 2012 to a maximum of 5.4% (2756/51,132) in 2018. It decreased 2.2% (1363/62,797) in 2022.

The percentage of quote tweets increased from 2% (877/41,730) in 2015 to 4% (2129/50,024) in 2017 before decreasing each year to a minimum of 2% (1106/62,797) in 2022. Table 1 displays the retweet and quote tweet data by year.

Table 1. Among an exhaustive enumeration of all tweets from US state public health agency accounts, the portions that were either retweets or quote tweets, as well as the portion mentioning or retweeting a CDC[a] account.

Year	Retweets (%)	Quote tweets (%)	Tweets with ≥1 mention of a CDC account (%)	Retweets of a CDC account (%)
2012	16.9	—[b]	4.2	2.1
2013	17.5	—	4.7	2.4
2014	18.4	—	4.8	2.9
2015	21.8	2.1	4.4	3.2
2016	26.0	3.3	6.0	4.1
2017	31.9	4.3	6.1	4.8
2018	31.8	3.7	7.0	5.4
2019	28.9	3.7	5.4	3.6
2020	29.4	2.5	4.8	3.6
2021	20.1	2.2	3.7	2.6
2022	17.3	1.8	3.5	2.2

[a] CDC: US Centers for Disease Control and Prevention.

[b] Not applicable.

Searchability

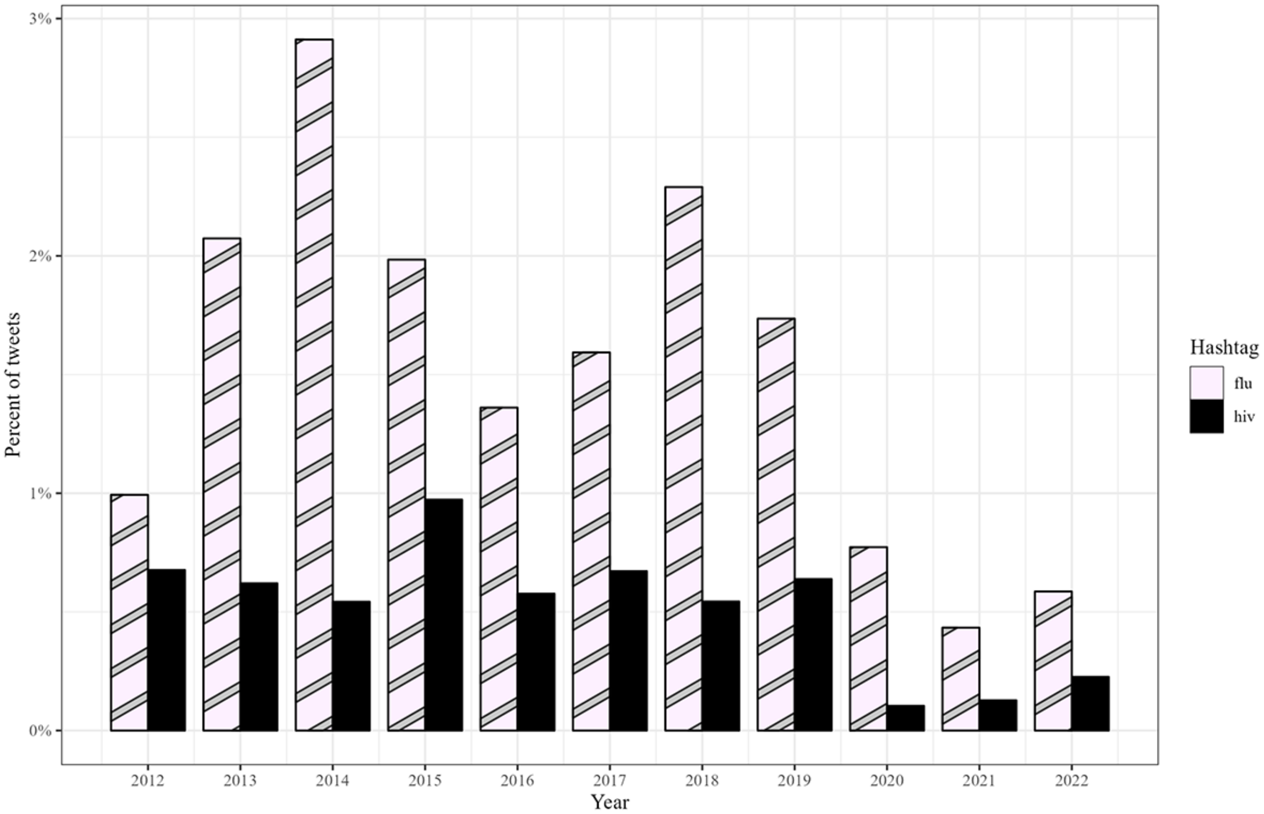
The percentage of tweets with at least 1 hashtag increased from a minimum of 36% (9019/25,276) in 2012 to a maximum of 58% (24,017/41,730) in 2015. It decreased to 38% (23,565/62,797) in 2022.

The hashtags #health, #healthde, #ebola, #vibriovulnificus, #zika, #flu, and #covid19 were each the most common hashtags at 1 point during the study period. The #flu and #covid19 hashtags were each the most common for 3 consecutive years. The top 10 hashtags each year included at least 1 reference to natural disasters or disease outbreaks like #sandy or #ebola. From 2012 through 2016, top hashtags included at least 1 reference to a public health event, such as #cdcchat. In 9 years, the top 10 included #netde, a Delaware-specific hashtag.

Nine of the top 10 hashtags in 2020 and 2021 referred to the COVID-19 pandemic. This number decreased to 5 in 2022. In 2020, 5 of the top 10 hashtags were state-specific COVID-19 hashtags like #covid19ma. This decreased to 2 in 2021 and 2022.

The hashtag #flu was in the top 10 every year in the study period. The hashtag #hiv was in the top 10 from 2012 through 2017 and in 2019. Figure 4 presents the percentage of tweets, including #flu or #hiv, each year.

Figure 4. Among an exhaustive enumeration of tweets from US state public health agency accounts, grouped bar charts displaying the portion using the #flu or the #hiv hashtag, for each year from 2012 through 2022.



Discussion

Principal Findings

We analyzed state public health agencies’ English-language tweets from 2012 through 2022 to provide a complete snapshot of Twitter use in the long term. We used Twitter API data as indicators for agencies’ engagement with key qualities of communication in a networked public: scalability, persistence, replicability, and searchability. We found increasingly fragmented communication over time, with less emphasis on replicability and persistence. This involved a drop in public interactions with other accounts in the form of retweets and quote tweets, which coincided with a dramatic increase in overall tweet volume and audience engagement during the COVID-19 pandemic.

Indicators of scalability and searchability on Twitter display a drastic shift at the beginning of the pandemic. We observed a dramatic increase in Twitter outputs and a focus on state-specific searchability amid a decreased focus on searchability overall. Though some states were especially active, the entire interquartile range of states’ mean Tweets per day shifted up throughout the pandemic. Audience engagement metrics followed a similar pattern, though still modest by marketing standards. Multiple COVID-19 hashtags were prominent from 2020 through 2022, with state-specific pandemic hashtags especially prominent in 2020. This contrasted with the prepandemic prominence of broader, more singular health emergency hashtags like #measles or #ebola.

Indicators of persistence and replicability did not display the same kind of COVID-19 pandemic–era shift. We observed more fragmented public health communication over time, relying less on replicating messages and traceable interactions with other accounts, on a time scale that did not align with the beginning of the pandemic. Mentions of state government officials became more prominent before the pandemic and remained so during the pandemic. The decrease in mentions of external accounts did not occur at the start of the pandemic. The rate of quote tweets decreased over time, starting before the pandemic. The rate of formal retweets followed a similar pattern but with a larger sudden decrease during the second year of the pandemic. The rate of mentions and retweets of CDC accounts decreased over time, starting before the pandemic.

Finally, an overarching finding in this study was between-state variation that did not fall neatly along the lines of geography, population, or a political binary of red and blue states. Florida, West Virginia, and Delaware were all highly active on Twitter for multiple years. Massachusetts, Wisconsin, and Ohio health agency communication styles included prominent use of state-specific pandemic hashtags. Tweets with record engagement were from Washington, Ohio, California, Florida, Wisconsin, and Alabama.

Limitations

Our findings should be interpreted in light of 4 limitations. First, we collected our data retrospectively, meaning Tweet deletion and changes in privacy settings potentially shaped our findings. By the nature of these underlying nonrandom mechanisms, it is not possible to estimate the magnitude of potential missing data or how missingness may have changed over the study period. However, we argue this is not a critical issue for state health agencies, whose communications are part of the public record. This could, however, impact our findings around retweets and quote tweets of other accounts. We argue this is not likely a significant limitation, as the scale of "standalone tweets" we observed toward the beginning of our study period was similar to measures of "original tweets" from Thackeray et al [15]. A second limitation is our use of API indicators to filter tweets by language, relying on Twitter's proprietary processing. Third, we focused solely on text features, meaning we missed trends in multimedia content, for example, reuploaded graphics. Finally, we studied only 1 platform. While our numerical results are not generalizable to other platforms, the patterns we identified in terms of our theoretical framework can be directly compared with data from other networked publics.

Comparison With Previous Work

Our findings about the long-term prominence of self-replies and standalone tweets align with findings on the one-way nature of public health communication on Twitter from the previous decade [14,15]. We argue this could actually be a form of effective tailoring, for example, if known journalists draw information from Tweets for dissemination more often than they draw from press releases. However, there is a need for more research to make sense of our findings, such as exploring whether public health tweets during the pandemic have been helpful for these audiences.

Our findings differ from those of a study from Bradford et al [17], whose work covered a study period contained within our own. This appears to be due to their inclusion of only 1 Twitter account from Maine, in contrast with our inclusion of both Maine's Department of Health and Human Services (DHHS) and its Center for Disease Control and Prevention (MeCDC) accounts. We made this decision after noting the MeCDC serves as the primary public-facing communication arm of the state DHHS. This aligns with the finding from Bradford et al [17] that a Maine account was barely active during their study period. Our conclusion about the overall volume of tweets also differs from Bradford et al [17], due purely to a matter of interpretation. Bradford et al [17] interpreted their observations as signs of low Twitter activity despite most accounts tweeting multiple times per day. We interpreted that same volume of tweets as high activity due in part to the sharp contrast with prepandemic levels. This highlights the need for more research on state health agencies' social media strategies to help interpret our findings and those of other quantitative descriptions of public health communication on social media. Despite our differing conclusions, we echo the call of Bradford et al [17] for more research into the determinants of social media use at state health agencies. For example, we do not know to what degree our findings reflect purposeful decisions within state health agencies or the effects of changes in platform ownership, communication budgets, political factors, or platform design changes over time.

Our findings align with research on US state health departments' use of Facebook. Previous work has found the overall volume of Facebook posts by state health departments to vary [24]. Previous work has also identified different message frames between CDC and state and local health department Facebook posts [25], which could relate to our observed trends in decreasing engagement with CDC tweets and Twitter accounts in the latter half of our study period. Though Twitter and Facebook are prominent social media platforms, future research should quantitatively analyze state health departments' use of other platforms as well.

Conclusions

Our findings warrant further research to elucidate the factors driving the variation we saw between states and over time. This includes research into potential platform-specific communication strategies, health communication training, platform policy, and network norms. Such research should include qualitative content analyses of social media content to elucidate the actual use of social media in public health agencies, not just intended strategies. Further quantitative research should examine the impact of political factors on public health communication on social media. This includes state and federal political factors like budget changes and political divisions around the practice of public health. Further research is also warranted to better understand various audiences' interactions with health agencies in networked publics, as well as their perceptions of state-level health communication.

Our findings raise further questions about what state health agencies' communication should look like and how to support it. For example, it might be impractical for state agencies to target communities that have more contact with local health departments, even when there may be more extensive communication resources at the state level. There is not a robust national infrastructure to fill this gap, much less monitor and evaluate health communication on social media during a public health emergency. This gap is a critical contextual factor to consider when interpreting our findings around between-state variation.

Our study exemplifies the use of theoretical frameworks from media studies to understand trends in public health communication on social media. Though we examined metrics from one social media platform, our use of the networked public framework allows us to draw broader inferences about health communication. This increases the relevance of our findings despite differing features across time and across platforms. This is in contrast to the existing descriptive work of US state health agencies on Twitter. Integration of a theoretical framework should be a key part of future research on public health communication on social media, so as to extend its use beyond the limited scope of one platform's features.

Finally, our study highlights the vulnerability of public health communication in networked publics. A social media platform's use is subject to corporate decisions and community norms. If a platform shuts down or its communities deliberately hide health information, then public health investment in that platform may be suddenly irrelevant. For example, it is unclear whether the drop-off in tweets we observed in 2022 was the result of strategic decisions, insufficient resources, or platform changes. Even the ability to efficiently study public health communication largely depends on corporate policies around data access. Our study highlights the importance of supporting health communication research. This includes regulation to ensure academic access to social media data, as well as building a robust communication infrastructure outside of private platforms.

Acknowledgments

This study was funded by the National Cancer Institute (grant 3U54CA156732-13S1). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the funder. Generative AI was not used in any portion of this study or in writing this manuscript.

Data Availability

The data analyzed during this study are available in a Harvard Dataverse repository [20]. The R code used to analyze the data and calculate descriptive statistics is available in an Open Science Framework repository [21].

Authors' Contributions

SRM contributed to conceptualization, methodology, software, formal analysis, data curation, writing (original draft), and visualization. SMNG managed conceptualization, methodology, supervision, and writing (review and editing). KME handled conceptualization, methodology, supervision, and writing (review and editing). KV contributed to conceptualization, methodology, supervision, and writing (review and editing).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist.
[DOCX File , 32 KB - jmir_v27i1e59786_app1.docx]

Multimedia Appendix 2
Usernames of Twitter accounts included in this study.
[XLS File (Microsoft Excel File), 49 KB - jmir_v27i1e59786_app2.xls]

Multimedia Appendix 3
Exhaustive summary data table.
[XLS File (Microsoft Excel File), 54 KB - jmir_v27i1e59786_app3.xls]

References

1. Boyd D. Social network sites as networked publics: Affordances, dynamics, and implications. In: Papacharissi Z, editor. A Networked Self: Identity, Community, and Culture on Social Network Sites. New York, NY: Routledge; 2011:47-66.
2. Gibson CM. The democratization of medical research and education through social media: The potential and the peril. JAMA Cardiol 2017;2(1):9-11. [doi: 10.1001/jamacardio.2016.4933] [Medline: 27973665]
3. Costanza-Chock S. Design narratives: From TXTMob to Twitter. In: Design Justice: Community-Led Practices to Build the Worlds We Need. Cambridge, MA: The MIT Press; 2020:103-134.
4. Baym NK, Burgess J. Twitter: A Biography. New York, NY: New York University Press; 2020:59-80.
5. 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;111(5):956-964. [doi: 10.2105/AJPH.2021.306154] [Medline: 33734838]
6. Hoffman BL, Colditz JB, Shensa A, Wolynn R, Taneja SB, Felter EM, et al. #DoctorsSpeakUp: Lessons learned from a pro-vaccine Twitter event. Vaccine 2021;39(19):2684-2691 [FREE Full text] [doi: 10.1016/j.vaccine.2021.03.061] [Medline: 33863574]
7. Harris JK, Choucair B, Maier RC, Jolani N, Bernhardt JM. Are public health organizations tweeting to the choir? Understanding local health department Twitter followership. J Med Internet Res 2014;16(2):e31 [FREE Full text] [doi: 10.2196/jmir.2972] [Medline: 24571914]
8. Crisis and emergency risk communication: Working with the media. Centers for Disease Control and Prevention. 2014. URL: https://emergency.cdc.gov/cerc/ppt/CERC_Working_with_the_Media.pdf [accessed 2024-01-04]

9. DePaula N, Hagen L, Roytman S, Alnahass D. Platform effects on public health communication: A comparative and national study of message design and audience engagement across Twitter and Facebook. JMIR Infodemiology 2022;2(2):e40198 [FREE Full text] [doi: 10.2196/40198] [Medline: 36575712]
10. Zhu JM, Sarker A, Gollust S, Merchant R, Grande D. Characteristics of Twitter use by state Medicaid programs in the United States: Machine learning approach. J Med Internet Res 2020;22(8):e18401 [FREE Full text] [doi: 10.2196/18401] [Medline: 32804085]
11. Slavik CE, Buttle C, Sturrock SL, Darlington JC, Yiannakoulias N. Examining tweet content and engagement of Canadian public health agencies and decision makers during COVID-19: Mixed methods analysis. J Med Internet Res 2021;23(3):e24883 [FREE Full text] [doi: 10.2196/24883] [Medline: 33651705]
12. Li L, Aldosery A, Vitiugin F, Nathan N, Novillo-Ortiz D, Castillo C, et al. The response of governments and public health agencies to COVID-19 pandemics on social media: A multi-country analysis of Twitter discourse. Front Public Health 2021;9:716333 [FREE Full text] [doi: 10.3389/fpubh.2021.716333] [Medline: 34650948]
13. Harris JK, Mansour R, Choucair B, Olson J, Nissen C, Bhatt J, Centers for Disease Control and Prevention. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013-2014. MMWR Morb Mortal Wkly Rep 2014;63(32):681-685 [FREE Full text] [Medline: 25121710]
14. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
15. Thackeray R, Neiger BL, Burton SH, Thackeray CR. Analysis of the purpose of state health departments' tweets: Information sharing, engagement, and action. J Med Internet Res 2013;15(11):e255 [FREE Full text] [doi: 10.2196/jmir.3002] [Medline: 24217361]
16. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. Am J Public Health 2013;103(9):1700-1707. [doi: 10.2105/AJPH.2012.301166] [Medline: 23865660]
17. Bradford NJ, Amani B, Walker VP, Sharif MZ, Ford CL. Barely tweeting and rarely about racism: Assessing US state health department Twitter use during the COVID-19 vaccine rollout. Ethn Dis 2022;32(3):257-264 [FREE Full text] [doi: 10.18865/ed.32.3.257] [Medline: 35909637]
18. Stake RE. The art of case study research. In: Research Questions. Thousand Oaks, CA: SAGE Publications, Inc; 1995:3-4.
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. Lancet 2007;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
20. Mendez SR. State public health agency, CDC, and FDA Tweets (2012 through 2022), V1. Harvard Dataverse. 2022. URL: https://doi.org/10.7910/DVN/VX4HK8 [accessed 2024-12-03]
21. Mendez S, Munoz-Najar S, Emmons KM, Viswanath K. Description of US state public health agencies, 2012 through 2022. Open Science Foundation. 2023. URL: https://osf.io/3hec6/ [accessed 2024-12-03]
22. Shu C. Twitter officially launches its “Retweet With Comment” feature. TechCrunch. 2015. URL: https://web.archive.org/web/20150408010147/https://techcrunch.com/2015/04/06/retweetception/ [accessed 2024-12-04]
23. CDC’s Social media channels. Centers for Disease Control and Prevention, Office of Communications. 2023. URL: http://web.archive.org/web/20231109023623/https://www.cdc.gov/digital-social-media-tools/Social-Media-Channels.html [accessed 2023-11-09]
24. Jha A, Lin L, Savoia E. The use of social media by state health departments in the US: Analyzing health communication through Facebook. J Community Health 2016;41(1):174-179. [doi: 10.1007/s10900-015-0083-4] [Medline: 26318742]
25. Meadows CZ, Meadows CW, Tang L. The CDC and state health department Facebook messages: An examination of frames and the extended parallel processing model. Communication Studies 2020;71(5):740-752. [doi: 10.1080/10510974.2020.1819839]

Abbreviations

API: application programming interface
CDC: US Centers for Disease Control and Prevention
DHHS: Department of Health and Human Services
MeCDC: Maine’s Center for Disease Control and Prevention
NIOSH: National Institute for Occupational Safety and Health
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Mavragani; submitted 22.04.24; peer-reviewed by D Boatman, N Hu; comments to author 17.09.24; revised version received 10.10.24; accepted 05.11.24; published 03.01.25.

Please cite as:
Mendez SR, Munoz-Najar S, Emmons KM, Viswanath K
US State Public Health Agencies' Use of Twitter From 2012 to 2022: Observational Study
J Med Internet Res 2025;27:e59786
URL: https://www.jmir.org/2025/1/e59786
doi:10.2196/59786
PMID:39752190

Original Paper

Optimizing Virtual Follow-Up Care: Realist Evaluation of Experiences and Perspectives of Patients With Breast and Prostate Cancer

Sarah Scruton[1], MSc; Geoff Wong[2], MD, PhD; Stephanie Babinski[1], MA, MPH; Lauren R Squires[1,3], MSc; Alejandro Berlin[4,5,6], MSc, MD; Julie Easley[7], PhD; Sharon McGee[8], PhD; Ken Noel[9], BSc; Danielle Rodin[4,5], MSc, MD; Jonathan Sussman[10], MSc, MD; Robin Urquhart[11], PhD; Jacqueline L Bender[1,3,12], PhD

[1]Cancer Rehabilitation and Survivorship, Department of Supportive Care, Princess Margaret Cancer Centre, Toronto, ON, Canada
[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
[3]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[4]Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada
[5]Radiation Medicine Program, Princess Margaret Cancer Centre, Toronto, ON, Canada
[6]TECHNA Institute, University Health Network, Toronto, ON, Canada
[7]Department of Medical Education, Dr Everett Chalmers Regional Hospital, Horizon Health Network, Fredericton, NB, Canada
[8]Department of Medicine, Division of Medical Oncology, The Ottawa Hospital and the University of Ottawa, Ottawa, ON, Canada
[9]The Walnut Foundation, Brampton, ON, Canada
[10]Department of Oncology, McMaster University, Hamilton, ON, Canada
[11]Department of Community Health and Epidemiology, Dalhousie University, Halifax, ON, Canada
[12]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Jacqueline L Bender, PhD
Cancer Rehabilitation and Survivorship
Department of Supportive Care
Princess Margaret Cancer Centre
610 University Ave
Toronto, ON, MG5 2C1
Canada
Phone: 1 416 581 8606
Email: Jackie.Bender@uhn.ca

Abstract

Background: Virtual follow-up (VFU) has the potential to enhance cancer survivorship care. However, a greater understanding is needed of how VFU can be optimized.

Objective: This study aims to examine how, for whom, and in what contexts VFU works for cancer survivorship care.

Methods: We conducted a realist evaluation of VFU among patients with breast cancer and prostate cancer at an urban cancer center during the COVID-19 pandemic. Realist evaluations examine how underlying causal processes of an intervention (mechanisms) in specific circumstances (contexts) interact to produce results (outcomes). Semistructured interviews were conducted with a purposive sample of patients ≤5 years after diagnosis. Interviews were audio-recorded and analyzed using a realist logic of analysis.

Results: Participants (N=24; n=12, 50% with breast cancer and n=12, 50% with prostate cancer) had an average age of 59.6 (SD 10.7) years. Most participants (20/24, 83%) were satisfied with VFU and wanted VFU options to continue after the COVID-19 pandemic. However, VFU impacted patient perceptions of the quality of their care, particularly in terms of its effectiveness and patient centeredness. Whether VFU worked well for patients depended on patient factors (eg, needs, psychosocial well-being, and technological competence), care provider factors (eg, socioemotional behaviors and technological competence), and virtual care system factors (eg, modality, functionality, usability, virtual process of care, and communication workflows). Key mechanisms that interacted with contexts to produce positive outcomes (eg, satisfaction) were visual cues, effective and empathetic communication, and a trusting relationship with their provider.

Conclusions: Patients value VFU; however, VFU is not working as well as it could for patients. To optimize VFU, it is critical to consider contexts and mechanisms that impact patient perceptions of the patient centeredness and effectiveness of their care. Offering patients the choice of in-person, telephone, or video visits when possible, coupled with streamlined access to in-person care when required, is important. Prioritizing and addressing patient needs; enhancing physician virtual socioemotional behaviors and technology competency; and enhancing VFU functionality, usability, and processes of care and communication workflows will improve patient perceptions of the patient centeredness and effectiveness of virtual care.

(J Med Internet Res 2025;27:e65148) doi:10.2196/65148

KEYWORDS

cancer; follow-up; virtual; outcomes; realist evaluation; survivorship

Introduction

According to report “Canadian Cancer Statistics 2023,” 1 in 2 Canadians will be diagnosed with cancer in their lifetime, and two-thirds of Canadians diagnosed with cancer today will live for ≥5 years after their diagnosis [1]. Among the most common cancers in men and women, the 5-year net survival rate is 87% for breast cancer (BC) and 95% for prostate cancer (PC) [1]. While these outcomes reflect advances in early detection and treatment, they have substantially increased the demand for follow-up care services. Optimal follow-up care consists of surveillance for cancer spread, recurrences, or secondary cancers; prevention and management of acute and persistent treatment side effects; and promotion of healthy behaviors to mitigate new and ongoing health concerns [2]. To be effective, this involves ongoing assessment and comprehensive care, including referral to specialized supportive care services to improve quality of life, reduce disability, and restore function [3,4].

In Canada, and internationally, follow-up cancer care is typically provided by oncologists in outpatient secondary or tertiary care settings. Evidence suggests that this model of in-person follow-up care with a specialist is not sustainable and is not working well for patients [5]. Oncology offices are often overcrowded with long wait times, short appointments, and high costs per visit [6]. Furthermore, the high burden of unmet posttreatment supportive care needs suggests that the existing model of follow-up care is not meeting patients’ needs [7,8]. Alternatives such as follow-up with primary care providers are comparable to oncologist-led follow-up in terms of detection of recurrence, survival, and quality of life [9,10]. However, prior work [6,11,12] shows that many survivors of cancer prefer to be seen by an oncologist for follow-up due to concerns about the quality and continuity of care.

One method proposed to improve the experience and effectiveness of cancer follow-up care is virtual care, defined as remote interactions between patients and care providers using technology (ie, video calls, telephone, email, etc) [13]. Virtual follow-up (VFU) could alleviate pressure on clinicians and health care services and better meet patient needs by providing patients with access to timely, convenient, and tailored follow-up care [14]. Before the COVID-19 pandemic, virtual care was infrequently used in Canada, mainly to provide care to rural and remote populations [13], and the evidence of its effectiveness was limited [15]. During the COVID-19 pandemic, there was rapid and widespread adoption of virtual care in Canada and globally [16]. In a prior study, we demonstrated that most patients with cancer were satisfied with VFU during the COVID-19 pandemic, but some did not want VFU to continue after the COVID-19 pandemic [17]. In addition, while virtual visits made care more accessible for some, other studies have shown that virtual care has exacerbated access disparities by widening the digital divide for those who lack the knowledge or technological means to access care virtually [18,19].

To sustain VFU as an option for patients with cancer, we need a better understanding from patients on how, for whom, and in what contexts this model of care works. Patient perspectives on how to optimize VFU care can inform updates to telehealth clinical practice guidelines, such as those offered by the American Association of Clinical Oncology (ASCO) early in the COVID-19 pandemic [20], to ensure patient-centered virtual care. Realist evaluations are ideally suited to answer these questions. Realist evaluations are a proven, theory-driven method that enable a rigorous understanding of context and its influence on causal processes (ie, mechanisms) to explain how and for whom an innovation works best [21,22]. This method is increasingly used to assess complex innovations in health care, particularly those that involve multiple interconnected components and are dependent on individual responses and the wider context for their success [23,24]. Therefore, we conducted a realist evaluation [21] to examine how, for whom, and in what contexts VFU works for patients with BC and PC and conclude with recommendations for optimizing VFU for patients with cancer.

Methods

Ethical Considerations

The study was approved by the University Health Network Institutional Research Ethics Board (ID 21-5397). Informed consent was obtained from all study participants before data collection, and all data were deidentified before analysis. All study participants received a CAD $25 honorarium to compensate for their time. At the time of the study, US $1 was equal to CAD $1.30.

Study Design

We conducted a realist evaluation following the steps by Pawson and Tilley [21] and RAMESES [23] quality and reporting guidelines.

Step 1 involved developing an initial program theory, which is a kind of conceptual framework, to explain how VFU could be delivered to work most effectively. This involved explaining how the underlying processes of VFU care (mechanisms) interact with specific circumstances (contexts) to produce results (outcomes). For this step, we used prior evidence to understand patient views [25] and the effectiveness of virtual versus in-person cancer care [15], theories from technology adoption [26-28] to understand factors that influence the adoption of VFU and theories from patient-provider communication [29] to determine the essential elements of effective clinical encounters, and the 6 domains of health care quality identified by the Agency for Healthcare Research and Quality [30] to understand the attributes of quality VFU. Using these information sources, we created a figure to illustrate how VFU can bring about the desired outcomes. We then held a 2-hour virtual workshop with project team members and stakeholders, consisting of clinicians, researchers, and patient partners (N=10) to obtain their views on the initial theory.

Step 2 involved collecting evidence on the contexts, mechanisms, and outcomes of VFU. For this, we (1) administered a web-based survey to patients at a cancer centre to obtain information on their sociodemographic characteristics, experience, and satisfaction with VFU [17]; (2) purposively recruited a subsample of survey respondents to participate in semistructured interviews to explore their views and experiences with VFU in greater depth; and (3) administered a short follow-up survey to collect information on interview participants' psychosocial well-being.

Steps 3 and 4 involved data integration and synthesis using a realist logic of analysis. The goal of this analysis was to determine whether data were functioning as context, mechanism, or outcome, and if so, within which context-mechanism-outcome configuration (CMOC) [23]. We then held a 2-hour virtual stakeholder workshop with a wider group of stakeholders (N=12) to obtain feedback on the relevance of our findings more broadly.

Setting and Participants

This study was conducted at the Princess Margaret (PM) Cancer Centre, a tertiary, university-affiliated, teaching hospital, which is a part of the University Health Network in Toronto, Canada. Before COVID-19, <1% of visits at PM were conducted virtually [31]. After the declaration of the COVID-19 pandemic, 68% of visits at PM were virtual [31]. The participants in this study comprised a subsample of patients with BC and PC who had completed the PM Virtual Care Evaluation Survey between May and July 2021 [17] and who agreed to participate in a follow-up interview. Interviews occurred between September and November 2021. Survey participants had to meet the following criteria: they must have received a cancer diagnosis, be aged >18 years, have participated in at least 1 virtual appointment (eg, by phone or video) in the last 12 months, and have a valid email address on file. We used the survey data to intentionally select interview participants who varied in age, ethnicity, geographic location, and satisfaction with VFU. In addition, to be eligible to participate in an interview, survey respondents must have agreed to be contacted about a follow-up interview, received a diagnosis of BC or PC within the last 5 years, and completed treatment for cancer or were receiving adjuvant hormone therapy. On the basis of previous research, a sample size of 15 per patient group (30 in total) was estimated to be sufficient [32,33]. However, data collection continued until theoretical saturation [34].

Data Collection

First, data on participants' sociodemographic characteristics and virtual care use and perceptions were extracted from the PM Virtual Care Evaluation Survey to describe the sample and inform each interview. Demographic data included age, whether they were born in Canada (yes or no), English as a first language (yes or no), race or ethnicity (10 response options), highest level of education (high school or less, college or technical school, university undergraduate, or postgraduate), and household income (CAD <$60,000, CAD $60,000-$99,000, or CAD ≥$100,000). Virtual care use and preferences were assessed by asking participants to indicate the type of virtual appointments they had received and would have liked to receive in the past 12 months (phone, video, or both). Patient satisfaction with virtual care was captured using a 5-point Likert response (very satisfied, satisfied, neutral, dissatisfied, or very dissatisfied). Patient desire for virtual care options after the COVID-19 pandemic was captured by asking participants to indicate their level of agreement with the statement "I would like to continue to have virtual options for some of my visits after the COVID-19 pandemic ends," using a 5-point Likert response option (strongly agree, agree, neutral, disagree, or strongly disagree).

Each participant then took part in one 45- to 60-minute semistructured interview based on the realist approach [21,22]. They were asked about their initial expectations and desired outcomes of VFU, how it worked or did not work for them, and how VFU could be made to work optimally. The interviews were conducted by a research coordinator with experience in qualitative interview methods (SB) who was trained in realist methodology by a realist expert on the team (GW). Interviews were audio-recorded and transcribed verbatim.

After the interview, participants completed a brief survey on their level of patient activation, anxiety, and depression. Patient activation was measured using the Patient Activation Measure [35], which scores a patient's knowledge, skills, and confidence for proactively managing their health and health care using a 13-item tool with 4-point Likert scales. The Patient Activation Measure scores (between 0 and 100) were converted into four levels of activation: (1) not believing (patient) activation is important (score of ≤47.0), (2) a lack of knowledge and confidence (score of 47.1-55.1), (3) beginning to take action (score of 55.2-67.0), and (4) taking action (score of ≥67.1). Anxiety was measured using the Generalized Anxiety Disorder 7-item [36], and depression was measured using the Patient Health Questionnaire-9 [37]. Anxiety and depression scores were converted into the following categories: minimal (score of 0-4), mild (score of 5-9), moderate (score of 10-14), moderately severe (depression score only [15-19]), and severe (score of 20-27 on Patient Health Questionnaire-9 or 15-21 on Generalized Anxiety Disorder 7-item). Fear of recurrence was measured using the Fear of Cancer Recurrence Inventory-Short Form [38], a 9-item severity scale that evaluates intrusive thoughts associated with fear of cancer recurrence using 4-point

Likert scales. A score of ≥13 on the Fear of Cancer Recurrence Inventory-Short Form denotes a clinically relevant fear of cancer recurrence.

Data Analysis

A realist logic of analysis was used to analyze interview transcripts [21], with the support of NVivo (version 14; Lumivero). The coding was both deductive (some codes were created in advance informed by the initial program theory) and inductive (some codes emerged from the data). The data were initially sorted into broad conceptual categories or "buckets" by 2 team members independently (SB and SS) through consultation with the lead author (JLB). The buckets were developed following data collection both deductively and inductively based on the initial program theory and what was found in the data. They were further analyzed by 1 team member (SS) to develop preliminary themes and CMOCs. As there were no meaningful differences across the 2 groups of BC and PC participants, the data were analyzed and reported jointly. The themes and CMOCs underwent an intensive refinement process with 3 research team members (JLB, GW, and SS) to ensure they were clear, concise, and representative of the interview data. Repetitive or redundant CMOCs were removed. The data were then used to refine the program theory. Descriptive statistics were used to summarize patient demographic characteristics, virtual care use, patient activation, and psychosocial well-being using R (version 4.1.2; R Foundation for Statistical Computing).

Results

Participant Characteristics

In total, 24 patients were interviewed; of these, 12 (50%) had been diagnosed and treated for BC and 12 (50%) had been diagnosed and treated for PC (Table 1). On average, the participants were aged 65.9 (SD 8.65) years, 100% (24/24) attended postsecondary school, and 71% (17/24) reported a household income of CAD ≥$100,000. Approximately half (11/24, 46%) of the sample identified as racialized, 50% (12/24) were born outside of Canada, and 20% (5/24) had a non-English first language. The most common treatment type was surgery (18/24, 75%), followed by radiation therapy (17/24, 71%) and chemotherapy (9/24, 37%). This was a highly activated sample, with most (18/24, 75%) at level 3 or 4 on the patient activation scale and most with relatively mild or minimal depression (18/24, 75%) or anxiety (20/24, 83%), and most participants (15/24, 63%) reported nonclinical levels of fear of recurrence.

Table 1. Participant characteristics (N=24).

Characteristic	Prostate cancer (n=12)	Breast cancer (n=12)	Total
Demographic characteristics			
Age (y), mean (SD)	65.92 (8.65)	53.25 (14.31)	59.58 (13.41)
Race or ethnicity, n (%)			
Arab or West Asian	0 (0)	1 (8)	1 (4)
Black	1 (8)	2 (17)	3 (13)
East Asian	0 (0)	1 (8)	1 (4)
Jewish	1 (8)	0 (0)	1 (4)
Latin American	0 (0)	2 (17)	2 (8)
Southeast Asian	0 (0)	2 (17)	2 (8)
South Asian	1 (8)	0 (0)	1 (4)
White	9 (75)	4 (33)	13 (54)
Born in Canada , n (%)			
Yes	5 (42)	7 (58)	12 (50)
English as a first language, n (%)			
Yes	10 (83)	9 (75)	19 (79)
Highest level of education received, n (%)			
High school or less	0 (0)	0 (0)	0 (0)
Postsecondary school	12 (100)	12 (100)	24 (100)
Household income (CAD $1.30=US $1), n (%)			
60,000-99,999	2 (17)	2 (17)	4 (17)
≥100,000	8 (67)	9 (75)	17 (71)
Prefer not to say	2 (17)	1 (8)	3 (12)
Clinical and psychosocial characteristics			
Treatments received, n (%)			
Drug or chemotherapy	0 (0)	9 (75)	9 (37)
Hormone therapy	1 (8)	5 (42)	6 (25)
Radiation therapy	6 (50)	11 (92)	17 (71)
Surgery	8 (67)	10 (83)	18 (75)
Other	1 (8)	0 (0)	1 (4)
Anxiety level , n (%)			
Minimal or mild	10 (83)	10 (83)	20 (83)
Moderate or severe	1 (8)	2 (17)	3 (13)
Depression level, n (%)			
Minimal or mild	11 (92)	7 (58)	18 (75)
Moderate, moderately severe, or severe	1 (8)	4 (33)	5 (21)
Level of patient activation , n (%)			
Not believing activation is important	2 (17)	1 (8)	3 (13)
A lack of knowledge or confidence	0 (0)	1 (8)	1 (4)
Beginning to take action	5 (42)	4 (33)	9 (37)
Taking action	4 (33)	5 (42)	9 (37)
Fear of recurrence, n (%)			
Clinical	4 (33)	3 (25)	7 (29)

Characteristic	Prostate cancer (n=12)	Breast cancer (n=12)	Total
Nonclinical	7 (58)	8 (67)	15 (63)
Missing	1 (8)	1 (8)	2 (8)

VFU Use and Preferences

All participants received a phone call appointment (Table 2). Most patients with BC also received a video appointment (20/24, 83%); none of the patients with PC received a video appointment. Overall, most participants (20/24, 83%) were satisfied or highly satisfied with their virtual care visits, and most participants (20/24, 83%) would want VFU options to continue after the COVID-19 pandemic. When asked which type of VFU appointment they would prefer in the future, most patients (15/24, 63%) indicated video appointments.

Table 2. Virtual follow-up (VFU) characteristics (N=24).

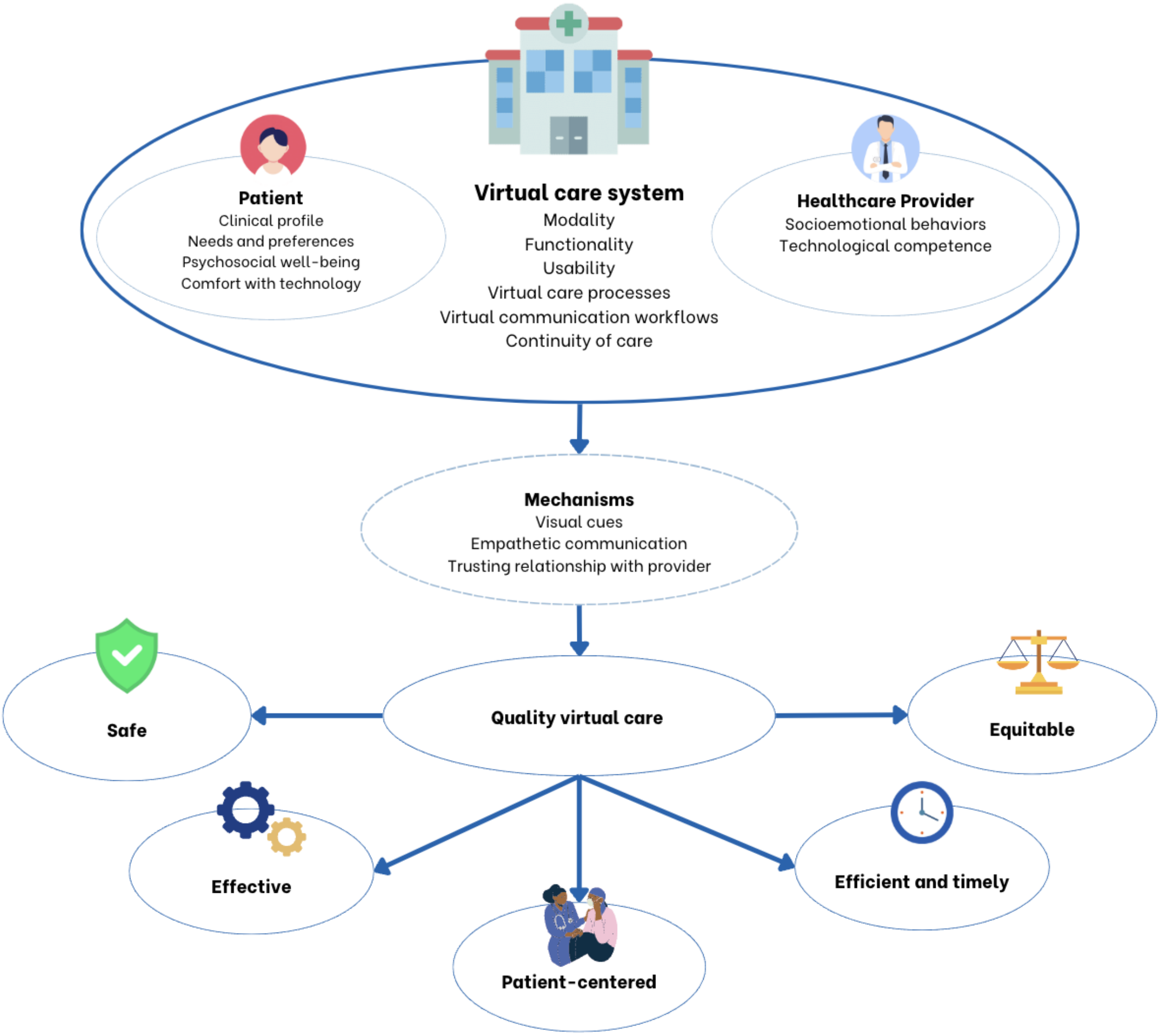
Characteristic	Prostate cancer (n=12)	Breast cancer (n=12)	All
Type of VFU appointment received, n (%)			
Only phone	12 (100)	2 (17)	14 (58)
Only video	0 (0)	0 (0)	0 (0)
Phone+video	0 (0)	10 (83)	10 (42)
Satisfaction with virtual care, n (%)			
Very satisfied	7 (58)	7 (58)	14 (58)
Satisfied	4 (33)	2 (17)	6 (25)
Dissatisfied	1 (8)	3 (25)	4 (17)
Would want the option for VFU after the COVID-19 pandemic, n (%)			
Strongly agree	6 (50)	7 (58)	13 (54)
Agree	4 (33)	3 (25)	7 (29)
Neutral	2 (17)	2 (17)	4 (17)
Disagree	0 (0)	0 (0)	0 (0)
Preferences for virtual care in the future, n (%)			
No preference	1 (8)	4 (33)	5 (21)
Phone	4 (33)	0 (0)	4 (17)
Video	7 (58)	8 (67)	15 (63)

Views and Experiences With VFU

Overview

We identified 8 overarching concepts that captured patients' views on how VFU could be made to work effectively. These 8 concepts consisted of 24 themes and 46 CMOCs (Multimedia Appendix 1). Due to the number of CMOCs that were developed, each CMOC is labeled with a code (eg, A1, A2) found in Multimedia Appendix 1 (for ease of reference). The findings related to each CMOC are labelled with a matching code throughout the Results section. The concepts, themes, and CMOCs were used to refine the program theory (Figure 1).

Figure 1. Refined program theory.



Effectiveness

Effectiveness of VFU appears to be influenced by patient-, physician-, and disease-related contextual factors. For example, VFU was considered ineffective if it replaced physical appointments all together, as physical examinations and speaking to a physician in person reduced patient anxiety about recurrence and management of ongoing side effects (A1). In-person appointments provided visual cues (ie, body language, facial expression, etc) that helped patients understand the meaning of information received from their provider and feel more confident in their care (A2). However, participants felt that this may be partly overcome through video appointments, as video allows patients and care providers to read each other's body language and facial expressions. This may be particularly important for patients experiencing high levels of anxiety about their cancer, as the lack of visual cues with phone calls causes VFU to be insufficiently reassuring (A3). Participants also felt that it was important for physicians to see the visual cues of the patient, as this would enable the physician to pick up on distress or concern that may not be verbally expressed by patients and act accordingly (A4). The final contextual factor described to impact effectiveness of VFU was the perceived loss of the appreciation of the seriousness of cancer. In-person appointments act as a reminder of the seriousness of the condition. Without in-person appointments, patients were concerned that they may become less vigilant in their self-care and self-management (eg, nonadherence to medication or forgetting self-examinations; A5).

Efficiency and Timeliness

If participants felt that the VFU they received was both efficient and timely, they were more likely to be satisfied with their care. These participants felt that VFU would likely improve the overall efficiency of follow-up care, as they believed hospitals could save time and money by scheduling more appointments per day (B1). Participants also believed that greater use of virtual appointments could improve the experience of in-person appointments (when required), as hospital waiting rooms would be less crowded and stressful (B2).

However, some participants expressed concerns that VFU could lead to delays in receiving care. Many described being given very long windows of time for their VFU appointments, which were difficult and frustrating to plan around (B3). In some cases, physicians called outside of these time windows, which led to missed appointments or patients having to speak to someone other than their principal care provider (B4). These types of situations were anxiety inducing for patients, particularly if they had concerns to share with their provider as they feared that their condition would worsen while waiting for a rescheduled appointment (B5). Issues with timeliness may also occur if patients have difficulty booking timely appointments due to long wait times for callbacks or uncertainty about whom to contact (B3).

Patient Centeredness

The largest conceptual bucket by far was related to patient centeredness. Participants described ways in which VFU can be more or less patient-centered compared to in-person care, and how this impacted how well VFU worked for them. This included references to personal values, preferences, and expectations of VFU and how these measured up to their experiences. A patient's preferences and expectations were influenced not only by their unique life circumstances but also by health system characteristics, such as continuity of care. Participants were concerned that the number of care providers involved in their care may increase with VFU due to physicians being busier with more appointments. Participants explained that receiving care from multiple care providers made them feel less cared for because they had less time to form trusting relationships with each provider (C1). This was particularly anxiety inducing for those who were newly diagnosed, as they were particularly vulnerable and in need of a strong physician-patient relationship (C2). In addition, they explained that receiving care by multiple care providers led to poor continuity and quality of care due to a lack of communication between care providers. This was frustrating for participants, as they felt they had to continually repeat their history to their care providers, and they were concerned that their care would be negatively impacted by their physician not having a comprehensive understanding of their case (C3 and C4).

Difficulty forming relationships with care providers was also heightened when participants had never been given the opportunity to meet their provider face-to-face, as they felt this hindered their ability to open up emotionally (C5). Some participants felt that the care they received was more holistic and empathetic when provided face-to-face, and they valued the ability to visually see that they had the full attention of their care provider (C6). Therefore, participants appreciated when their care providers would make use of video calls for their VFU appointments. However, most participants also expressed that their VFU appointments felt rushed compared to in-person appointments, which made them feel less cared for as they were unable to take the time needed to process information and ask questions (C7). Many participants felt their VFU appointments only addressed their clinical or medical issues, as opposed to the personal issues that accompany a cancer diagnosis, which caused them to have unmet needs (C8). This may be because the nature of VFU tends to favor information provision rather than emotional support (C9).

When patients leave active treatment, some described feeling "lost in the system." This may be increased with VFU, as not entering hospitals means that it may be more difficult for patients to access support when needed (C10). Participants felt that VFU may be improved if it included straightforward processes to obtain information, support, and care when needed, which would prevent them from feeling abandoned. This could include the ability to communicate with their care provider via technology (ie, email) in between appointments (C11 and C12). In addition, participants expected their care providers to continue to provide referrals to supportive care programs, but this was not always the case with VFU (C13).

Finally, participants appreciated that VFU improved the convenience of care. Patients did not have to travel to the hospital for their appointment, which saved them time and money (C14). They also appreciated that they could attend their appointments from anywhere, including from the comfort of their own home, which avoided stressful and slow waiting rooms (C15). This also meant they could multitask while waiting for appointments, meaning VFU was less disruptive to their schedule (C15). VFU also allowed family members to be included in VFU appointments, which was particularly valuable when COVID-19 prevented family members from attending in-person appointments (C16). Many participants felt that these advantages of VFU outweighed the benefits of in-person care.

Equity

Follow-up care works best when it is equitable for every patient. Participants noted that VFU may limit access to care for those with certain sociodemographic characteristics such as a disability, older age, or low socioeconomic status. For example, participants expressed concern that someone with hearing loss may face difficulties speaking over the phone, and someone with vision loss may struggle with video appointments (D1). Some participants also considered that some individuals may not have access to fast internet or that those with a lower level of education may be unable to search for medical information on the internet if they are not given time to ask questions during VFU appointments (D2 and D3).

Safety

Participants noted that VFU could improve safety by decreasing their risk of being exposed to infectious diseases such as COVID-19 (E2). However, some participants felt that VFU may be less safe because it removed the ability to be physically examined by an oncologist, potentially resulting in poorer care. Turning elsewhere for care (such as the emergency room) caused participants to feel they were being inadequately cared for (E1).

Patient Characteristics

Whether VFU worked well depended on patient characteristics that impacted their suitability and preferences for virtual care. This included comfort with technology, as some participants described challenges that they and others may have with navigating and using VFU technology (F1). Participants who had experience working within a virtual environment had an easier time using VFU and therefore were more likely to be satisfied with it (F2).

Other patient factors that influenced whether VFU was suitable included clinical characteristics such as mental health, disease characteristics, or comorbidities. Those with complex issues (ie, fear of recurrence, severe side effects, anxiety, and recently completed treatment) reported a greater need for reassurance, acknowledgment, and help coping with their adversities, along with a higher reliance on physical examinations, making them a poorer candidate for VFU (F3). The opposite was true for those who experienced fewer side effects and whose appointments mainly addressed simple routine check-ups, making them ideal candidates for VFU (F4).

At the same time, VFU seemed to overcome access barriers experienced by those who are less likely to seek health care. Participants who described themselves as being reluctant to access the health care system (eg, due to a lack of time, downplaying their condition, not the type to seek care, etc) appreciated the convenience and efficiency of VFU, which they explained motivated them to access health care (F5). Finally, a patient's preference for VFU seemed to depend on their coping style. Those who described themselves as requiring considerable emotional support (ie, emotional coping) felt that VFU did not work as well for them because VFU tends to focus on information provision over emotional support (F6). In contrast, those who did not require as much emotional support (ie, problem-focused coping), appreciated the simple, quick information-provision appointments (F7).

Physician Characteristics

Many participants described the importance of physician competence and bedside manner regardless of whether the appointment was virtual or in person. If a physician did not appear to be comfortable with the VFU technology, participants judged the physician as less competent in providing care and were concerned that this could impact the quality of their care (G1). In addition, if a physician was less empathetic, patient, or comforting during VFU compared to in-person appointments, participants stated that they would not prefer this type of care as they would feel less cared for (G2). Most participants described feeling more rushed and less listened to during phone appointments, as if the appointments were only for the purpose of sharing information such as test results and not for the purpose of addressing emotional concerns (G3). Some participants felt that "webside manner" is something that could be improved with proper training (G4).

Virtual Care System Characteristics

The final factor found to impact whether VFU works well for patients was the characteristics of the virtual care system itself. The most discussed topic was VFU modality, as most patients were not offered the choice between having their VFU appointments via phone or video. Some participants were okay with this, as they felt the VFU modality that they were offered was effective for their needs. However, others were disappointed that they were not given a choice in the modality. They felt that being provided the option for video appointments would have increased the quality of their care as they could have benefitted from face-to-face communication (H1). That said, this was only true if the VFU technology worked well (functionality) and was easy to use (usability). Some participants were frustrated with the VFU technology because it was difficult for them to use and did not work well for them (H2). However, when VFU technology does work well, participants described being appreciative of the fact that they could easily access appointments and test results and communicate with their care providers. Participants felt that physicians should be responsible for offering fallback plans if the VFU technology did not work well, such as what to do when a call drops, so that they need not be fearful of the technology malfunctioning (H3). Some participants noted the lack of virtual care processes and communication channels (virtual communication workflows). This caused issues when they did not know how to contact support staff and schedule appointments, when they did not know if their appointment would occur outside of the scheduled time window, and when they did not know what to do if their VFU call or internet connection was dropped or the technology malfunctioned (H4).

Recommendations

The study findings point to several strategies that health care providers and health care systems can use to enhance the patient centeredness and perceived effectiveness of VFU appointments for patients. We have outlined these patient-derived recommendations in Textbox 1, which align with ASCO's guidelines for telehealth in oncology [38] and provide additional considerations to optimize VFU for patients.

Textbox 1. Patient-derived recommendations for improving virtual follow-up (VFU) care.

1. Offer patients the choice of an in-person or virtual visit
2. Offer patients the choice of virtual visit modality (eg, telephone or video)
3. Ensure there is an easy-to-use and efficient process for patients to schedule an in-person visit if needed
4. Clearly communicate how the appointment can be continued if the technology fails and provide technical support
5. Ensure that virtual visit booking and scheduling processes are flexible enough to cater to patient's changing needs and expectations and reduce wait times where possible
6. Make time to listen to patient concerns and provide emotional support during virtual visits
7. Demonstrate empathy and provide your full attention using active listening skills during virtual visits
8. Demonstrate competency with virtual care technology
9. Ensure patients are provided resources and referrals to survivorship care programs during or after a virtual visit
10. Improve the quality of VFU technology and work flows to create a seamless and reliable care experience for patients

Discussion

Principal Findings

To our knowledge, this is the first realist evaluation of VFU for cancer care. This study has helped to answer for whom and in which contexts VFU is most suitable and has identified ways in which VFU could be optimized to better meet the needs of patients with cancer receiving follow-up care. Overall, the patients with BC and PC in this study described themselves as being satisfied with VFU during the COVID-19 pandemic and wanted VFU options to continue after the COVID-19 pandemic. However, VFU impacted patient perceptions of the quality of their care, particularly in terms of its effectiveness and patient centeredness. Whether VFU worked well for patients depended on patient factors (eg, needs, psychosocial well-being, and technological competence), provider factors (eg, socioemotional behaviors and technological competence), and virtual care system factors (eg, modality, functionality, usability, and virtual processes of care and communication workflows). Key mechanisms that interacted with contexts to produce positive outcomes (eg, satisfaction and reassurance) were visual cues, clear empathetic communication, and a trusting relationship with their provider.

Is Virtual Care Suitable for All Cancer Follow-Up Appointments?

While follow-up appointments have been identified by patients with cancer and health care providers and recommended by clinical practice guidelines as suitable for virtual care delivery because of their relative brevity and simplicity [20,39,40], they can be anxiety provoking and complex depending on the circumstances. For example, moderate to severe "scanxiety" (the fear, stress, and anxiety in anticipation of surveillance tests in follow-up cancer care) is present in as many as 28% of survivors of cancer [41]. In a prior study, we reported that survivors of cancer who experienced distress (anxiety or depression) were less likely to be satisfied with VFU and less likely to want VFU in the future compared to those who did not report distress [17]. This study has revealed that patients experiencing distress may not be satisfied with VFU because the quality of emotional support provided virtually is less effective. Participants felt more comforted during in-person visits because they had more time to discuss their concerns and the type of emotional support received was more effective because the visual cues of the provider conveyed empathy and understanding. Interestingly, a systematic review comparing virtual to in-person cancer care on the psychosocial outcomes of patients with cancer found that virtual care from physicians or nurses was more beneficial for patients during active treatment than follow-up [42]. However, a systematic review comparing virtual versus in-person cancer care found virtual psychosocial counseling to be equally effective for follow-up care [15]. The ASCO telehealth guidelines recommend providing individualized orientation and instruction to patients and care providers on the specific technology that will be used for the virtual care interaction [20]. Our findings highlight a need for care provider training in empathetic virtual communication as well, which has been expressed by care providers themselves [43], accompanied with seamless referrals to psychosocial professionals for follow-up psychosocial issues.

For Whom and in Which Contexts VFU Is Most Suited

Importantly, this study has identified for whom and in which contexts VFU is most suited—a knowledge gap that has been identified by many [39,44,45]. In doing so, the study findings align with and considerably expand upon the recommendations of the European Society of Medical Oncology on which patients should be offered telehealth [46]. In its current form, this study has found that VFU is more suited for patients with cancer who are doing well—physically and emotionally—often termed "the well follow-up patient" [11]. From a clinical perspective, this includes patients with no evidence of disease; with limited or well managed treatment side effects including no-to-mild depression, anxiety, and fear of recurrence; who have problem-focused coping styles; and who require limited emotional support from professionals to deal with the effects of cancer and its treatment. From a sociodemographic perspective, this also includes patients who are comfortable speaking in English (for appointments that are in English); who do not have auditory, visual, cognitive, or physical impairments that limit their effective use of VFU technology; who have access to and experience with VFU technology; and who have the education and skills needed to proactively manage their health and health care. VFU is also well suited for patients with cancer with busy schedules and many competing demands, as they appreciate the convenience of virtual visits. It may even increase health care access for those who are typically reluctant to seek care due to the inconvenience of in-person visits.

Hence, it its current form, if VFU was the only option, it would likely increase disparities in access to care for some patients with cancer. Other studies have demonstrated this to be the case. Notably, a study of health care use during the COVID-19 pandemic in the United States found that Black patients had 0.6 times of the adjusted odds of accessing care through telemedicine compared to White patients [19]. However, subgroup analysis revealed that younger Black women were more likely to access care through telemedicine, particularly for urgent care issues. In contrast, a study of access disparities conducted at the same institution as this study identified no differences in virtual care use based on patient demographics [47]. This may be due to the higher socioeconomic status of their study sample, as only 7% identified as low income. Importantly, they found that regardless of visit type, patients who were structurally marginalized by ethnocultural, situational, and residential status, as well as gender, were less satisfied with their care whether it was virtual or in person. These findings highlight the need for proactive and continuous efforts to identify and intervene to address health disparities in access, experience, and outcomes.

How Health Care Providers Can Optimize Virtual Care for Patients

This study also identified strategies that care providers can use to enhance the patient centeredness and perceived effectiveness of VFU for patients. In fact, health care provider behavior was the key contextual factor that triggered the mechanisms (eg, assessment of visual cues, providing full attention, etc) that led to positive outcomes (eg, patient responses such as trust, reassurance, feeling cared for, and satisfaction). VFU was considered ineffective for patients if care providers did not give patients the option of choosing in-person versus virtual appointments; were not competent using virtual care technology; did not give patients their full attention and time; were less empathetic, patient, and comforting during the virtual visit; or focused on collecting or providing information at the expense of emotional support. Collectively, except for technological competence, these behaviors reflect socioemotional behaviors critical for the effectiveness of the patient-physician encounter. Drawing on social interaction theory, Roberts and Aruguete [48] have shown that patients mostly recognize and react to physician socioemotional behaviors (eg, concern, affection, and attention displayed verbally and nonverbally) due to a lack of understanding of physician task behaviors (eg, explanation of etiology, symptoms, and treatment). In a series of experiments using videotapes of varied levels of physician task versus socioemotional behavior, they demonstrated that high levels of socioemotional behaviors increased patient self-disclosure, trust, satisfaction, and likelihood of recommending the physician. Participants in this study recognized that physician "webside manner" could be improved through training. Likewise, a study of key informants representing Canadian provincial and national health care organizations with expertise in virtual care delivery [43] identified a need to train care providers in how to use technology and integrate virtual care into practice, adapt their clinical skills (including examination skills) to virtual care, build rapport through good communication with patients, and understand when an in-person visit is necessary based on the nature of the appointment and patient contextual factors.

How Health Care Systems Can Optimize Virtual Care for Patients

System transformation is needed to deliver virtual care optimally and equitably to patients. First, health care systems need to enhance virtual care infrastructure and workflows for virtual care to realize its potential for greater efficiency. From a patient perspective, this includes ensuring that VFU technology is accessible, easy to use, and reliable. To increase accessibility, high technology (eg, video) and low technology (telephone and text) VFU solutions are required. Most Canadian studies of virtual care during the COVID-19 pandemic [17,30,49], including this study, reported greater use of telephone than video for virtual care. This was likely due to limited virtual care infrastructure, problems encountered using video, and limited video technology proficiency of patients and care providers. As virtual care technology innovates, low technology options should remain and be offered to patients to ensure equitable access, as broadband connectivity remains inconsistent across Canada [50]. However, the virtual care reimbursement structure needs to change to adequately reimburse care providers for telehealth services. Currently, Ontario's virtual care funding model limits telephone-based care by compensating physicians less for telephone visits compared to in-person visits or video visits [51]. Consequently, physicians might opt to limit their use of telephone visits, potentially leading to reduced access to care for those without high-speed internet access. Establishment of virtual care processes and workflows, along with patient education materials, are needed to seamlessly navigate patients through their virtual care appointment. From a patient perspective, this includes ensuring patients are given the option to choose their virtual visit modality, informed how to use VFU technology optimally, and what to do and who to contact if the technology fails or they require help. Support for technological issues needs to be provided by the system itself, to avoid the responsibility falling on health care providers. These strategies will enhance patient reactions and responses to virtual care, which will result in better outcomes by increasing patient confidence in their ability to use virtual care and patient confidence in the health system's ability to support them if the technology malfunctions or does not work as intended.

Limitations

This study has certain limitations. First, while theory-driven research can provide transferable context-specific understandings of phenomena (in the case of this study, VFU for cancer), inferences about transferability must be tested. In other words, while we may assume that a CMOC operates in a different setting due to the presence of similar mechanisms, this claim should be tested using locally collected primary data. Any testing is aided by the fact that CMOCs are a form of middle-range theory and are expressed in a way that permits empirical testing using primary data. Though our sample was diverse in terms of race, ethnicity, first language, etc, the participants were predominantly high-income, highly educated, and highly activated individuals. As such, our findings may not fully represent the perspective of less activated or less educated patients who may have lower health literacy, be less involved in their care, and face more challenges with virtual care. It is possible that these individuals would differ in their satisfaction with VFU and its suitability as a method of follow-up care. This highlights a need for future research to better understand how VFU could be optimized for populations who face systemic and structural barriers to the use of virtual care. Finally, it may be a limitation that a small subset of participants (6/24, 25%) did not experience an in-person appointment within 12 months of participation. Though it is likely they received an in-person appointment before this date, it is possible that they lacked experience with in-person cancer care with which to compare virtual cancer care. However, we found no obvious differences in the findings between participants who did and did not report in-person appointments during the 12-month period.

Conclusions

Patients value VFU as a part of their care; however, VFU is not working as well as it could for them. To optimize VFU, it is critical to consider the contextual factors and underlying mechanisms that influence patient perceptions of the patient centeredness and effectiveness of their care. Offering patients the choice of in-person, telephone, or video visits when possible, coupled with streamlined access to in-person care when required, is important. Prioritizing and addressing patient needs; enhancing physician online socioemotional behaviors and technology competency; and enhancing VFU functionality, usability, and processes of care and communication workflows will improve the patient centeredness of virtual care. By improving these contextual factors, VFU can be aligned with patient needs and preferences, resulting in effective, efficient, safe, and equitable follow-up care.

Acknowledgments

The authors would like to gratefully acknowledge all participants who took the time to participate in this study as well as the patient partners who supported this study. This study was funded by the Canadian Institutes of Health Research (grant TT7 128272).

Authors' Contributions

JLB, GW, AB, JE, SM, KN, DR, JS, and RU contributed to the conceptualization of the project. JLB, SS, SB, LRS, and AB contributed to data curation. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU contributed to formal analysis. JLB acquired funding. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU contributed to study investigation. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU contributed to the methodology. JLB, SS, and SB were involved in project administration. JLB contributed to resources and software and supervised the study. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU contributed to data validation. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU prepared the original draft. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU reviewed and edited the manuscript.

Conflicts of Interest

JLB is the Associate Editor for *JMIR Cancer*. All other authors have no conflicts to declare.

Multimedia Appendix 1
Conceptual buckets, themes, and context-mechanism-outcome configurations.
[DOCX File , 23 KB - jmir_v27i1e65148_app1.docx]

References

1. Canadian cancer statistics 2023. Canadian Cancer Society. 2023. URL: https://cancer.ca/en/cancer-information/resources/publications/canadian-cancer-statistics-2023 [accessed 2024-03-19]
2. Hewitt M, Greenfield S, Stovall E. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC: National Academies Press; 2005.
3. Silver JK, Baima J, Mayer RS. Impairment-driven cancer rehabilitation: an essential component of quality care and survivorship. CA Cancer J Clin 2013 Sep;63(5):295-317 [FREE Full text] [doi: 10.3322/caac.21186] [Medline: 23856764]
4. Cheville AL, Mustian K, Winters-Stone K, Zucker DS, Gamble GL, Alfano CM. Cancer rehabilitation: an overview of current need, delivery models, and levels of care. Phys Med Rehabil Clin N Am 2017 Feb;28(1):1-17. [doi: 10.1016/j.pmr.2016.08.001] [Medline: 27912990]
5. Urquhart R, Cordoba W, Bender J, Cuthbert C, Easley J, Howell D, et al. Risk stratification and cancer follow-up: towards more personalized post-treatment care in Canada. Curr Oncol 2022 May 03;29(5):3215-3223 [FREE Full text] [doi: 10.3390/curroncol29050261] [Medline: 35621651]
6. Mayer EL, Gropper AB, Neville BA, Partridge AH, Cameron DB, Winer EP, et al. Breast cancer survivors' perceptions of survivorship care options. J Clin Oncol 2012 Jan 10;30(2):158-163 [FREE Full text] [doi: 10.1200/JCO.2011.36.9264] [Medline: 22162585]
7. Paterson C, Robertson A, Smith A, Nabi G. Identifying the unmet supportive care needs of men living with and beyond prostate cancer: a systematic review. Eur J Oncol Nurs 2015 Aug;19(4):405-418. [doi: 10.1016/j.ejon.2014.12.007] [Medline: 25613370]
8. Shakeel S, Tung J, Rahal R, Finley C. Evaluation of factors associated with unmet needs in adult cancer survivors in Canada. JAMA Netw Open 2020 Mar 02;3(3):e200506 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0506] [Medline: 32142127]
9. Grunfeld E, Fitzpatrick R, Mant D, Yudkin P, Adewuyi-Dalton R, Stewart J, et al. Comparison of breast cancer patient satisfaction with follow-up in primary care versus specialist care: results from a randomized controlled trial. Br J Gen Pract 1999 Sep;49(446):705-710 [FREE Full text] [Medline: 10756611]
10. Grunfeld E, Levine MN, Julian JA, Coyle D, Szechtman B, Mirsky D, et al. Randomized trial of long-term follow-up for early-stage breast cancer: a comparison of family physician versus specialist care. J Clin Oncol 2006 Feb 20;24(6):848-855. [doi: 10.1200/JCO.2005.03.2235] [Medline: 16418496]
11. Kwan JY, Croke J, Panzarella T, Ubhi K, Fyles A, Koch A, et al. Personalizing post-treatment cancer care: a cross-sectional survey of the needs and preferences of well survivors of breast cancer. Curr Oncol 2019 Apr;26(2):e138-e146 [FREE Full text] [doi: 10.3747/co.26.4131] [Medline: 31043819]

12. Murchie P, Norwood PF, Pietrucin-Materek M, Porteous T, Hannaford PC, Ryan M. Determining cancer survivors' preferences to inform new models of follow-up care. Br J Cancer 2016 Dec 06;115(12):1495-1503 [FREE Full text] [doi: 10.1038/bjc.2016.352] [Medline: 27802453]
13. Virtual care in Canada: environmental scan from CPAC. Canadian Partnership Against Cancer. URL: https://www.partnershipagainstcancer.ca/topics/virtual-care-canada/ [accessed 2024-04-29]
14. Pham Q, Hearn J, Gao B, Brown I, Hamilton RJ, Berlin A, et al. Virtual care models for cancer survivorship. NPJ Digit Med 2020;3:113 [FREE Full text] [doi: 10.1038/s41746-020-00321-3] [Medline: 32923690]
15. Singh S, Fletcher GG, Yao X, Sussman J. Virtual care in patients with cancer: a systematic review. Curr Oncol 2021 Sep 08;28(5):3488-3506 [FREE Full text] [doi: 10.3390/curroncol28050301] [Medline: 34590602]
16. Husak L, Sovran V, Ytsma A, Comeau M. Impact of the COVID-19 pandemic on virtual care: a major shift for physicians and patients. Healthc Q 2022 Oct;25(3):11-13. [doi: 10.12927/hcq.2022.26948] [Medline: 36412522]
17. Bender JL, Scruton S, Wong G, Abdelmutti N, Berlin A, Easley J, et al. Virtual follow-up care among breast and prostate cancer patients during and beyond the COVID-19 pandemic: association with distress. Cancer Med 2024 Feb;13(4):e6948 [FREE Full text] [doi: 10.1002/cam4.6948] [Medline: 38466233]
18. Virtual care in Canada: progress and potential. Canadian Medical Association. URL: https://www.cma.ca/virtual-care-canada-progress-and-potential-report-virtual-care-task-force-0 [accessed 2024-04-29]
19. 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]
20. Zon RT, Kennedy EB, Adelson K, Blau S, Dickson N, Gill D, et al. Telehealth in oncology: ASCO standards and practice recommendations. JCO Oncol Pract 2021 Sep;17(9):546-564. [doi: 10.1200/OP.21.00438] [Medline: 34319760]
21. Pawson R, Tilley N. Realistic Evaluation. Thousand Oaks, CA: Sage Publication; 1997.
22. Pawson R, Manzano-Santaella A. A realist diagnostic workshop. Evaluation 2012 Apr 22;18(2):176-191. [doi: 10.1177/1356389012440912]
23. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. BMC Med 2016 Jun 24;14(1):96 [FREE Full text] [doi: 10.1186/s12916-016-0643-1] [Medline: 27342217]
24. Wong G, Westhorp G, Greenhalgh J, Manzano A, Jagosh J, Greenhalgh T. Quality and reporting standards, resources, training materials and information for realist evaluation: the RAMESES II project. Health Serv Deliv Res 2017;5(28):2-134 [FREE Full text] [doi: 10.3310/hsdr05280] [Medline: 29072890]
25. Traversa P. 36th annual CAPO conference: advocating for all: psychosocial oncology at the intersections of equity, diversity, and inclusion, 8-10 June 2021. Curr Oncol 2021 Jul 15;28(4):2579-2698 [FREE Full text] [doi: 10.3390/curroncol28040234] [Medline: 34287261]
26. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
27. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
28. 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]
29. Feldman-Stewart D, Brundage MD. A conceptual framework for patient-provider communication: a tool in the PRO research tool box. Qual Life Res 2009 Feb;18(1):109-114. [doi: 10.1007/s11136-008-9417-3] [Medline: 19043804]
30. Six domains of health care quality. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/talkingquality/measures/six-domains.html [accessed 2024-03-19]
31. Berlin A, Lovas M, Truong T, Melwani S, Liu J, Liu ZA, et al. Implementation and outcomes of virtual care across a tertiary cancer center during COVID-19. JAMA Oncol 2021 Apr 01;7(4):597-602 [FREE Full text] [doi: 10.1001/jamaoncol.2020.6982] [Medline: 33410867]
32. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and availability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
33. Ando H, Cousins R, Young C. Achieving saturation in thematic analysis: development and refinement of a codebook. Compr. Physiol 2014 Feb 20;3:03.CP.3.4. [doi: 10.2466/03.cp.3.4]
34. Lincoln YS, Guba EG. Naturalistic Inquiry. Thousand Oaks, CA: Sage Publications; 1985.
35. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
36. 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]
37. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

38. Simard S, Savard J. Fear of cancer recurrence inventory: development and initial validation of a multidimensional measure of fear of cancer recurrence. Support Care Cancer 2009 Mar;17(3):241-251. [doi: 10.1007/s00520-008-0444-y] [Medline: 18414902]
39. Li CZ, Borycki EM, Kushniruk AW. Connecting the world of healthcare virtually: a scoping review on virtual care delivery. Healthcare (Basel) 2021 Oct 05;9(10):1325 [FREE Full text] [doi: 10.3390/healthcare9101325] [Medline: 34683005]
40. Abdelmutti N, Powis M, Macedo A, Liu Z, Bender JL, Papadakos J, et al. Virtual cancer care beyond the COVID-19 pandemic: patient and staff perspectives and recommendations. JCO Oncol Pract 2024 May;20(5):643-656. [doi: 10.1200/OP.23.00254] [Medline: 38266201]
41. Bui KT, Liang R, Kiely BE, Brown C, Dhillon HM, Blinman P. Scanxiety: a scoping review about scan-associated anxiety. BMJ Open 2021 May 26;11(5):e043215 [FREE Full text] [doi: 10.1136/bmjopen-2020-043215] [Medline: 34039571]
42. Caminiti C, Annunziata MA, Di Giulio P, Isa L, Mosconi P, Nanni MG, et al. Psychosocial impact of virtual cancer care through technology: a systematic review and meta-analysis of randomized controlled trials. Cancers (Basel) 2023 Mar 31;15(7):2090 [FREE Full text] [doi: 10.3390/cancers15072090] [Medline: 37046750]
43. Curran V, Hollett A, Peddle E. Training for virtual care: what do the experts think? Digit Health 2023;9:20552076231179028 [FREE Full text] [doi: 10.1177/20552076231179028] [Medline: 37274369]
44. Young K, Xiong T, Pfisterer KJ, Ng D, Jiao T, Lohani R, et al. A qualitative study on healthcare professional and patient perspectives on nurse-led virtual prostate cancer survivorship care. Commun Med (Lond) 2023 Nov 02;3(1):159 [FREE Full text] [doi: 10.1038/s43856-023-00387-6] [Medline: 37919491]
45. Turner K, Bobonis Babilonia M, Naso C, Nguyen O, Gonzalez BD, Oswald LB, et al. Health care providers' and professionals' experiences with telehealth oncology implementation during the COVID-19 pandemic: a qualitative study. J Med Internet Res 2022 Jan 19;24(1):e29635 [FREE Full text] [doi: 10.2196/29635] [Medline: 34907900]
46. Curigliano G, Banerjee S, Cervantes A, Garassino MC, Garrido P, Girard N, et al. Managing cancer patients during the COVID-19 pandemic: an ESMO multidisciplinary expert consensus. Ann Oncol 2020 Oct;31(10):1320-1335 [FREE Full text] [doi: 10.1016/j.annonc.2020.07.010] [Medline: 32745693]
47. Safavi AH, Lovas M, Liu ZA, Melwani S, Truong T, Devonish S, et al. Virtual care and electronic patient communication during COVID-19: cross-sectional study of inequities across a Canadian tertiary cancer center. J Med Internet Res 2022 Nov 04;24(11):e39728 [FREE Full text] [doi: 10.2196/39728] [Medline: 36331536]
48. Roberts CA, Aruguete MS. Task and socioemotional behaviors of physicians: a test of reciprocity and social interaction theories in analogue physician-patient encounters. Soc Sci Med 2000 Feb;50(3):309-315. [doi: 10.1016/s0277-9536(99)00245-2] [Medline: 10626757]
49. Watson L, Qi S, Delure A, Link C, Photitai E, Chmielewski L, et al. Virtual cancer care during the COVID-19 pandemic in Alberta: evidence from a mixed methods evaluation and key learnings. JCO Oncol Pract 2021 Sep;17(9):e1354-e1361 [FREE Full text] [doi: 10.1200/OP.21.00144] [Medline: 34351822]
50. Report 2: connectivity in rural and remote areas. Office of the Auditor General of Canada. URL: https://www.oag-bvg.gc.ca/internet/English/parl_oag_202303_02_e_44205.html [accessed 2023-07-19]
51. Virtual care 1: comprehensive and limited virtual care services. Ontario Ministry of Health. URL: http://www.ontario.ca/document/education-and-prevention-committee-billing-briefs/virtual-care-1-comprehensive-and-limited [accessed 2023-07-19]

Abbreviations

ASCO: American Association of Clinical Oncology
BC: breast cancer
CMOC: context-mechanism-outcome configuration
PC: prostate cancer
PM: Princess Margaret
VFU: virtual follow-up

Edited by A Mavragani; submitted 06.08.24; peer-reviewed by K Le Du, J Li; comments to author 08.10.24; revised version received 28.10.24; accepted 07.11.24; published 03.01.25.

Please cite as:
Scruton S, Wong G, Babinski S, Squires LR, Berlin A, Easley J, McGee S, Noel K, Rodin D, Sussman J, Urquhart R, Bender JL
Optimizing Virtual Follow-Up Care: Realist Evaluation of Experiences and Perspectives of Patients With Breast and Prostate Cancer
J Med Internet Res 2025;27:e65148
URL: https://www.jmir.org/2025/1/e65148
doi:10.2196/65148
PMID:

Original Paper

Stress Reduction in Perioperative Care: Feasibility Randomized Controlled Trial

Haridimos Kondylakis[1*], PhD; Irene Alice Chicchi Giglioli[2*], PhD; Dimitrios Katehakis[1*], MSc; Hatice Aldemir[2*], PhD; Paul Zikas[3*], PhD; George Papagiannakis[1*], PhD; Santiago Hors-Fraile[2*], PhD; Pedro L González-Sanz[2*], MD; Konstantinos Apostolakis[1*], PhD; Constantine Stephanidis[1*], Prof Dr; Francisco J Núñez-Benjumea[4*], PhD; Rosa M Baños-Rivera[5*], Prof Dr; Luis Fernandez-Luque[2*], PhD; Angelina Kouroubali[1*], PhD

[1]FORTH-ICS, Heraklion, Greece

[2]Adhera Therapeutics (United States), Wake Forest, CA, United States

[3]ORamaVR SA, Heraklion, Greece

[4]Innovation & Data Analysis Unit, Institute of Biomedicine of Seville, IBiS/Virgen Macarena University Hospital/CSIC/University of Seville, Seville, Spain

[5]Polibienestar Institute, University of Valencia, Valencia, Spain

[*]all authors contributed equally

Corresponding Author:
Haridimos Kondylakis, PhD
FORTH-ICS
N Plastira 100
Vassilika Vouton
Heraklion, GR70013
Greece
Phone: 30 2810391449
Email: kondylak@ics.forth.gr

Abstract

Background: Patients undergoing surgery often experience stress and anxiety, which can increase complications and hinder recovery. Effective management of these psychological factors is key to improving outcomes. Preoperative anxiety is inversely correlated with the amount of information patients receive, but accessible, personalized support remains limited, especially in preoperative settings. Face-to-face education is often impractical due to resource constraints. Digital health (DH) interventions offer a promising alternative, enhancing patient engagement and empowerment. However, most current tools focus on providing information, overlooking the importance of personalization and psychological support.

Objective: This study aimed to assess the viability of a DH intervention known as the Adhera CARINAE DH Program. This program is specifically designed to offer evidence-based and personalized stress- and anxiety-management techniques. It achieves this by using a comprehensive digital ecosystem that incorporates wearable devices, mobile apps, and virtual reality technologies. The intervention program also makes use of advanced data-driven techniques to deliver tailored patient education and lifestyle support.

Methods: A total of 74 patients scheduled for surgery across 4 hospitals in 3 European countries were enrolled in this study from September 2021 to March 2022. Surgeries included cardiopulmonary and coronary artery bypass surgeries, cardiac valve replacements, prostate or bladder cancer surgeries, hip and knee replacements, maxillofacial surgery, and scoliosis procedures. After assessment for eligibility, participants were randomized into 2 groups: the intervention group (n=23) received the Adhera CARINAE DH intervention in addition to standard care, while the control group (n=27) received standard care alone. Psychological metrics such as self-efficacy, self-management, and mental well-being were assessed before and after the intervention, alongside physiological markers of stress.

Results: The intervention group demonstrated significant improvements across several psychological outcomes. For example, Visual Analogue Scale Stress at the hospital improved at admission by 5% and at hospital discharge by 11.1% and Visual Analogue Scale Pain at admission improved by 31.2%. In addition, Hospital Anxiety and Depression Scale Anxiety after surgery improved by 15.6%, and Positive and Negative Affect Scale-Negative at hospital admission improved by 17.5%. Overall, patients in the intervention study spent 17.12% less days in the hospital. Besides these individual scores, the intervention group shows more positive relationships among the psychological dimensions of self-efficacy, self-management, and mental well-being, suggesting that the CARINAE solution could have a positive effect and impact on the reduction of stress and negative emotions.

Conclusions: Our results provide an important first step toward a deeper understanding of optimizing DH solutions to support patients undergoing surgery and for potential applications in remote patient monitoring and communication.

Trial Registration: ClinicalTrials.gov NCT05184725; https://clinicaltrials.gov/study/NCT05184725

International Registered Report Identifier (IRRID): RR2-10.2196/38536

(J Med Internet Res 2025;27:e54049) doi:10.2196/54049

KEYWORDS

CARINAE; digital health; patient empowerment; stress and anxiety management; mHealth; virtual reality; feasibility; perioperative; randomized controlled trial; surgery; risk; patient empowerment; wearable devices; patient education; mobile app; psychological; self-efficacy; self-management; well-being; patient monitoring

Introduction

Patients undergoing surgical operations commonly experience symptoms of severe stress, anxiety, and fear due to the potentially threatening nature of surgeries [1,2]. Family caregivers, who play a crucial role in caring for these patients, also face emotional distress and physical challenges [3]. Surgeons and health care professionals use various stress coping strategies, as stressors can affect surgical performance and lead to complications [4]. Psychological support and patient education have proven effective in reducing stress and anxiety in surgical settings [5,6]. Providing information about the surgical procedure is essential for both patients and caregivers, as it helps decrease anxiety levels and surgical complications [7,8]. Patient empowerment, where patients actively participate in managing their diseases, leads to better self-care management and improved outcomes in terms of satisfaction, cost, health status, and function [9,10]. Addressing caregiver strain is also vital for pediatric patients [11] and those with special health care needs [12].

Being in unfamiliar surroundings and facing preoperative requirements can add to feelings of confusion and stress. It is common for individuals anticipating surgery to experience uncertainty, fear, hesitation, and anxiety. These emotions can affect patients' well-being and their ability to adhere to instructions throughout their perioperative journey. Research shows [13] that patient experience is closely tied to emotional health and surgical outcomes. Failure to address anxiety and stress adequately can lead to unnecessary discomfort, prolonged hospital stays, higher costs, and suboptimal clinical results.

Preventing presurgical anxiety can significantly contribute to positive outcomes in terms of health and well-being. Various strategies and techniques have been used to manage preoperative stress and anxiety, ranging from simple methods like listening to music and basic relaxation techniques to more involved approaches such as providing information, consulting with nurses, and using advanced interventions like patient education programs and modern information and communication technology tools [14,15].

For instance, allowing preoperative patients the opportunity to listen to music before surgery can prove to be an effective intervention in reducing anxiety and aiding patients in coping with what could potentially be a challenging or stressful procedure. Music has demonstrated its ability to alleviate anxiety among preoperative patients on numerous occasions [8]. However, it is noteworthy that only a small fraction of patients receive adequate stress relief support before undergoing surgery.

Digital health (DH) interventions have significantly supported enhancing health condition awareness and mental health management, using both nonimmersive systems, for example, mobile apps, and immersive systems, for example, virtual reality (VR) [16]. These interventions have shown promise in supporting patients in managing anxiety, stress, and pain [17-19]. Especially for VR, studies [20-23] have demonstrated benefits in various health care areas, including stress and pain reduction [23,24], medical practitioner training, patient counseling, cognitive rehabilitation, physical therapy, dentistry, mental health management, and surgery, as well as in managing pain and anxiety in pediatric patients [21,22,25]. VR's distractive properties make it an efficient tool for stress relief and pain management [26]. In addition, VR has been shown to reduce perceptions of anxiety in preoperative patients [27].

On the other hand, DH interventions generate vast amounts of data that can be used for personalization through artificial intelligence techniques [28-33]. These techniques capitalize on remote activity recognition and monitoring in order to provide recommendations, enabling the personalization of patient interventions based on their unique behavioral and health needs [34,35]. Perioperative stress varies greatly among patients, depending on the severity of their illness and the type of surgery required. Personalized interventions targeting stress management should take these individual differences into account, incorporating health behavior theories. Health Recommender Systems, like the I-Change model, guide the personalization of educational and behavioral interventions [36,37].

Independent of the progress in the domain, the feasibility of combining VR with mobile-based technologies as the primary means of delivering a DH intervention remains poorly understood [38]. This study aims to explore the feasibility of a DH intervention named CARINAE, which combines evidence-based perioperative stress management, anxiety, and pain relief techniques grounded in behavioral science. The intervention leverages mobile and VR technologies to help patients manage stress and anxiety during surgery, promoting healthy recovery, extensively reported already in another paper [39]. Through self-reported measurements and a control group, the study assesses the impact of the CARINAE platform on perioperative stress, aiming to determine the effectiveness of the intervention.

Methods

Overview

In this section, we summarize the design and the protocol followed for the intervention. For more information on the details of the protocol and the technological solution, the interested reader is forwarded to the relevant paper [39].

Ethical Considerations

This study was approved by the Institutional Review Board of the four hospitals participating in the study (Hospital Universitario Reina Sofia, Spain; Instituto di Ricovero e Cura per Anziani, Italy; Sant Joan de Déu Hospital, Spain; and Fundació Parc Taulí, Spain) and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. Additionally, measures were taken to ensure the confidentiality and anonymity of patient data throughout the study. The study posed no significant risks to participants, and they were free to withdraw at any time without affecting their standard of care. Finally, no financial compensation was provided to the participants of the study.

Study Setting

The study, following ethical boards' approval, was conducted at the following European hospitals: (1) Hospital Universitario Reina Sofia (SAS; Spain)—Cardiothoracic Surgery Department; (2) Instituto di Ricovero e Cura per Anziani (INRCA; Italy)—Urology Department; (3) Sant Joan de Déu Hospital (SJD; Spain)—Orthopedics and Traumatology Department for Children; and (4) Fundació Parc Taulí (Parc Taulí; Spain)—Orthopedics and Traumatology Department for Adults. A fifth hospital had granted ethical approval to conduct the study; however, due to external reasons, no patients were recruited from this hospital, and so it did not participate.

The recruitment process started in September 2021 and finished in March 2022.

Clinical investigators prescreened the eligibility of the participants that they had available in their pool of participants that were proposed for one of the surgeries in the inclusion criteria that are reported in the sequel. Whenever a potentially eligible participant was identified, he or she was invited to the study, either by phone or during the routine consultation, whatever was more convenient. As soon as the patient showed interest in taking part in the study, she or he was referred to the research coordinator of the study who facilitated him or her by providing the patient information letter and the informed consent form and who solved any questions and concerns that the patient might have. Upon the signature of the informed consent, the patient was considered recruited for the trial.

Study Design

The CARINAE DH platform was tested in a multicentric trial conducted in 4 clinical settings across two European countries. The trial used a stratified randomized controlled design and aimed to address two main research questions.

RQ1: The first question seeks to determine the extent to which CARINAE impacts patients' stress, anxiety, and pain levels compared to those who receive the standard of care only. Additionally, the study investigated the effects on secondary outcomes, such as well-being and overall quality of life. As a side effect, the study investigated the impact of CARINAE on caregivers' stress and anxiety based on the groups they were assigned to.

RQ2: The second research question focuses on evaluating the overall usability of the CARINAE solution based on feedback and experiences from patients, caregivers, and health care professionals involved in the trial.

Eligibility Criteria

Inclusion Criteria

Participants aged 12 to 65 years who underwent various surgeries, including cardiopulmonary bypass surgery, coronary artery bypass surgery, cardiac valve replacement, prostate, kidney, or bladder cancer surgery, hip or knee replacement, maxillofacial surgery, orthognathic surgery, or scoliosis. Adult participants were required to have an Android smartphone and demonstrate basic digital literacy, while children's caregivers also needed to have an Android smartphone and basic digital literacy.

Exclusion Criteria

Participants who could not provide informed consent, communicate effectively in the native language, demonstrate basic digital literacy, exhibited symptoms of dementia, had allergies to dedicated wearable materials like steel and silicone, were pregnant, or were already enrolled in another clinical trial.

Interventions

Overview

The eligible participants were randomly allocated to either the experimental group or the control group using block randomization with a block size of 4. The randomization process was facilitated through the Sealed Envelope web-based tool [31]. The DH solution consisted of three distinct components: a mobile app, a VR component, and a clinical web application. Participants in the experimental group used the first and second components, while the third component was exclusively used by health care professionals for those patients. Both the control and experimental groups underwent the same visit schedule.

Control Group

The control group received standard care, which included four visits with the health care provider. During these visits, patients in the control group received instructions on diet and healthy lifestyle habits. In current health care settings, it is not common to provide patients with stress and anxiety relief support during the perioperative period. Assessments for the control group took place during the following visits: (1) the initial visit, where the health care provider communicated the need for surgery to the patient (2-4 weeks before the surgery); (2) hospital admission, which occurred 1-3 days before the surgery; (3) hospital discharge, approximately 1 week after the surgery; and (4) remote follow-up 14 days after the surgery. After each visit, patients were administered several questionnaires capturing primary and secondary outcomes, as well as covariates.

Experimental (Intervention) Group

The participants in the experimental group received the digital solution CARINAE during the first visit, along with training on how to use the tool effectively. They were allowed to take CARINAE home and use it as frequently as desired. Following each of the 4 standard care visits, the experimental group completed the same questionnaires as the control group.

Sample Size Calculation

The study was a feasibility clinical trial, with the number of participants established before the beginning of the project by the clinics. According to the study protocol reported by Kondylakis et al [39], participants were randomized to achieve a balance between the 2 groups (intervention and control groups) according to the type of surgery and baseline characteristics (covariates). Considering a 1:1 random allocation, significance level of .05 (2-sided), and 80% power, 60 participants were needed to detect a 10% difference in stress, anxiety, and pain between the intervention and control groups. Sample size calculations are estimated using G*Power (version 3.1.9.2; University of Dusseldorf).

Outcome Variables

The primary outcome variables in this study were stress, anxiety, and pain, which were measured using paper-and-pencil questionnaires. The secondary outcome variables included overall quality of life, emotional status, mental well-being, self-efficacy perception, and patient activation during and after the hospital stay.

For assessing the primary outcome variables, the following questionnaires were used after each standard care visit and administered on paper:

- Patients' and caregivers' self-reported stress measured using a Visual Analogue Scale (VAS) at baseline, admission for surgery, hospital discharge, and 2 weeks after surgery [40].
- Patients' self-reported pain measured using a VAS at baseline, admission for surgery, hospital discharge, and 2 weeks after surgery [41].
- Patients' Hospital Anxiety and Depression (HADS) measured at admission for surgery, hospital discharge, and 2 weeks after surgery [42].

For assessing the secondary outcome variables, the following questionnaires were used after each standard care visit and administered on paper:

- Patients' health-related quality of life measured using the EQ-5D-3L questionnaire at baseline, admission for surgery, and clinical discharge [43].
- Patients' emotional status measured using the Positive and Negative affect Scale (PANAS) at baseline, admission for surgery, hospital discharge, and 2 weeks after surgery [44].
- Patients' and caregivers' mental well-being measured using the "Short Warwick Edinburgh Mental Well-Being Scale" (SWEMWBS) at baseline and 2 weeks after surgery [45].
- Patients' and caregivers' self-efficacy measured using the General Self-Efficacy in short form (GSE) questionnaire at baseline and 2 weeks after surgery [46].
- Patients' activation status measured using the Patient Activation Measure short form (PAM-13) at baseline, admission for surgery, and 2 weeks after surgery [47].

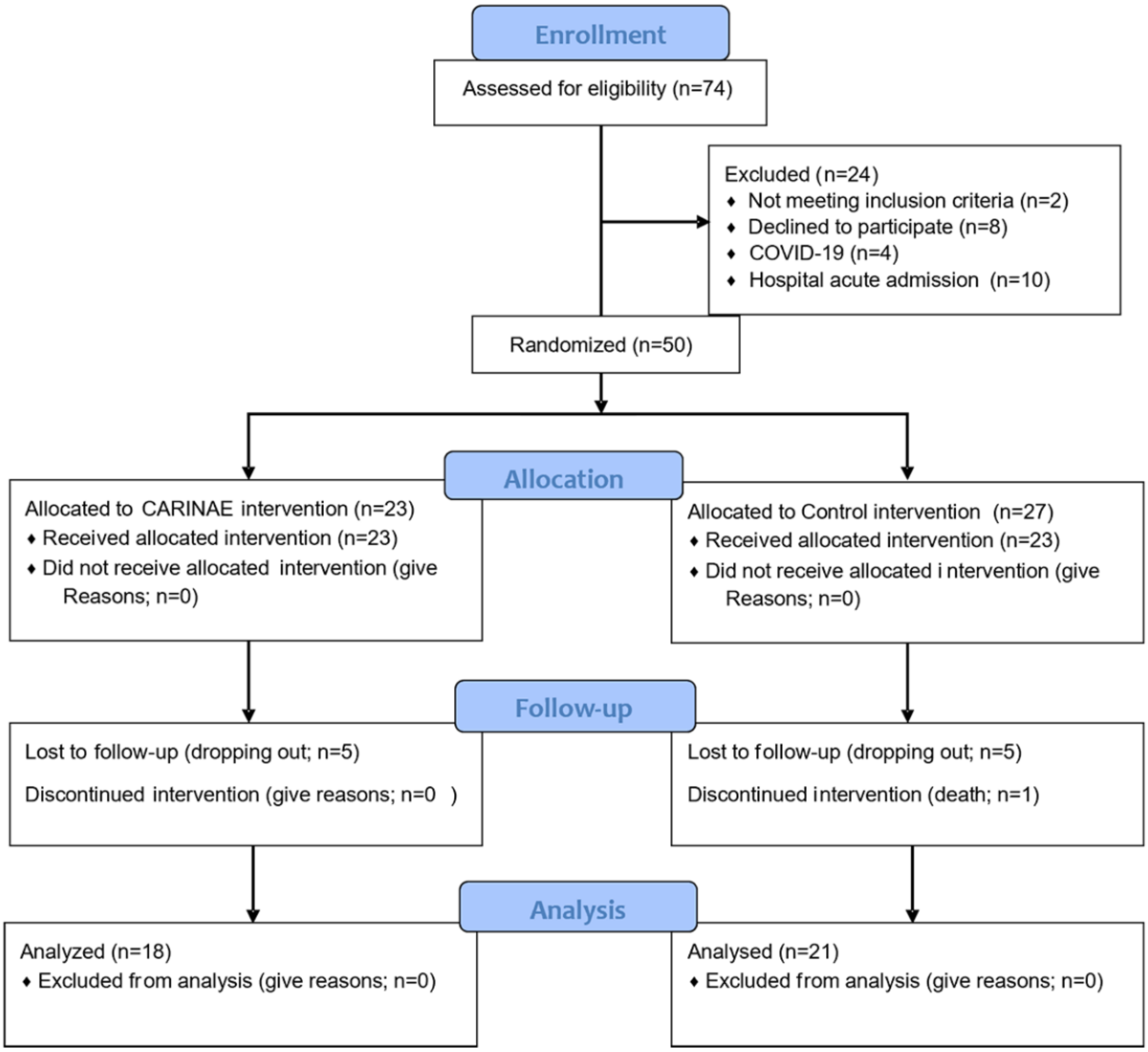
All these questionnaires have proven to be reliable and effective in measuring the respective variables.

Results

Overview

The diagram of the study is shown in Figure 1. From September 2021 to March 2022, a total of 74 patients were assessed for eligibility and 24 patients were excluded according to the following: (1) not meeting inclusion criteria (n=2); (2) declined to participate (n=8); (3) COVID-19 urgency (n=4); and (4) hospital acute admission (n=10). A total of 50 patients participated in the study, 23 patients were allocated to the CARINAE group (intervention group) and 27 participants were allocated to the control group. In total, 10 patients dropped out of the study (5 patients in the intervention group and 5 patients in the control group). In addition, 1 control group patient died during the surgery. A total of 39 patients finalized the trial and their data have been analyzed (21 patients in the control group and 18 patients in the intervention group; see the CONSORT [Consolidated Standards of Reporting Trials] flowchart in Figure 1 and CONSORT checklist in Multimedia Appendix 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of the trial.



The intervention group consisted of 18 patients (10 male and 8 female patients; mean age 45.38, SD 26.2, range 12-91 years) and the control group consisted of 21 patients (11 male and 10 female patients; mean age 56.66, SD 27, range 12-91 years). Furthermore, intervention group surgeries included 1 patient undergone coronary bypass, 1 cardiac valve replacement, 4 scoliosis, 2 hip and 2 knee replacements, 3 prostate cancer, 2 bladder cancer surgeries, and 3 maxillofacial surgery types (see Multimedia Appendix 2 for sociodemographic characteristics).

The control group consisted of 3 patients undergone cardiac valve replacement, 2 scoliosis, 2 hip replacements, 4 knee replacements, 2 prostate cancer, 4 bladder cancer, and 4 maxillofacial surgery types (see Multimedia Appendix 2 for sociodemographic characteristics). The baseline table highlights a potential bias with regard to the age of the participants in both intervention and control groups (ca. 10 years), see the impact of that in the Discussion/Limitations section.

In addition, 22 caregivers participated in the study (11 in the control group and 11 in the intervention group). A total of 5 caregivers dropped out of the study (4 in the control group and 1 in the intervention group). A total of 17 caregivers finalized the trial and their data have been analyzed. This difference in drop out might be resulting from the lack of incentive to stay in the control group.

Finally, 12 health care professionals answered the three questionnaires (VAS, SWEMWBS, and Caregiver GSE), and their data were analyzed; 5 health care professionals belonged to INRCA—Italy, 4 to the SAS—Cordoba, Spain, 2 to the HSJD—Barcelona, Spain and 1 to the Parc Taulí, Spain.

Descriptive Statistics

Multimedia Appendix 2 presents descriptive statistics for the most important sociodemographic characteristics such as age, sex, digital expertise, educational level, type of medication, type of surgery, and previous comorbidities. Most of the participants had basic (n=13) or advanced digital skills (n=17) and 8 of them considered themselves experts. Regarding previous comorbidities, 11 of them had cancer in the past, 7 of them had cardiovascular disease, 1 had diabetes mellitus, 1 had renal disease, and 4 of them had mental diseases.

Clinical Outcome Parameters

Regarding the clinical variables, descriptive statistical metrics along with results of univariate analysis between intervention and control groups by *t* test method are presented in Multimedia Appendix 3.

Among all the variables in Multimedia Appendix 3, the VAS values for stress were lower for the intervention groups in all the overall cases (prehospitalization, hospitalization, and discharge) but they did not reach significance. There is only one that demonstrates statistical differences between control and intervention groups: HADS Depression Score during hospital admission in Parc Taulí. For the rest variables, the hypothesis of equal mean between intervention and control groups cannot be rejected due to the high *P* value outcome of the statistical tests. In addition, the mean VAS Stress, VAS Pain, HADS Depression, HADS Anxiety, PANAS Positive, and PANAS Negative score evolution throughout the 4 hospital visits follow-up period are illustrated in Figures 2-7.

Figure 2. Evolution of mean VAS Stress score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). VAS: Visual Analogue Scale.

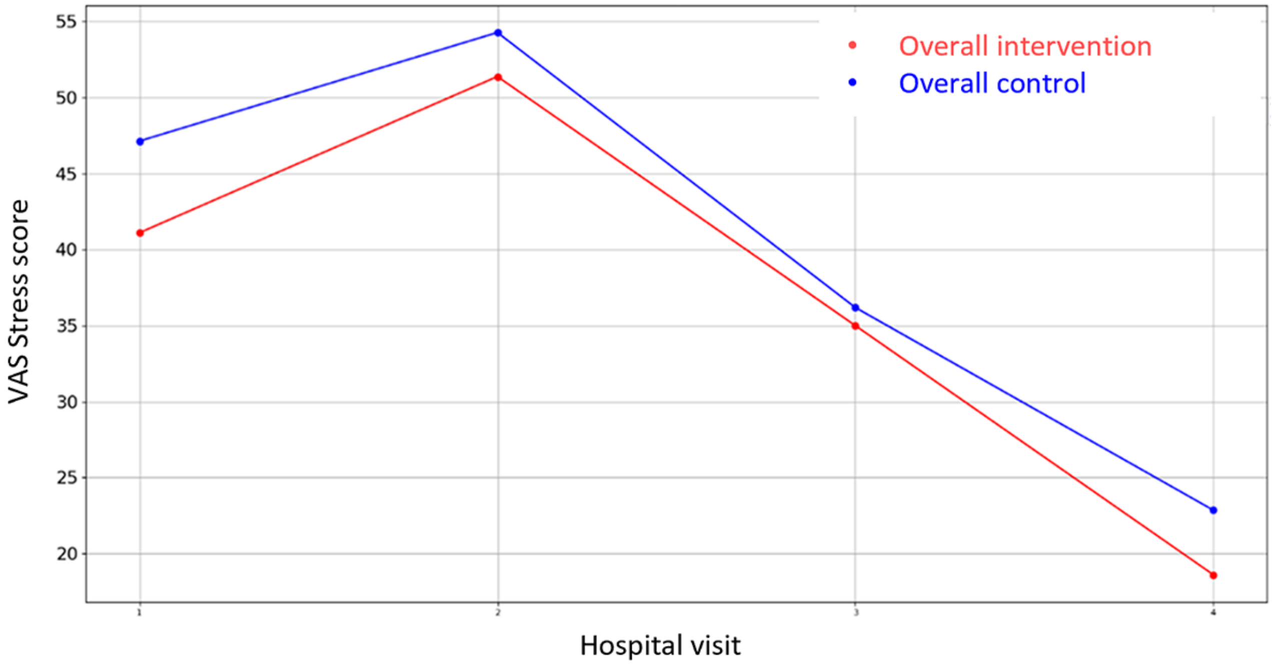


Figure 3. Evolution of mean VAS Pain score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). VAS: Visual Analogue Scale.

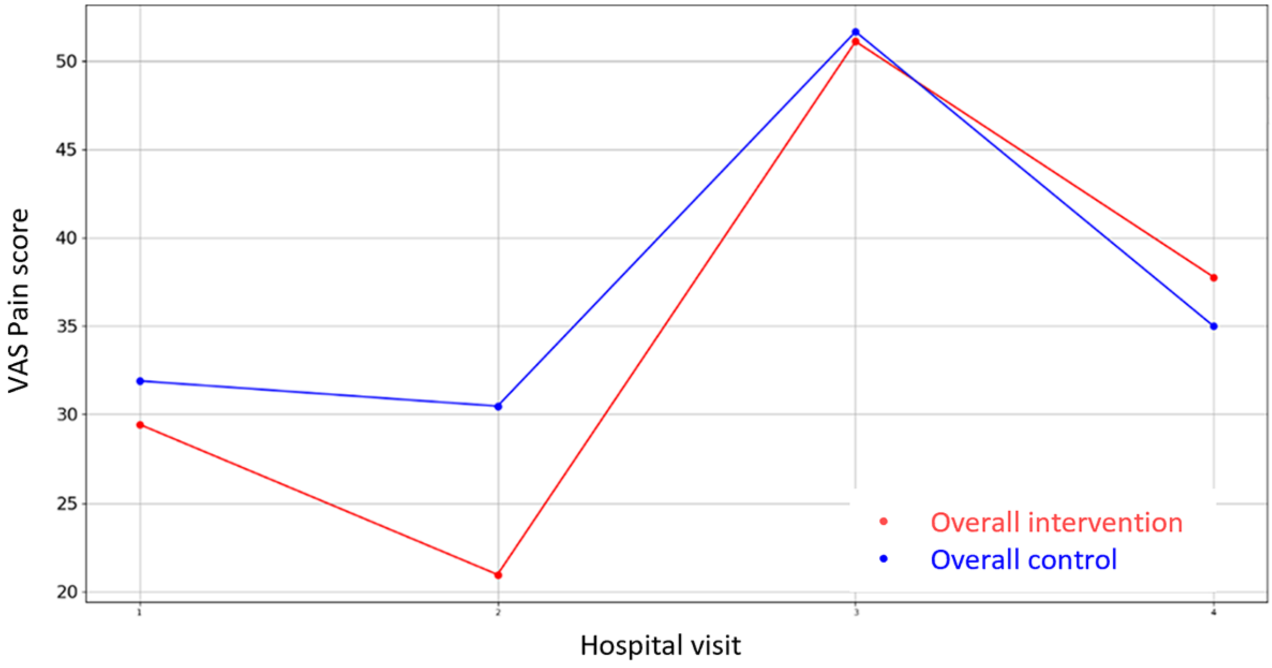


Figure 4. Evolution of mean HADS Depression score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). HADS: Hospital Anxiety and Depression Scale.

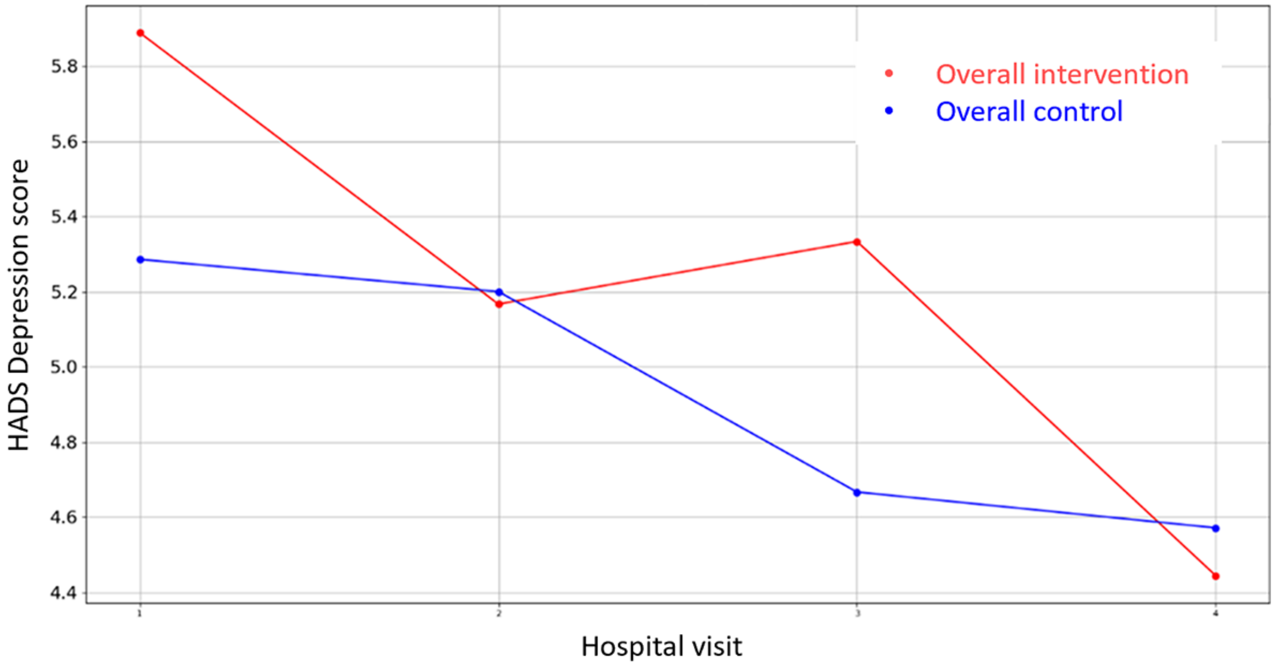


Figure 5. Evolution of mean HADS Anxiety score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). HADS: Hospital Anxiety and Depression Scale.

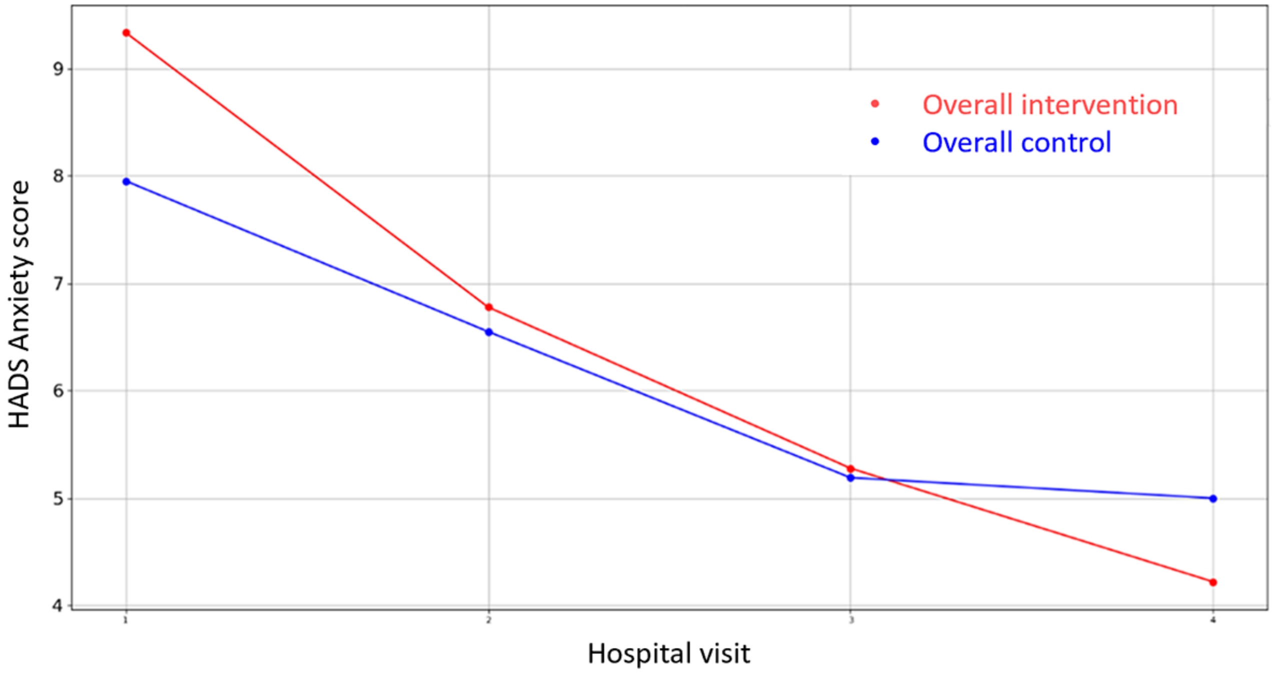


Figure 6. Evolution of mean PANAS Positive score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). PANAS: Positive and Negative Affect Scale.

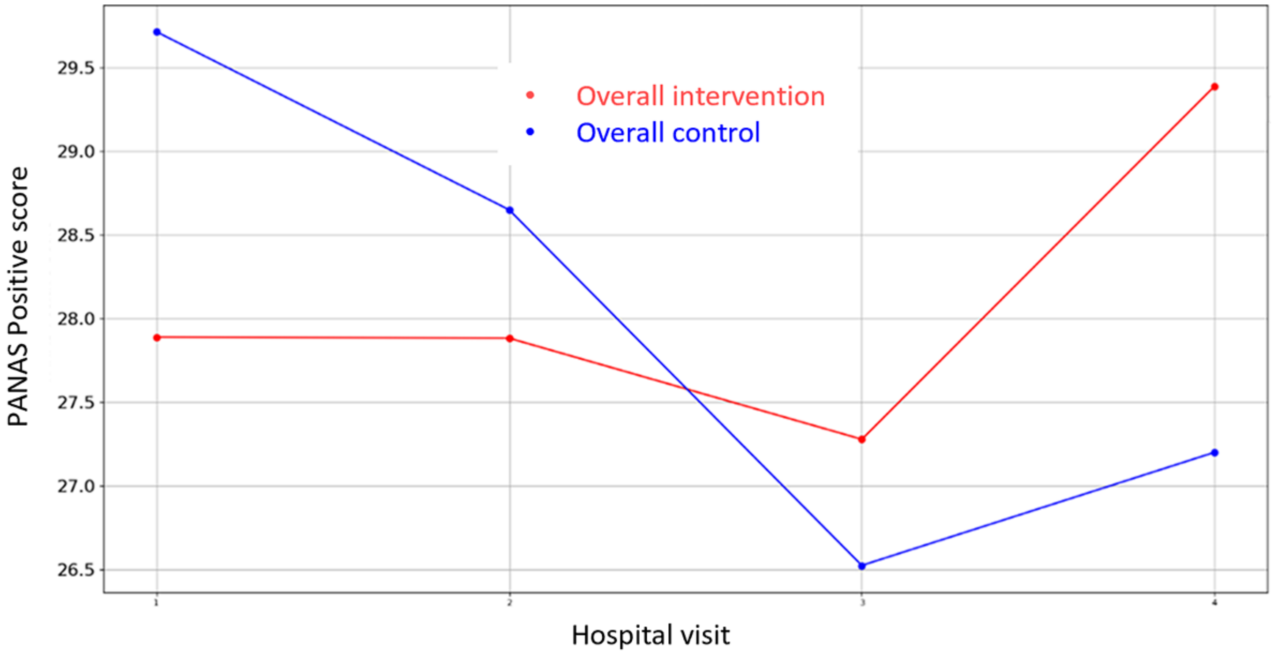
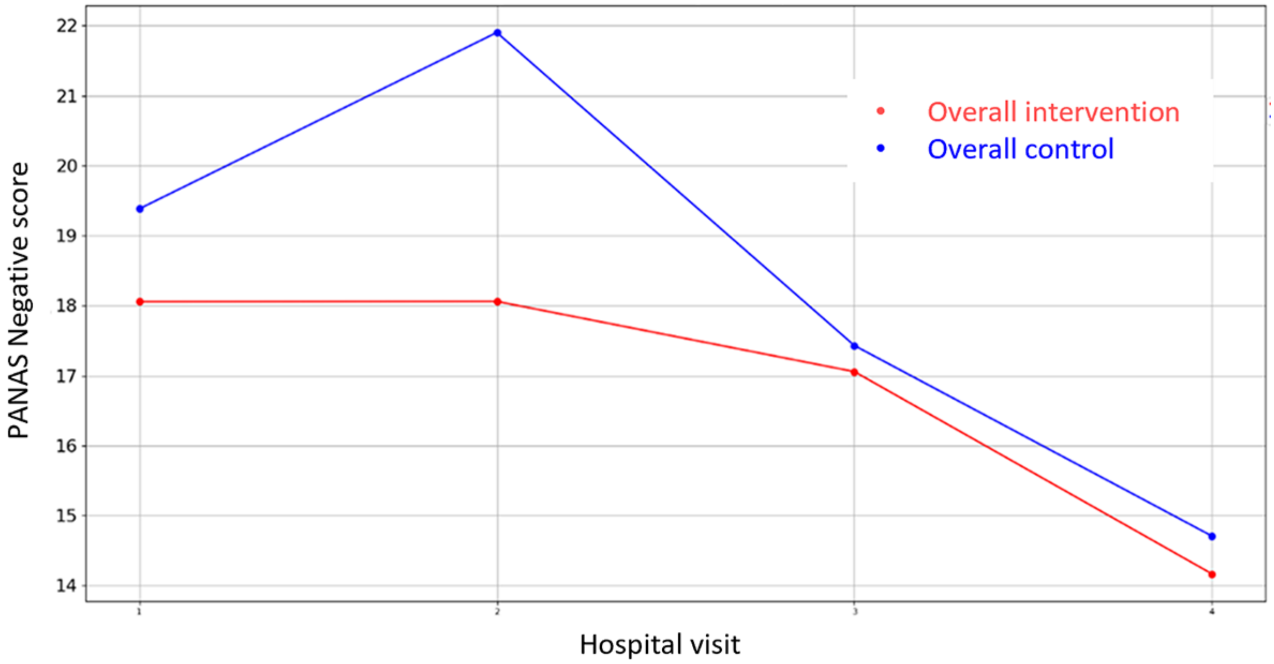


Figure 7. Evolution of mean PANAS (Positive and Negative Affect Scale) Negative score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery).



Likewise, regarding the nonclinical outcome parameters of Multimedia Appendix 4, the same statistical methodology has been designed and executed to identify the most statistically significant variables. In this case, there are three variables where the hypothesis of equal means can be rejected (*P*<.05), that is, PAM-13 Hospital Admission item 13, VAS Stress—hospital discharge, GSE, and the corresponding questionnaires' (PAM-13, GSE, SWEMWBS, QOLmean, VAS Stress Caregiver, SWEMWBS Caregiver, and GSE Caregiver) score evolution graphics are presented in Figures 8-16.

Figure 8. Evolution of mean PAM-13 score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). PAM-13: Patient Activation Measure 13.

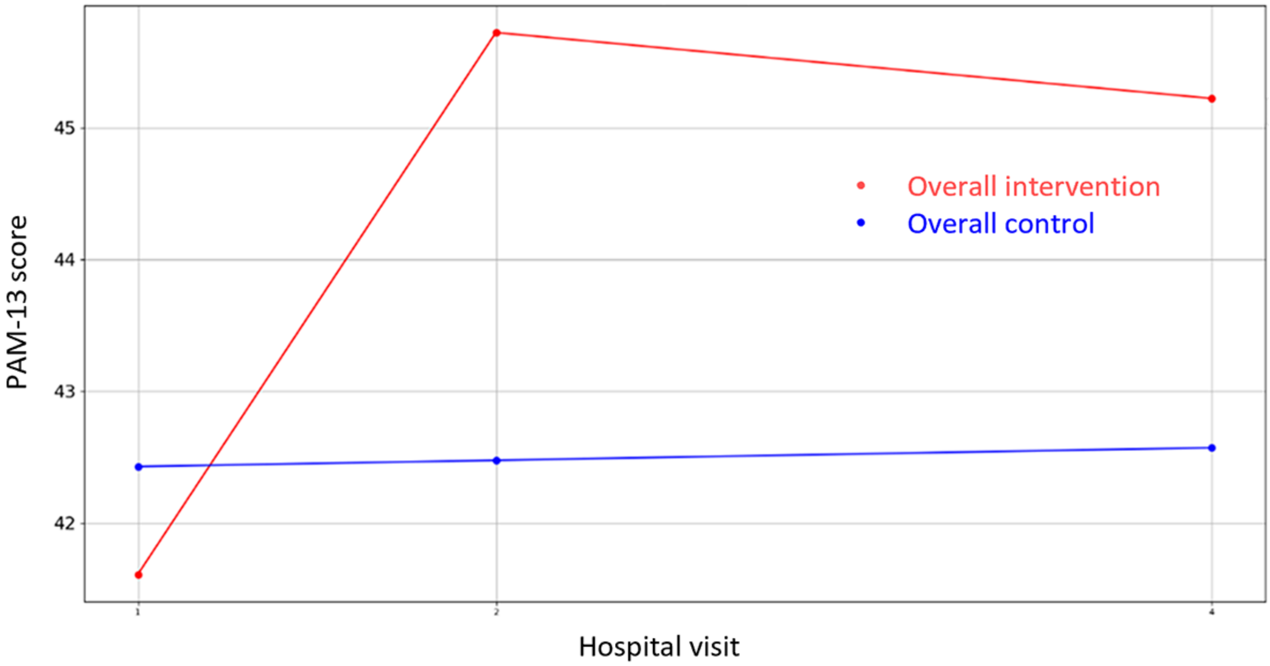


Figure 9. Evolution of mean GSE score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). GSE: General Self-Efficacy.

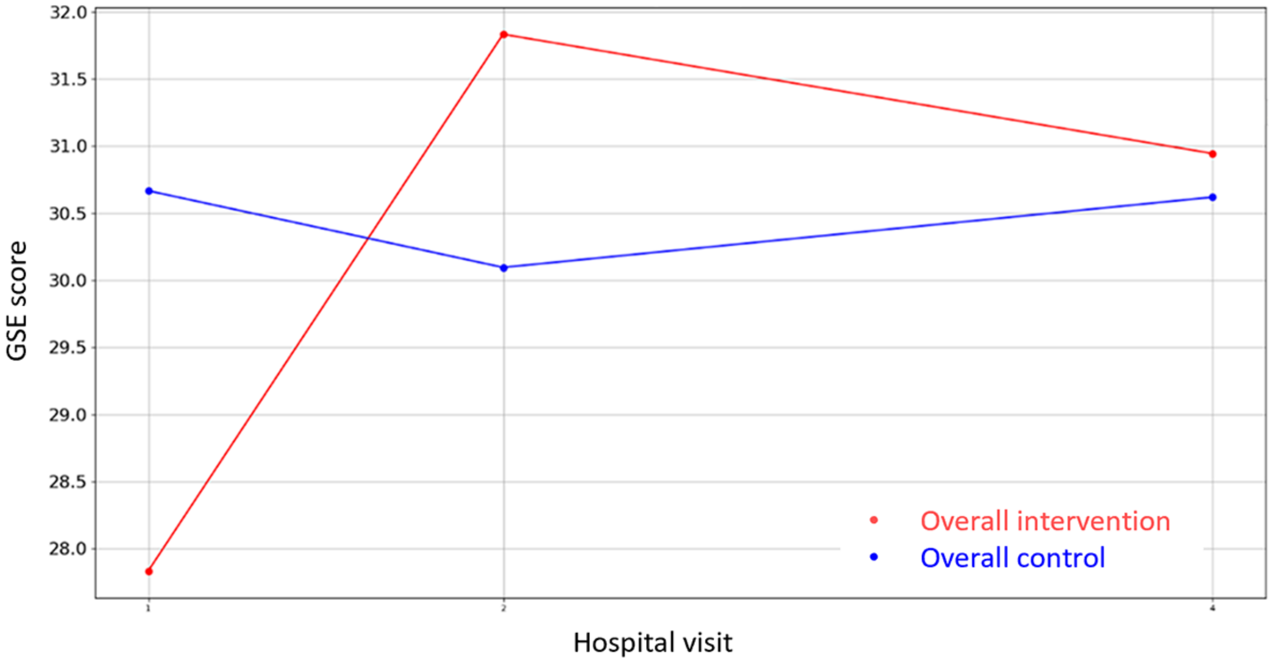


Figure 10. Evolution of mean SWEMWBS score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale.

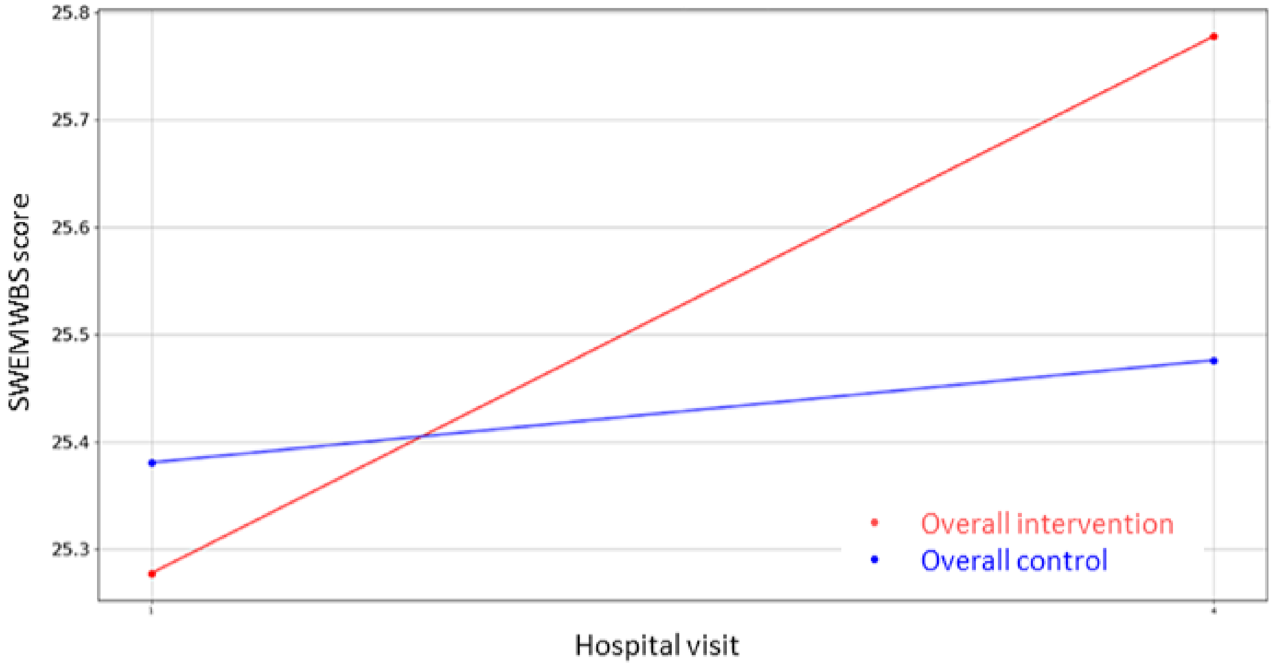


Figure 11. Evolution of mean SWEMWBS Metric score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale.

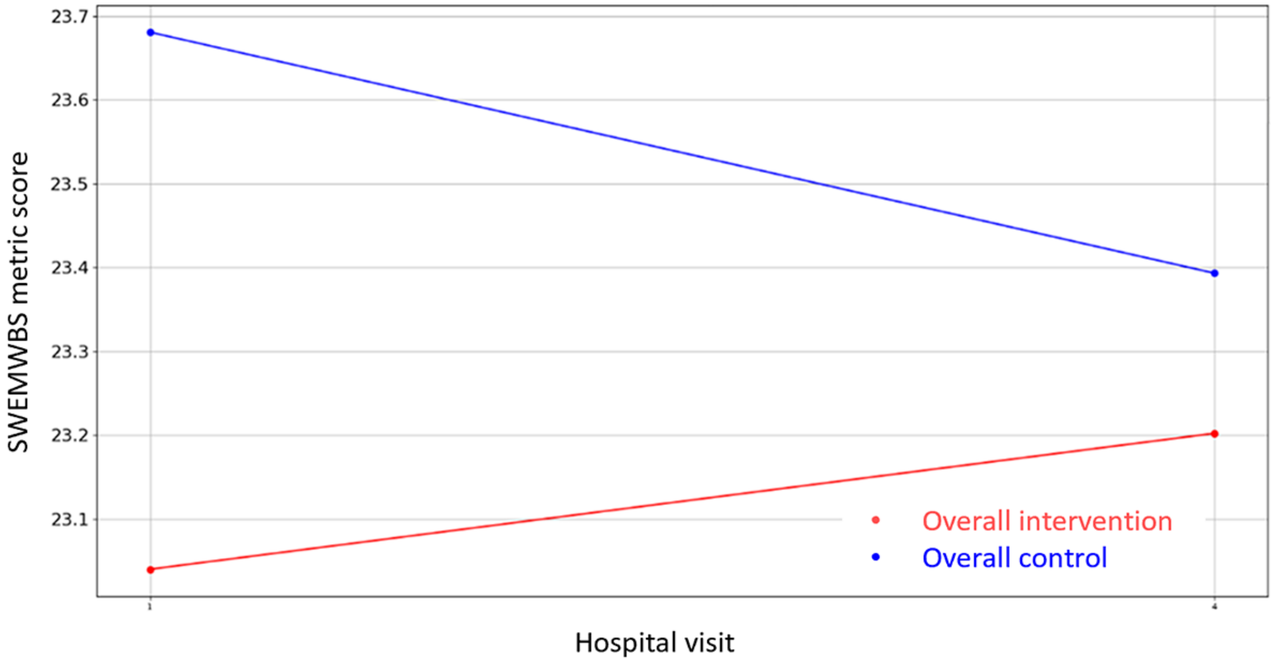


Figure 12. Evolution of mean QOL score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). QOL: Quality of Life.

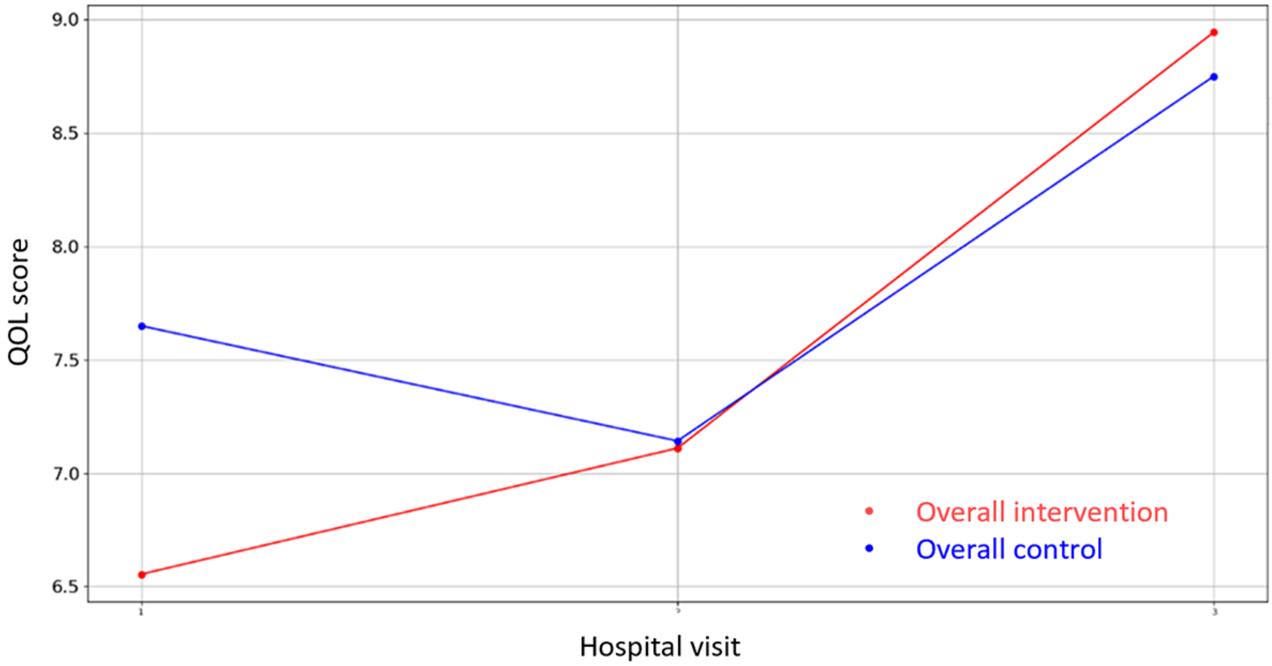


Figure 13. Evolution of mean QOL VAS score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). QOL: quality of life; VAS: Visual Analogue Scale.

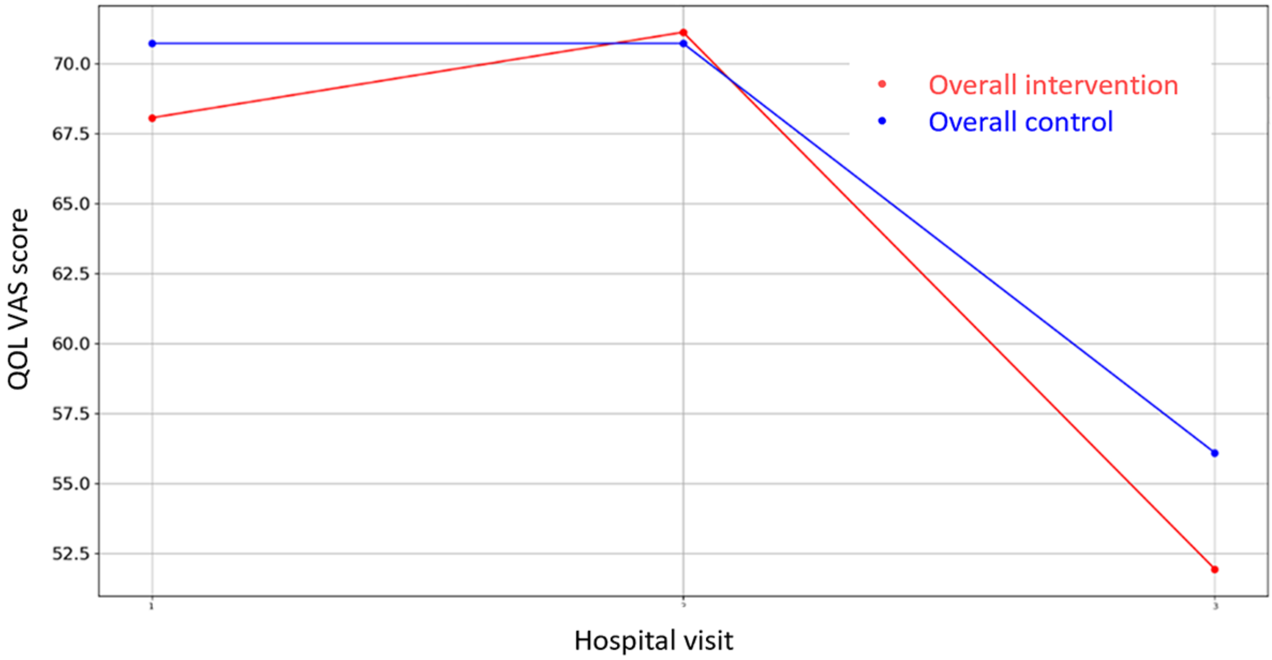


Figure 14. Evolution of Caregiver’s VAS Stress score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). VAS: Visual Analogue Scale.

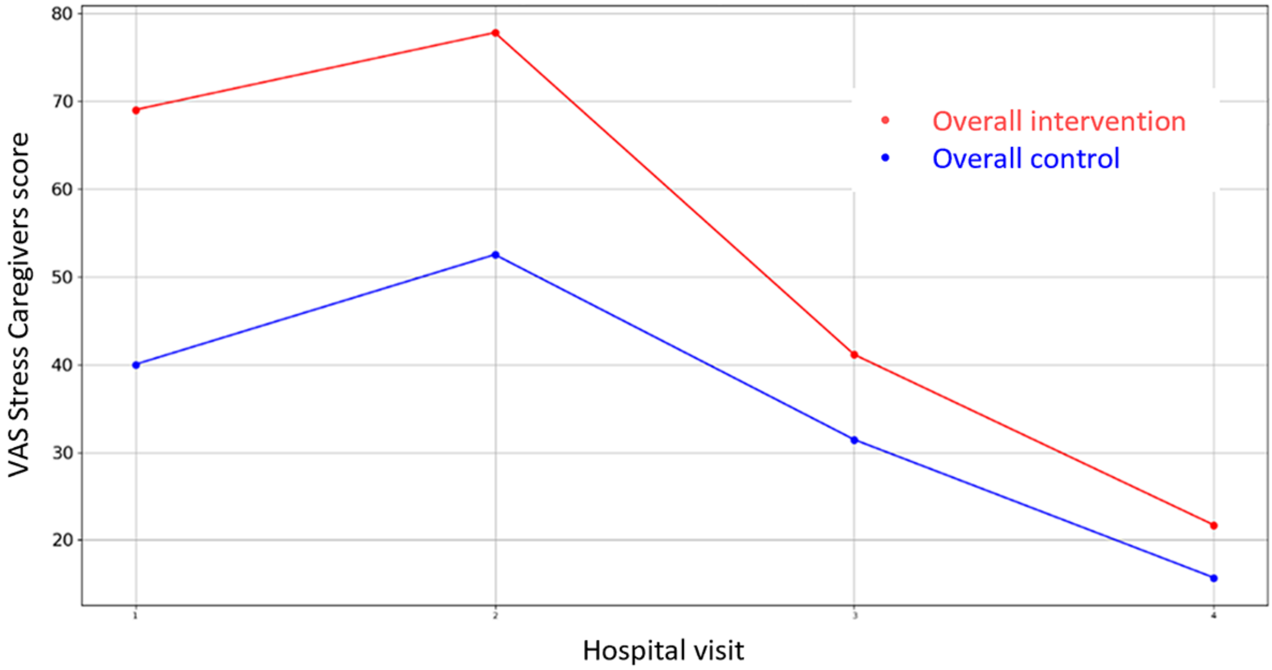


Figure 15. Evolution of Caregiver’s SWEMWBS score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale.

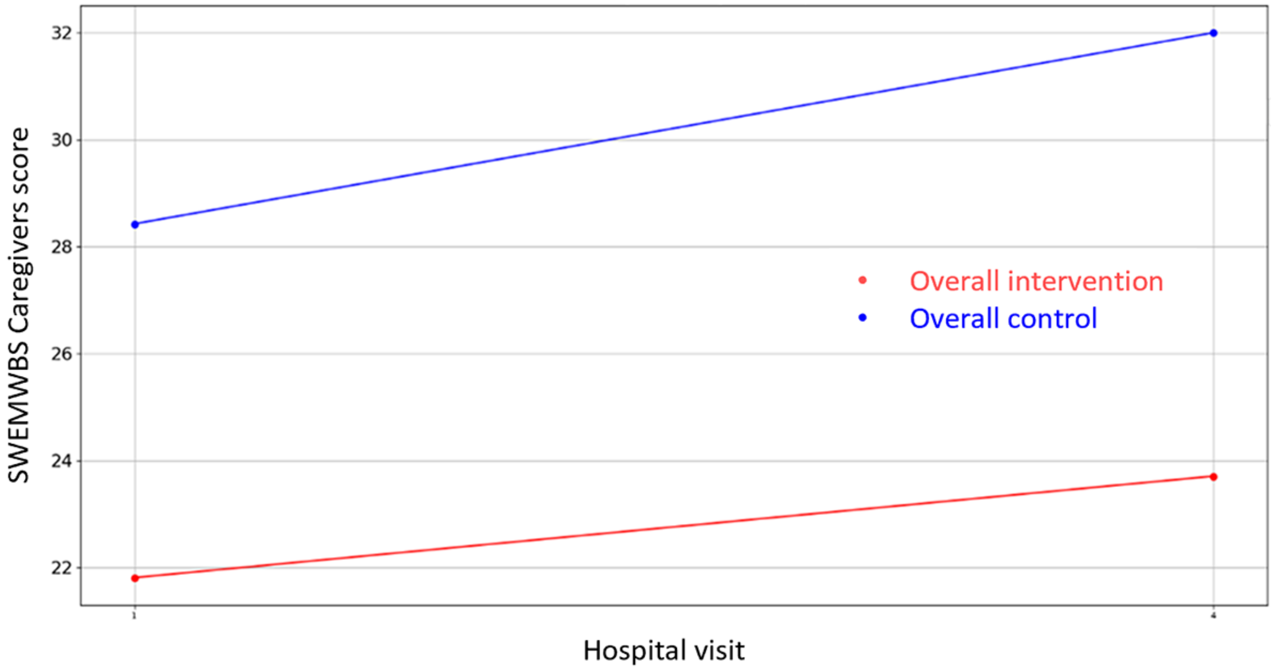
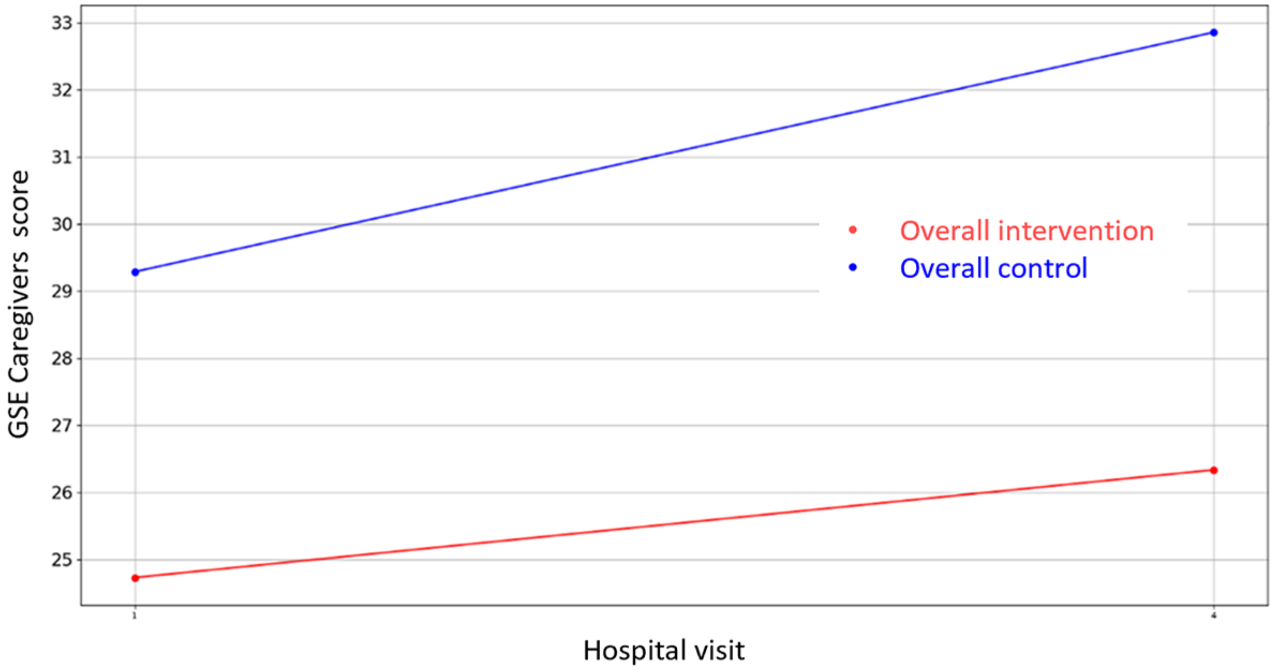


Figure 16. Evolution of Caregiver's GSE score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). GSE: General Self-Efficacy.



Surgery's Exploratory Analysis

ANOVA Analysis

This section aims to investigate any possible statistical effect of the type of surgery on the patients between the intervention and control groups. It uses the mixed-ANOVA statistical tests applied to each unique surgery type group. Thus, depending on the surgery, ANOVA results are as follows:

- Hip replacement surgery: $F_{29,29}$=1.38; P=.19; η_p^2=0.58
- Knee replacement surgery:$F_{29,116}$=259; P=.06; η_p^2=0.27
- Maxilofacial surgery: $F_{29,145}$=0.67; P=.89; η_p^2=0.11
- Cardiac valve replacement surgery: $F_{29,58}$=0.8, P=.73; η_p^2=0.28
- Bladder surgery: $F_{29,116}$=0.38; P=.99; η_p^2=0.08
- Prostate surgery: $F_{29,87}$=0.73; P=.82; η_p^2=0.19
- Scoliosis surgery: $F_{29,116}$=0.13; P=.007; η_p^2=0.32

The *P* value obtained from ANOVA analysis in scoliosis is significant ($P<.05$), and therefore, we conclude that there are significant differences among intervention and control patients undergone that type of surgery.

Tukey Approach

For further investigation of the scoliosis data, an essential pair-wise correlational analysis has been implemented to specify the exact pair of variables whose statistics are significantly different between the intervention and control groups. One of the most commonly used post hoc tests is Tukey's test, which allows us to make pairwise comparisons between the means of each group while controlling for the family wise error rate.

Thus, we would conclude that there is a statistically significant difference ($P<.05$) between the means of groups: VAS Stress (hospital discharge) – VAS Pain (postoperation day) VAS Stress (hospital discharge) – HADS Depression (enrollment) VAS Stress (hospital discharge) – HADS Depression (hospital admission) VAS Stress (hospital discharge) – HADS Depression (hospital discharge) VAS Stress (hospital discharge) – HADS Depression (postoperation day) VAS Stress (hospital discharge) – HADS Anxiety (enrollment) VAS Stress (hospital discharge) – HADS Anxiety (hospital admission) VAS Stress (hospital discharge) – HADS Anxiety (hospital discharge) VAS Pain (enrollment) – VAS Pain (postoperation day) VAS Pain (hospital discharge) – VAS Pain (postoperation day) VAS Pain (hospital discharge) – HADS Depression (enrollment) VAS Pain (hospital discharge) – HADS Depression (hospital admission) VAS Pain (hospital discharge) – HADS Depression (hospital discharge) VAS Pain (hospital discharge) – HADS Depression (postoperation day) VAS Pain (hospital discharge) – HADS Anxiety (enrollment) VAS Pain (hospital discharge) – HADS Anxiety (hospital admission) VAS Pain (hospital discharge) – HADS Anxiety (hospital discharge) HADS Depression (enrollment) – PAM (enrollment).

Linear Mixed Models

Apart from ANOVA tests, to identify any connections between the input data clustered by type of surgery, another statistical approach has been tested, the linear mixed modeling technique.

In this exploratory analysis VAS Stress, VAS Pain, HADS Depression, HADS Anxiety and PANAS questionnaires' scores have been submitted to a linear mixed models (LMM) analysis wherein "time," "group (intervention vs control)," their interaction (time × group), age, and sex have been included as fixed factors. On the other hand, "subjects" and "surgery" have been submitted as random factors. The most significant LMM model's results appear to be the ones with dependent variables the questionnaire VAS Stress and VAS Pain. The corresponding results are presented in Tables 1 and 2.

Table 1. Results of the linear mixed models for the VAS[a] Stress.

	Coefficient	SE	*P* value
Hospital admission score	8.59	5.7	.13
Hospital discharge score	–8.72	5.7	.13
Postoperation visit score	–23.46	5.7	<.001
Sex (female)	11.92	5.06	.02
Groups (intervention/control)	–8.99	4.616	.05
Age	–1.58	0.448	.001

[a] VAS: Visual Analogue Scale.

Table 2. Results of the linear mixed models for the VAS[a] Pain.

	Coefficient	SE	*P* value
Hospital admission score	–4.69	5.195	.36
Hospital discharge score	20.64	5.195	<.001
Postoperation visit score	5.513	5.195	.29
Sex (female)	0.85	4.6	.85
Groups (intervention/control)	–2.22	4.2	.59
Age	–0.28	0.349	.42

[a] VAS: Visual Analogue Scale.

Feature Importance Analysis

Feature importance refers to techniques that calculate a score for all the input features for a given model—the scores simply represent the “importance” of each feature. A higher score means that the specific feature will have a larger effect on the model that is being used to predict a certain variable. The following figures correspond to the computation of the most important factors linked to the principal questionnaires of this study (VAS Stress, VAS Pain, HADS Depression, HADS Anxiety, PANAS) using the XGBoost Regressor. The results are shown in the Multimedia Appendices 5-10.

Adherence and Engagement Analysis to CARINAE App

Adherence to VAS Pain and Stress, as well as to Wound Healing questionnaires into the CARINAE app have been analyzed to show the actions performed by the patients and the results show an overall adherence on VAS Pain of 47.23%, on VAS Stress of 68.41%, and Wound Healing of 2.5%. In the following table (Table 3), the adherence rates are also presented per hospital.

Further, the engagement table to the CARINAE app is presented in Table 4 for the various hospitals showing the total amount of interactions and patient sessions.

Table 3. Mean adherence to VAS[a] Pain, VAS Stress, and Wound Healing questionnaires into the CARINAE app.

Hospital	Days in hospital	Days in study	Completed VAS Pain	VAS Pain rate	Completed VAS Stress	VAS Stress rate	Wound Healing rate
INRCA[b]	3	33	23.33	73%	23	71.13%	3%
Parc Taulí	4	37	12.33	24.15%	15.6	30.50%	0%
Hospital Sant Joan de Déu	2.2	23.4	9.2	40.40%	20.2	79%	4.1%

[a] VAS: Visual Analogue Scale.

[b] INRCA: Instituto di Ricovero e Cura per Anziani.

Table 4. Engagement to CARINAE app in terms of patient and caregiver interactions, number of sessions, and session duration.

Hospital	Patients total interactions	Caregivers total interactions	Patients session duration (in seconds), mean (SD)	Caregivers session duration (in seconds), mean (SD)	Patients total number of sessions	Caregivers total number of sessions
INRCA[a]	526	N/A[b]	4.402 (132)	N/A	375	N/A
Parc Taulí	46	14	1.789 (54)	76 (22)	136	10
Hospital Sant Joan de Déu	264	43	3.456 (104)	2.143 (64)	190	167
SAS	34	N/A	87.683 (2630)	N/A	8	N/A

[a] INRCA: Instituto di Ricovero e Cura per Anziani.

[b] N/A: not applicable.

Discussion

Principal Findings

The main aim of this study was to test the feasibility and efficacy of a personalized stress and anxiety patient empowerment DH solution (CARINAE) on patients undergoing surgery compared to a control group. Furthermore, this study aimed to show the impact of the CARINAE solution on quality of life, emotional status, mental well-being, and self-efficacy, activation status on patients' knowledge, skills, and confidence for self-management. Besides, the CARINAE solution has been assessed by health care professionals and patients and caregivers allocated to the intervention group, in terms of engagement.

According to our results, at baseline, patients of both groups reported generally normal levels of pain and stress, with higher levels of stress. Patients undergoing coronary bypass and cardiac valve replacement surgeries indicated higher levels of stress before the surgery (VAS>80), followed by hip and knee replacement surgeries in both pain and stress (VAS>60) compared to the other patients' surgeries. Similar stress values can be observed by caregivers. Regarding depression, patients of both groups reported general normal levels but they indicated to experience anxiety (HADS score between 7 and 10). In detail, depression is primarily observed by older patients undergoing prostate and bladder cancer removal surgery, with values between the abnormal range (HADS depression abnormal range 11-21) and secondarily, with borderline values (HADS depression borderline range 8-10), by patients undergoing coronary bypass and cardiac valve replacement surgeries. In contrast, anxiety is more generalized among various patients, and the results showed that older patients undergoing an operation to remove prostate or bladder cancer report very high values, in the abnormal range (HADS anxiety >11), followed by the youngest undergoing scoliosis and maxillofacial surgeries (HADS anxiety=10), and hip and knee surgeries, as well as coronary bypass and cardiac valve replacement surgeries with borderline levels of anxiety (HADS anxiety borderline range 8-10). Regarding positive and negative affect scores, patients of both groups showed normal levels of positive affect (mean scores normal levels 33.3, SD 7.2) and higher values regarding negative affect (mean scores normal levels 17.4, SD 6.2). Regarding patients' knowledge, skills, and confidence for self-management, no patient of either group found it important to take action with respect to the situation and the self-efficacy perception resulted quite high in most patients and caregivers (GSE>25). Finally, with respect to mental well-being and quality of life, at baseline patients and caregivers of both groups present quite high rates of mental well-being and quality of life. With respect to the latter two measures, it can be emphasized that patients who are older and have undergone more delicate surgeries, such as coronary bypass and cardiac valve replacement surgeries showed lower levels than other surgeries and age. A potential bias is the difference in age between the control and intervention groups, the control group has on average 10 years more. In a small sample, this bias might have occurred inevitably, but it might be indicative of a selection bias. In future studies, we should consider the possibility of including a "control intervention" that also requires the use of some simple mobile-based technology (eg, watching educational videos on YouTube).

Looking at the relationship between the variables, overall patients show most strong associations between negative emotions, anxiety, and stress, especially in the preoperative phase (recruitment and hospital admission), which in some cases are also maintained in the postoperative phase (hospital discharge and postoperative follow-up visit), but nevertheless, the presence of a strong inverse relationship between mental well-being and anxiety at the postoperative follow-up visit suggests that as anxiety decreases, patients' mental well-being increases again.

Examining the 2 groups separately, the control group shows greater relationships between the more negative psychological dimensions, such as anxiety, stress, negative emotions, and depression, whereas the intervention group shows more positive relationships between the psychological dimensions of self-efficacy, self-management, and mental well-being, suggesting that CARINAE solution could have a positive effect and impact on the reduction of stress and negative emotions. In detail, the control group showed levels of stress, depression, anxiety, and negative emotions that remained constant throughout the entire perioperative process, and only at the end of the process could changes in the perception of greater mental and physical well-being be observed. There were positive trends in the intervention group, although not significant from a statistical point of view. The intervention group shows a greater perception of mental well-being, self-efficacy, and self-management during the entire perioperative process and not only at the end of the process. Considering the clinical outcome parameters on pain, stress, anxiety, and depression, a significant difference between the two groups has been found in the depression subscale regarding the hip and knee surgeries, still showing values within the normal range. Although no other statistically significant differences were found, it can be seen how the values with respect to pain remain low at the time of hospitalization and how they increase after the operation and then decrease again at the follow-up visit. Pain is known to be related to stress levels. The highest postoperative pain values are perceived more in those patients who underwent a more invasive operation (eg, coronary bypass and cardiac valve replacement surgeries) than in those whose operation was less invasive (prostate and bladder cancer). As expected, as far as stress was concerned, no significant differences were found, but the values increased at hospital admission, compared to baseline, and these values decreased at discharge and more so at the follow-up visit. Looking at the individual operations, the cardiac valve replacement surgeries and hip and knee replacement surgeries on admission to the hospital are the ones that cause the most stress for patients. Finally, the values of anxiety and depression remained within the normal ranges, and with respect to the baseline, it can be observed that in patients undergoing prostate or bladder cancer removal surgery who had high levels, the levels of both parameters steadily decreased at various stages of the perioperative process. We should notice that the project and the intervention aim at reducing stress levels, and not clinical anxiety disorder or major depression. This is quite important since a narrower focus on major mental health disorders might reduce the applicability of the intervention to most of the population who undergoes surgery. We should be aware that the exploratory analyses might, due to multiple testing, be subject to type 1 error, and cautious interpretation of the analysis is warranted.

Regarding the nonclinical outcome parameters, no significant differences have been found. However, the positive effect values remain stable and in the normal range during the perioperative process, increasing more after the hospital discharge, while negative values decrease after the operation and at the follow-up visit. Regarding patients' knowledge, skills, and confidence for self-management, no patient of either group found it important to take action with respect to the situation, and the self-efficacy perception resulted higher in patients and caregivers after the surgeries than the baseline (GSE>30). Finally, with respect to mental well-being and quality of life, in the follow-up visit, patients and caregivers of both groups presented higher rates of mental well-being and lower rates of quality of life than baseline, suggesting a faster recovery process of mental well-being in terms of recovery of feeling relaxed, optimistic, thinking clearly, and dealing with problems well and a slower recovery of quality of life in terms of recovery of mobility, autonomy, and usual activities. These results are not conclusive, and a bigger study might be needed to identify the level of effectiveness of the intervention. However, major aspects deserve further attention prior to a larger study. The low awareness about the importance of self-management might be an indicator of low health literacy levels. It might be wise to consider strategies that as part of the intervention address awareness of lifestyle factors for recovery even before the intervention itself.

The founded differences between the two groups may be linked to the young age of the participants and consequently to the related hormonal and neurodevelopment changes. Indeed, these changes are currently conceptualized in terms of imbalance between systems supporting reactivity and regulation, specifically nonlinear changes in reactivity networks and linear changes in regulatory networks [48].

In addition, as identified by LMMs, stress and pain showed other statistical differences in accordance with other parameters, showing that experiences of stress seem to differ according to age and sex. The level of stress rises with aging and is higher in female patients than in male patients. In addition, both stress and pain also differed depending on the phase of the perioperative process in which patients find themselves. In detail, pain presented more effect and impact at hospital discharge, showing the highest levels of the perioperative process, and stress at the end of the entire process (at follow-up visit), showing the lowest levels of the perioperative process. Further analyzing the importance of the various features using XGBoost revealed that stress, anxiety, pain, and negative affect measures are interrelated as expected and that physical activity is also important.

Finally, according to the adherence to the psychometric questionnaires in the CARINAE app, we observed that younger and older participants were more involved in the self-evaluation regarding stress with 79% and 71%, respectively. On pain, older participants were involved more than younger with 73% of responses by the older and 40% by the younger. These data are also reflected in the use of the CARINAE app, where the greatest interactions were found among older and younger patients. These data suggest, on the one hand, a greater provision of time by older people, and on the other hand, extensive use of technology by younger people and overall high involvement in the use of CARINAE for both ages.

Comparison to Literature

Although no significant differences have been found between groups, advances in DH interventions are playing great support in enhancing awareness about health conditions and for the management of mental health by relying on both nonimmersive (eg, such as mobile apps) and immersive systems (eg, VR) [16]. These include support for patients in the management of anxiety, stress [37], and pain [18]. On one hand, mobile apps for perioperative processes are becoming a hot topic, allowing to provide psychoeducational contents, mental well-being activities to reduce pain and stress, up-to-date information, tracking personal health data, reminding and engaging patients, and communicating in a cost-effective way. On the other, VR has been used in multiple health care applications including reducing stress and pain, training medical practitioners, patient counseling, cognitive rehabilitation, physical therapy in medicine, and for diagnostic and treatment needs in dentistry, mental health management, and surgery [19]. A multiuser immersive VR system was developed and used during presurgical discussions in a prospective patient cohort undergoing cerebrovascular surgery [20]. An immersive VR intervention adopted in pediatric patients to manage pain and anxiety provided a new, easy, and cost-effective intervention that can be applied to other painful and stressful medical procedures [49]. Pain is a highly distressing symptom for patients in all clinical settings and stress and anxiety levels influence it. VR applications have proven to be efficient in stress relief and pain management, mainly due to their distractive properties [22].

In the literature, there is a lack of understanding of the feasibility of combining VR with mobile-based technologies as the main channel of the provision of a DH intervention [36]. In this study, what is explored is the feasibility of a DH intervention that leverages the latest mobile and VR technologies within the use case of helping patients manage stress and anxiety during surgery while promoting healthy recovery.

The findings of this project, reveal the importance of addressing this type of intervention as a service design approach with a strong focus on implementation aspects. A recent guideline by the World Health Organization in DH reinforces the importance of addressing training and supporting the environment as key success aspects in DH implementations [50]. In addition, infrastructure was a key element (eg, connectivity issues). The workforce is not only crucial from a service delivery point of view, but also a key channel to facilitate patient DH literacy which in this very complex setup with multiple devices and features is especially important.

Strengths and Limitations

Several strategies and techniques have been proposed to manage preoperative stress and anxiety that can be effective in supporting patients to cope with a wide range of stressful health situations. In the current health care settings, however, it is not very common to provide patients with stress and anxiety relief support prior to a surgical procedure. Usually, VR-enhanced solutions focus only on providing informative content, neglecting the importance of patient empowerment with a more robust educational curriculum. This study has been among the first to evaluate the potential effectiveness of a comprehensive technology in reducing perioperative stress and anxiety. CARINAE provided a unique combination of endpoints and the integration of knowledge deriving from several domains, including stress or anxiety management, patient empowerment, communication with medical professionals, adaptation to illness, self-regulation and self-management, and adaptation to medical procedures. The study has been successful in terms of identifying the key impacts of such type of intervention and provided enough exploratory insights to redesign the intervention and establish a new randomized controlled trial design to provide more conclusive data on efficacy. At the same time, this study allowed us to identify implementation issues that need to be addressed prior to larger studies and implementations. For example, larger trials with more participants are advisable for future work, which will facilitate the application of ANCOVA analysis.

This solution allowed participants to receive constant feedback to improve their appraisal and coping skills in an entertaining and motivating manner. It focuses on patient empowerment through active participation in the process and is dynamically adapted according to operation type, patient preferences, needs, and medical history all the way through the preclinical phase, admission, and discharge, in a continuous and personalized way. At the same time, it facilitated effective interactions between patients and health care professionals, through user-friendly and intelligent communication. It used the spaced learning methodology to help patients understand and learn the diverse aspects of their surgical process, from presurgery requirements to recovery steps with stress management all along the process, and provides multichannel anxiety and stress relief personalized content.

CARINAE solution, combining mobile health and VR technologies with a web app, provided positive preliminary results in reducing perioperative stress and creating effective collaborations between physicians or surgeons and their patients whilst supporting them in improving their knowledge in related domains. CARINAE has shown the potential to improve physical and emotional reactions to a stressor, such as surgical operations, to increase the levels of calmness to promote a sense of well-being and to empower patients in preoperative conditions. Information provided through the platform advances and enhances health literacy and digital competence and increases the participation of the patient in the decision-making process. Integration with third-party applications has facilitated the exchange of important information between patients and physicians, as well as between personal applications and clinical health systems.

While the findings of this study are interesting and valuable, it has some limitations. First, the small sample size, as well as the high variability of the ages and surgeries might limit the generalizability of the results. Given the pragmatic nature of this feasibility study, its small sample size was anticipated, further compounded by emergent challenges amid the COVID-19 crisis. The study did not aim to have a conclusive answer in terms of effectiveness, but to gather insight on whether it can be done as suggested in related literature [51]. As such, this study provides insights to support the design of a future randomized controlled trial. While a prospective observational design was considered, its susceptibility to confounding due to temporal changes prompted the inclusion of a control group, albeit with a heightened risk of limited significance. Moving forward, future research should prioritize more robust, adequately powered clinical studies to offer comprehensive insights.

Furthermore, pain and anxiety management guidelines might differ from one clinical setting to another, thus providing limited evidence based on raw cross-comparisons between different health care providers. Finally, the different durations of each surgery and each process limited the engagement data to the DH solution. Third, no data about use have been gathered from the VR app, limiting the quantitative performance rates and engagement to the VR component.

Recommendations for Future Research, Development, and Clinical Practice

An area of future improvement that both this feasibility study and other DH solutions demonstrate the need for is long-term patient engagement. Indeed, according to the literature, as mobile health apps grow more prevalent, many have low attrition rates, reducing the relevance of data collection for the study [51]. Potential hurdles to long-term involvement include the patients' functional condition and the workload associated with performing in-app tasks. Our findings are aligned with those studies. This is especially true for procedures such as cancer and cardiac surgeries, where the target population is older and potentially more functionally constrained than younger patient populations. To address the issue of long-term engagement, potential solutions include developing even more simplified versions of the solution that are easier to use for those unfamiliar with technology, using SMS text messaging–only modes of communication, and using shorter surveys that take less time to complete. For younger patients, on the contrary, more videos and images instead of text could engage more. Additionally, it may empower patients further if they are shown the outcomes of their data and are able to visualize and quantify in real-time, how their condition has improved. Previous studies have noted that patients provided positive feedback when presented with their study results and often noted reluctance to participate because they had not received personalized feedback from their care teams in previous app-based studies [51]. In this study, differences across sites might be due to differences in populations but also on how the solutions were promoted or introduced in each site. When integrating new technology into any health care setting, provider input may significantly streamline the integration of applications into already complex clinical workflows. Continually enhancing the experience and usability, from both the patient and provider perspectives, will allow the realization of the full potential of DH solutions for patients undergoing interventional procedures.

Conclusions

We evaluated the feasibility of using CARINAE, a comprehensive DH solution, for patients undergoing surgery. This study examined the user experience and engagement with the solution, as well as the ability to collect patient-reported outcomes. Our results provide an important first step toward a deeper understanding of optimizing DH solutions to support patients undergoing surgery and for potential applications in remote patient monitoring and communication. This study highlighted how DH may be integrated into major procedures to improve patient education, emotional management, and engagement, and serve as a versatile clinical and research tool. Future studies will aim to test CARINAE in a larger cohort to establish its potential impact on health care resource utilization.

Acknowledgments

This study was part of the STARS-EU-PCP project that has received funding from the European Union's Horizon 2020 Research and Innovation Program (grant: 727585).

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

IACG, HA, SH-F, PLG-S, FJN-B, and LF-L developed the patient view of the digital solution and coordinated clinical studies across all hospitals. In addition, they performed the initial data analysis. HK, DK, and AK developed the clinician view of the digital solution and contributed to the analysis of the results and their interpretation. RMB-R focused on the analysis of the results from the psychological perspective. GP, PZ, KA, and CS focused on the development of the virtual reality technology of the solution and contributed to the analysis of the results. All authors contributed to the writing of the manuscript, and reviewed, edited, and approved the final manuscript for submission.

Conflicts of Interest

IACG, HA, SH-F, and PLG-S are employees of Adhera Therapeutics, which owns the Adhera CARINAE DH Program. LF-L and SH-F were co-founders of Salumedia Labs, which was acquired by Adhera Therapeutics. All other authors declare no conflicts of interest.

Multimedia Appendix 1
The completed CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 1291 KB - jmir_v27i1e54049_app1.pdf]

Multimedia Appendix 2
Baseline characteristics of the control and intervention arms, both overall and per hospital.
[DOCX File , 43 KB - jmir_v27i1e54049_app2.docx]

Multimedia Appendix 3
Clinical outcome parameters of the control and intervention arms, both overall and per hospital.
[DOCX File , 42 KB - jmir_v27i1e54049_app3.docx]

Multimedia Appendix 4

Non-clinical outcome parameters of the control and intervention arms, both overall and per hospital.

[DOCX File , 63 KB - jmir_v27i1e54049_app4.docx]

Multimedia Appendix 5

HADS (Hospital Anxiety and Depression) Depression Feature Importance Analysis by XGBoost.

[PNG File , 21 KB - jmir_v27i1e54049_app5.png]

Multimedia Appendix 6

Panas Positive Feature Importance Analysis by XGBoost.

[PNG File , 19 KB - jmir_v27i1e54049_app6.png]

Multimedia Appendix 7

PANAS Negative Feature Importance Analysis by XGBoost.

[PNG File , 50 KB - jmir_v27i1e54049_app7.png]

Multimedia Appendix 8

VAS Pain Feature Importance Analysis by XGBoost.

[PNG File , 21 KB - jmir_v27i1e54049_app8.png]

Multimedia Appendix 9

HADS Anxiety Feature Importance Analysis by XGBoost.

[PNG File , 20 KB - jmir_v27i1e54049_app9.png]

Multimedia Appendix 10

VAS Stress Feature Importance Analysis by XGBoost.

[PNG File , 53 KB - jmir_v27i1e54049_app10.png]

References

1. Kumar A, Das S, Chauhan S, Kiran U, Satapathy S. Perioperative anxiety and stress in children undergoing congenital cardiac surgery and their parents: effect of brief intervention—a randomized control trial. J Cardiothorac Vasc Anesth 2019 May;33(5):1244-1250. [doi: 10.1053/j.jvca.2018.08.187] [Medline: 30243867]
2. Mathew JP, Machovec KA. A new paradigm of anxiolysis. J Cardiothorac Vasc Anesth 2019 May;33(5):1251-1252. [doi: 10.1053/j.jvca.2018.09.022] [Medline: 30529182]
3. Anton NE, Montero PN, Howley LD, Brown C, Stefanidis D. What stress coping strategies are surgeons relying upon during surgery? Am J Surg 2015;210(5):846-451. [doi: 10.1016/j.amjsurg.2015.04.002] [Medline: 26105802]
4. Sherman DW. A review of the complex role of family caregivers as health team members and second-order patients. Healthcare (Basel) 2019;7(2):63 [FREE Full text] [doi: 10.3390/healthcare7020063] [Medline: 31022837]
5. Kondylakis H, Bucur A, Dong F, Renzi C, Manfrinati A, Graf N. iManageCancer: developing a platform for empowering patients and strengthening self-management in cancer diseases. IEEE; 2017 Presented at: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS); June 22-24, 2017; Thessaloniki, Greece p. 755-760. [doi: 10.1109/CBMS.2017.62]
6. Cheung KL, Durusu D, Sui X, de Vries H. How recommender systems could support and enhance computer-tailored digital health programs: a scoping review. Digital Health 2019;5:2055207618824727 [FREE Full text] [doi: 10.1177/2055207618824727] [Medline: 30800414]
7. Nigussie S, Belachew T, Wolancho W. Predictors of preoperative anxiety among surgical patients in Jimma University specialized teaching hospital, South Western Ethiopia. BMC Surg 2014;14:67 [FREE Full text] [doi: 10.1186/1471-2482-14-67] [Medline: 25189274]
8. Thompson M, Moe K, Lewis CP. The effects of music on diminishing anxiety among preoperative patients. J Radiol Nurs 2014;33(4):199-202. [doi: 10.1016/j.jradnu.2014.10.005]
9. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
10. Katehakis DG, Kondylakis H, Koumakis L, Kouroubali A, Marias K. Integrated care solutions for the citizen: personal health record functional models to support interoperability. EJBI 2017;13(1):51-58. [doi: 10.24105/ejbi.2017.13.1.8]
11. Chi N, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2015;21(1):37-44 [FREE Full text] [doi: 10.1177/1357633X14562734] [Medline: 25475220]

12. Pilapil M, Coletti DJ, Rabey C, DeLaet D. Caring for the caregiver: supporting families of youth with special health care needs. Curr Probl Pediatr Adolesc Health Care 2017;47(8):190-199. [doi: 10.1016/j.cppeds.2017.07.003] [Medline: 28803827]
13. Kiecolt-Glaser J, Page G, Marucha P, MacCallum R, Glaser R. Psychological influences on surgical recovery: perspectives from psychoneuroimmunology. Am Psychol 1998 Nov;53(11):1209-1218. [doi: 10.1037//0003-066x.53.11.1209] [Medline: 9830373]
14. Gürsoy A, Candaş B, Güner Ş, Yılmaz S. Preoperative stress: an operating room nurse intervention assessment. J Perianesth Nurs 2016;31(6):495-503. [doi: 10.1016/j.jopan.2015.08.011] [Medline: 27931701]
15. Ibrahim MS, Khan MA, Nizam I, Haddad FS. Peri-operative interventions producing better functional outcomes and enhanced recovery following total hip and knee arthroplasty: an evidence-based review. BMC Med 2013;11:37 [FREE Full text] [doi: 10.1186/1741-7015-11-37] [Medline: 23406499]
16. East ML, Havard BC. Mental health mobile apps: from infusion to diffusion in the mental health social system. JMIR Ment Health 2015;2(1):e10 [FREE Full text] [doi: 10.2196/mental.3954] [Medline: 26543907]
17. Kondylakis H, Kouroubali A, Koumakis L, Rivero-Rodriguez A, Hors-Fraile S, Katehakis D. Designing a novel technical infrastructure for chronic pain self-management. Stud Health Technol Inform 2018;249:203-207. [doi: 10.3233/978-1-61499-868-6-203] [Medline: 29866983]
18. Kondylakis H, Hors-Fraile S, Koumakis L, Kouroubali A, Notas G, Rivero-Rodriguez A, et al. An innovative, information and communication technology supported approach, towards effective chronic pain management. In: Alternative Pain Management: Solutions for Avoiding Prescription Drug Overuse. Hershey, Pennsylvania, USA: IGI Global; 2020:21.
19. Pillai AS, Mathew PS. Impact of virtual reality in healthcare: a review. Virtual Augment Real Ment Health Treat 2019:17-31. [doi: 10.4018/978-1-5225-7168-1.ch002]
20. Sugiyama T, Clapp T, Nelson J, Eitel C, Motegi H, Nakayama N, et al. Immersive 3-dimensional virtual reality modeling for case-specific presurgical discussions in cerebrovascular neurosurgery. Oper Neurosurg (Hagerstown) 2021;20(3):289-299 [FREE Full text] [doi: 10.1093/ons/opaa335] [Medline: 33294936]
21. Wong CL, Lui MMW, Choi KC. Effects of immersive virtual reality intervention on pain and anxiety among pediatric patients undergoing venipuncture: a study protocol for a randomized controlled trial. Trials 2019;20(1):369 [FREE Full text] [doi: 10.1186/s13063-019-3443-z] [Medline: 31221208]
22. Scates D, Dickinson JI, Sullivan K, Cline H, Balaraman R. Using nature-inspired virtual reality as a distraction to reduce stress and pain among cancer patients. Environ Behav 2020;52(8):895-918. [doi: 10.1177/0013916520916259]
23. Prabhu VG, Stanley L, Morgan R. A biofeedback enhanced adaptive virtual reality environment for managing surgical pain and anxiety. Int J Semant Comput 2020;14(3):375-393. [doi: 10.1142/s1793351x20400152]
24. Prabhu VG, Stanley L, Morgan R, Shirley B. Designing and developing a nature-based virtual reality with heart rate variability biofeedback for surgical anxiety and pain management: evidence from total knee arthroplasty patients. Aging Ment Health 2024;28(5):738-753. [doi: 10.1080/13607863.2023.2270442] [Medline: 37850735]
25. Ryu JH, Park JW, Nahm FS, Jeon YT, Oh AY, Lee HJ, et al. The effect of gamification through a virtual reality on preoperative anxiety in pediatric patients undergoing general anesthesia: a prospective, randomized, and controlled trial. J Clin Med 2018;7(9):284 [FREE Full text] [doi: 10.3390/jcm7090284] [Medline: 30227602]
26. Nayarisseri A, Khandelwal R, Tanwar P, Madhavi M, Sharma D, Thakur G, et al. Curr Drug Targets 2021;22(6):631-655. [doi: 10.2174/1389450122999210104205732] [Medline: 33397265]
27. Hendricks TM, Gutierrez CN, Stulak JM, Dearani JA, Miller JD. The use of virtual reality to reduce preoperative anxiety in first-time sternotomy patients: a randomized controlled pilot trial. Mayo Clin Proc 2020;95(6):1148-1157. [doi: 10.1016/j.mayocp.2020.02.032] [Medline: 32498771]
28. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y. Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: a systematic review. J Biomed Inform 2018;87:138-153 [FREE Full text] [doi: 10.1016/j.jbi.2018.09.002] [Medline: 30267895]
29. Li L, Novillo-Ortiz D, Azzopardi-Muscat N, Kostkova P. Digital data sources and their impact on people's health: a systematic review of systematic reviews. Front Public Health 2021;9:645260 [FREE Full text] [doi: 10.3389/fpubh.2021.645260] [Medline: 34026711]
30. Ho D, Quake SR, McCabe ER, Chng WJ, Chow EK, Ding X, et al. Enabling technologies for personalized and precision medicine. Trends Biotechnol 2020;38(5):497-518 [FREE Full text] [doi: 10.1016/j.tibtech.2019.12.021] [Medline: 31980301]
31. Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision medicine, AI, and the future of personalized health care. Clin Transl Sci 2021;14(1):86-93 [FREE Full text] [doi: 10.1111/cts.12884] [Medline: 32961010]
32. Bertoncello C, Colucci M, Baldovin T, Buja A, Baldo V. How does it work? Factors involved in telemedicine home-interventions effectiveness: a review of reviews. PLoS One 2018;13(11):e0207332 [FREE Full text] [doi: 10.1371/journal.pone.0207332] [Medline: 30440004]
33. De Croon R, Van Houdt L, Htun NN, Štiglic G, Abeele VV, Verbert K. Health recommender systems: systematic review. J Med Internet Res 2021;23(6):e18035 [FREE Full text] [doi: 10.2196/18035] [Medline: 34185014]
34. de Vries H. An integrated approach for understanding health behavior; the I-change model as an example. PBSIJ 2017;2(2):555585. [doi: 10.19080/pbsij.2017.02.555585]

35. Hors-Fraile S, Malwade S, Luna-Perejon F, Amaya C, Civit A, Schneider F, et al. Opening the black box: explaining the process of basing a health recommender system on the I-change behavioral change model. IEEE Access 2019;7:176525-176540. [doi: 10.1109/access.2019.2957696]
36. Repetto C, Gaggioli A, Pallavicini F, Cipresso P, Raspelli S, Riva G. Virtual reality and mobile phones in the treatment of generalized anxiety disorders: a phase-2 clinical trial. Pers Ubiquit Comput 2011;17(2):253-260. [doi: 10.1007/s00779-011-0467-0]
37. Kouroubali A, Kondylakis H, Karadimas E, Kavlentakis G, Simos A, Baños RM, et al. Digital health tools for preoperative stress reduction in integrated care. Eur J Biomed Inform 2019;16(2):7-13 [FREE Full text]
38. Kouroubali A, Kondylakis H, Koumakis L, Papagiannakis G, Zikas P, Katehakis DG. iSupport: building a resilience support tool for improving the health condition of the patient during the care path. Stud Health Technol Inform 2019;261:253-258. [Medline: 31156125]
39. Kondylakis H, Giglioli IAC, Katehakis DG, Aldemir H, Zikas P, Papagiannakis G, et al. A digital health intervention for stress and anxiety relief in perioperative care: protocol for a feasibility randomized controlled trial. JMIR Res Protoc 2022;11(11):e38536 [FREE Full text] [doi: 10.2196/38536] [Medline: 36445734]
40. Mitchell AM, Crane PA, Kim Y. Perceived stress in survivors of suicide: psychometric properties of the perceived stress scale. Res Nurs Health 2008;31(6):576-585. [doi: 10.1002/nur.20284] [Medline: 18449942]
41. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: visual analog scale for pain (VAS Pain), numeric rating scale for pain (NRS Pain), mcGill pain questionnaire (MPQ), short-form mcGill pain questionnaire (SF-MPQ), chronic pain grade scale (CPGS), short form-36 bodily pain scale (SF-36 BPS), and measure of intermittent and constant osteoarthritis pain (ICOAP). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S240-S252 [FREE Full text] [doi: 10.1002/acr.20543] [Medline: 22588748]
42. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
43. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
44. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
45. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the warwick-edinburgh mental well-being scale (WEMWBS): a rasch analysis using data from the Scottish health education population survey. Health Qual Life Outcomes 2009;7:15 [FREE Full text] [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
46. Schwarzer R. General perceived self-efficacy in 14 cultures. URL: https://userpage.fu-berlin.de/~health/world14.htm [accessed 2024-04-01]
47. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
48. Flannery JE, Giuliani NR, Flournoy JC, Pfeifer JH. Neurodevelopmental changes across adolescence in viewing and labeling dynamic peer emotions. Dev Cogn Neurosci 2017;25:113-127 [FREE Full text] [doi: 10.1016/j.dcn.2017.02.003] [Medline: 28262423]
49. Wong CL, Lui MMW, Choi KC. Effects of immersive virtual reality intervention on pain and anxiety among pediatric patients undergoing venipuncture: a study protocol for a randomized controlled trial. Trials 2019;20(1):369 [FREE Full text] [doi: 10.1186/s13063-019-3443-z] [Medline: 31221208]
50. WHO Guideline. Recommendations on digital interventions for health system strengthening. 2023. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2024-11-20]
51. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DH: digital health
GSE: General Self-Efficacy
HADS: Hospital Anxiety and Depression
INRCA: Instituto di Ricovero e Cura per Anziani
LMM: linear mixed model
PAM-13: Patient Activation Measure 13
PANAS: Positive and Negative Affect Scale
SAS: Hospital Universitario Reina Sofia
SJD: Sant Joan de Déu Hospital

SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale
VAS: Visual Analogue Scale
VR: virtual reality

Edited by A Mavragani; submitted 27.10.23; peer-reviewed by V Girishan Prabhu, A Dimitrios; comments to author 12.03.24; revised version received 02.05.24; accepted 14.10.24; published 07.01.25.

Please cite as:
Kondylakis H, Giglioli IAC, Katehakis D, Aldemir H, Zikas P, Papagiannakis G, Hors-Fraile S, González-Sanz PL, Apostolakis K, Stephanidis C, Núñez-Benjumea FJ, Baños-Rivera RM, Fernandez-Luque L, Kouroubali A
Stress Reduction in Perioperative Care: Feasibility Randomized Controlled Trial
J Med Internet Res 2025;27:e54049
URL: https://www.jmir.org/2025/1/e54049
doi:10.2196/54049
PMID:

Original Paper

Topic Modeling of Nursing Issues in the Media During 4 Emerging Infectious Disease Epidemics in South Korea: Descriptive Analysis

Jungok Kim[1*], PhD; Eun Kyoung Yun[1*], PhD

College of Nursing Science, Kyung Hee University, Seoul, Republic of Korea
[*] all authors contributed equally

Corresponding Author:
Eun Kyoung Yun, PhD
College of Nursing Science
Kyung Hee University
26, Kyungheedae-ro
Dongdaemun-gu
Seoul, 02447
Republic of Korea
Phone: 82 2 961 2348
Fax: 82 2 961 9398
Email: ekyun@khu.ac.kr

Abstract

Background: Emerging infectious disease disasters receive extensive media coverage and public attention. Nurse burnout and attrition peak during health crises such as pandemics. However, there is limited research on nursing issues related to repeated emerging infectious disease crises over time.

Objective: The purpose of this study was to analyze and draw implications from changes in key nursing issues reported by the news media during the outbreaks of severe acute respiratory syndrome (SARS; 2003), influenza A (2009), Middle East respiratory syndrome (MERS; 2015), and COVID-19 (2020) in Korea using topic modeling.

Methods: A total of 51,489 news articles were extracted by searching for the keywords “nursing” or “nurse” in the title or body of articles published from April 2003 to May 2021 (during new infectious disease outbreaks) in the open integrated database. The selected news articles were preprocessed then analyzed for text and structure using a 3-step keyword analysis method, latent Dirichlet allocation topic modeling, and keyword network analysis.

Results: Among the 51,489 news articles collected with the search terms “nursing” and “nurse,” 17,285 (33.6%) were selected based on the eligibility criteria and used in the final analysis. Using topic modeling, we derived 5 topics each for SARS, influenza A, and MERS and 6 topics for COVID-19. The themes commonly identified through topic modeling and keyword network analysis across the 4 epidemics were “response to emerging infectious diseases in Korea,” “demand for nurses,” “vulnerability in the work environment,” and “roles and responsibilities of nurses.” Although the topic names were the same, the meanings implied by the comprehensive keywords for each epidemic varied depending on the epidemic and the times.

Conclusions: Analysis of the identified themes and associated keyword network revealed that issues related to nurse shortages, working conditions, and poor treatment were not unique to the COVID-19 pandemic but rather recurring themes from previous epidemics. Our findings can be used to inform strategies to improve the professional roles, work environment, and treatment of nurses during health crises. Suggestions for future nursing-related policy impact and change research are also provided.

(J Med Internet Res 2025;27:e60446) doi:10.2196/60446

KEYWORDS

topic modeling; news articles; nursing issues; text analysis; emerging infectious disease

Introduction

An emerging infectious disease (EID) is a newly recognized or previously known disease characterized by new virulence or spread in previously unaffected areas [1]. Since the 2000s, South Korea has had 4 EIDs: severe acute respiratory syndrome (SARS) in 2003, novel swine-origin influenza A (H1N1) in 2009, Middle East respiratory syndrome (MERS) in 2015, and COVID-19 in 2020. Among them, influenza A and COVID-19 were declared as pandemics by the World Health Organization (WHO). Antiviral treatments were provided free of charge in Korea starting in August 2009 during the SARS outbreak, after which the number of infections and deaths were no longer regularly recorded. Before then, 270 deaths and 2417 infections were recorded, and at the time of manuscript submission, SARS had resulted in an additional 0 deaths and 4 infections according to official reports. Although the exact number of infections has not been counted since August 2009, more than 700,000 cases had been prescribed antiviral drugs from August 2009 to December 2009. During the MERS period, there were 39 deaths and 186 infections. Regarding COVID-19, 34,572,552 confirmed cases and 35,605 deaths were reported from January 20, 2020, to August 30, 2023, the period involving full monitoring [2].

EID catastrophes receive extensive media coverage and public attention. Indeed, public interest in nurses peaks during health crises, such as infectious disease epidemics [3]. EID outbreaks increase public recognition of and attention to the dedication of nurses working on the front lines of the disaster response [4]. Meanwhile, nurses experience physical and psychological burnout as well as professional ethical conflicts owing to the risk of infection, heavy workloads, patient violence, and the stigma of being a carrier of infection [5-8]. Rising rates of nurse burnout and resignation during inadequate infectious disease crisis responses have become global concerns [9,10]. The loss of skilled nurses may lead to a decline in the quality of nursing care and organizational nursing capacity, which may pose a risk to national health crisis responses and public health.

Although infectious disease outbreaks can be traumatic for individuals and communities, they can also lay the groundwork for health system transitions after the outbreak ends, through institutional and policy improvements that identify and address health system weaknesses revealed during the response. With respect to issues in the nursing workforce, scholars have suggested that societal consensus for improvements in working conditions and institutional arrangements for health resources and protection are needed to minimize nurse burnout and attrition in special situations, such as infectious disease outbreaks [6,8,11,12]. At the national level, discussions about appropriate investments and compensation systems for the training and deployment of infectious disease professionals and collaboration with the medical community are needed, whereas at the societal level, increasing public awareness of health workers and building public consensus through media coverage have been emphasized as obligations [7,13,14]. Media's reports of issues in the field of infectious diseases play an important role in forming society's perceptions of an infectious disease response, nursing, and nurses. Problems that are not resolved over a long period of time, occur repeatedly, and increase public interest become social issues, which become part of a public agenda that the general public recognizes as appropriate for the government to solve. This public agenda is then developed and set as a specific policy agenda.

Previous studies analyzing media coverage of infectious disease outbreaks focused on public perceptions of nursing portrayals and issues [3,4,15,16] and were mostly limited to specific epidemic periods. However, these studies did not track how nursing issues were portrayed during recurrent outbreaks since the 2000s, the unique characteristics of each period, and how characteristics have changed. In the context of COVID-19, which was declared as an endemic by the WHO, policies that address the underlying issues and promote a better future urgently need to be developed. Prior studies using content analysis techniques to analyze data collected through standardized questionnaires or media inevitably had limited data to reflect public perceptions and trends. To compensate for this, it is necessary to grasp the meaning inherent in big data such as online news and social media, and topic modeling and keyword network analysis methods, which have been widely used recently to analyze unstructured text in big data, overcome the limitations of manual analysis by humans [17-19].

Therefore, this study aimed to analyze the changes in major nursing issues reported by the news media when 4 new infectious diseases occurred in the 2000s using topic modeling and keyword network analysis methods and derive the implications. Specifically, we aimed to identify the issues that should be considered to effectively respond to new infectious disease crises that may occur in the future and ensure public safety. Our findings could provide the basis for the development of desirable infectious disease response strategies.

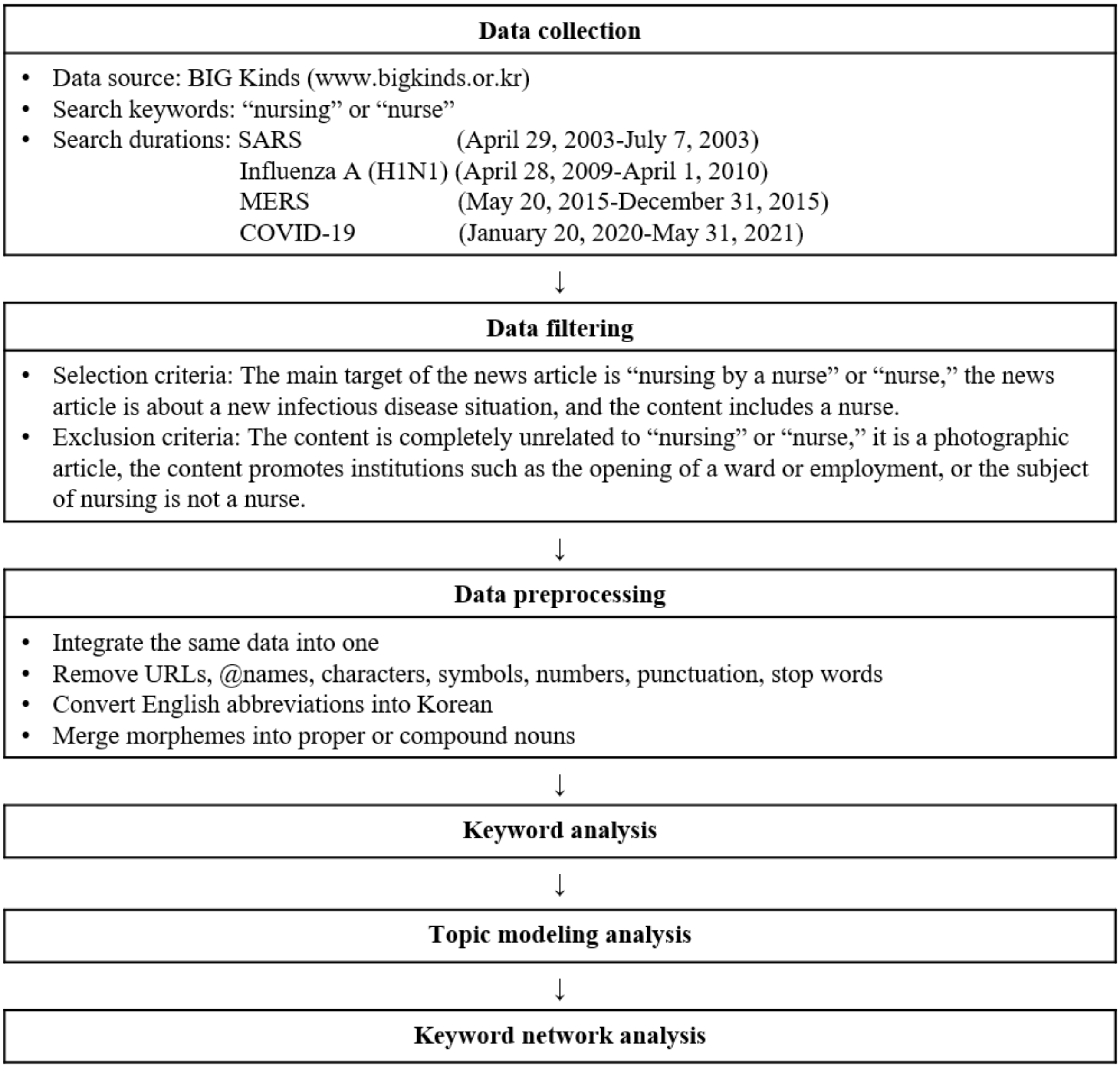
Methods

Data Extraction

We collected data from the Korea Press Foundation's article information website [20], an open-access database containing news articles from 54 media organizations. We collected data published between April 2003 and May 2021 corresponding to the periods of EID epidemics. Specifically, we analyzed news articles published during the SARS, influenza A, MERS, and COVID-19 outbreaks.

A total of 51,489 news articles were extracted by searching for "nursing" or "nurse" as the keyword in the title or text of the article during the outbreak of each new infectious disease. Although the terms "nursing" and "nurse" appeared in the text of these articles, irrelevant content, such as promotional material, was excluded using the selection and removal criteria detailed in Figure 1. These criteria were informed by advice from experts in the field of nursing research and clinical experience. A total of 17,285 news articles were selected using these criteria and used for the final analysis. Figure 1 details the data extraction process used in this study.

Figure 1. Study process. MERS: Middle Eastern respiratory syndrome; SARS: severe acute respiratory syndrome.



Data collection

- Data source: BIG Kinds (www.bigkinds.or.kr)
- Search keywords: “nursing” or “nurse”
- Search durations: SARS (April 29, 2003-July 7, 2003)
  Influenza A (H1N1) (April 28, 2009-April 1, 2010)
  MERS (May 20, 2015-December 31, 2015)
  COVID-19 (January 20, 2020-May 31, 2021)

↓

Data filtering

- Selection criteria: The main target of the news article is “nursing by a nurse” or “nurse,” the news article is about a new infectious disease situation, and the content includes a nurse.
- Exclusion criteria: The content is completely unrelated to “nursing” or “nurse,” it is a photographic article, the content promotes institutions such as the opening of a ward or employment, or the subject of nursing is not a nurse.

↓

Data preprocessing

- Integrate the same data into one
- Remove URLs, @names, characters, symbols, numbers, punctuation, stop words
- Convert English abbreviations into Korean
- Merge morphemes into proper or compound nouns

↓

Keyword analysis

↓

Topic modeling analysis

↓

Keyword network analysis

Data Preprocessing

We preprocessed the metadata of the selected news articles using the *KoNLP* and *tm* packages for natural language processing in R version 4.1.0. Specifically, we removed terms, including symbols, special characters, names of private individuals, and media company names. English terms, such as “SARS” and “OECD,” were occasionally removed during terminology preprocessing; these terms were converted into their Korean equivalents to maintain consistency. During morpheme analysis, complex nouns like “living treatment center” were segmented into “living,” “treatment,” and “center,” causing loss of meaning. To address this, these terms were reconstituted as proper or compound nouns based on N-gram analysis and the context of the original article. Abbreviated Korean terms were replaced with their official names to ensure terminological consistency.

Keyword Analysis

We used the *tm* package in R version 4.1.0 to calculate word frequencies and determine the importance of specific terms using the term frequency-inverse document frequency metric value. Statistical weights were calculated to assess the significance of specific terms within the corpus for keyword analysis of the news articles.

Topic Modeling

Topic modeling classifies and integrates topics in extensive document sets using probabilistic statistical methods to identify latent high-level concepts, extract the topics implicit in the documents, and derive common high-level concepts latent in the topics [21]. We used the latent Dirichlet allocation (LDA) algorithm of the *topicmodels* package of R version 4.1.0 to analyze the topics embedded in the documents. The coding of the analysis was based on the work of Grün and Hornik [16] and open-source code on the GitHub site.

We configured the LDA algorithm by setting the document extraction parameter (α) to 50/k, and the previous parameter delta (or eta) of the word distribution constituting the topic was set to 0.1. Gibbs sampling was used to estimate the topic distribution for the document, with 1000 repetitions used for Gibbs sampling [16,22].

We also conducted a quantitative analysis to select the appropriate number of topics. We calculated perplexity values ranging from 2 to 20 using the *topic models* and *LDAvis* packages of the R program. The topic modeling analysis was repeated by applying different numbers of topics based on the perplexity values. For the analysis, we repeated the adjustment of the λ value of the LDA visualization and reviewed the original text of the news articles to extract a list of the 30 most representative keywords for each topic.

Keyword Network Analysis

Keyword network analysis visualizes keyword relationships by calculating co-occurrence frequencies and metrics such as connection centrality, proximity centrality, and mediation centrality [23]. We performed keyword network analysis using the *Network*, *sna*, and *ggplot2* packages in the R program to improve the specific content and structure of the topics and the words constituting the topics derived from thematic modeling. Considering the visibility of the network around thematic keywords, the correlation coefficient was established, and the relationship between the simultaneous appearance of words was analyzed.

The top 15 keywords were selected as the main keywords for each topic through quantitative analyses including topic modeling analysis, keyword network analysis, and verification by nursing experts, and the topic names reflected the same higher-level concept to examine the trends of the topics and keywords that make up the topic by period of the new infectious disease based on the main keywords. In this study, 15 keywords for the new infectious disease topic and topic are presented.

Ethical Considerations

All news article data analyzed in this study were publicly available and not subject to copyright restrictions. The Institutional Review Board of Kyung Hee University approved this study (IRB number KHSIRB-21-334(EA)).

Results

News and Keyword Counts

After preprocessing the data from a total of 17,285 news articles, we extracted 1974 words for the SARS epidemic, 9520 words for influenza A, 15,639 words for MERS, and 24,888 words for COVID-19, resulting in a total of 52,021 words. The numbers of news articles and keyword types analyzed increased progressively across successive infectious disease outbreaks, and the average daily number of news articles was the highest during the MERS period.

Top Topics and Keywords During Epidemics

Table 1 presents the keywords and topic names identified for each epidemic period. We selected 5 topics each for SARS (perplexity=1599.241, λ=0.7), influenza A (perplexity=977.101, λ = 0.8), and MERS (perplexity=1717.432, λ=0.7) and 6 topics for COVID-19 (perplexity=2515.911, λ=0.6). We sorted the topics in the order of the largest percentage of word clusters that make up the topic and identified them as themes.

Table 1. Themes and keywords of the topics.

Themes and topics	Top keywords contributing to the topic model	Token, %
Theme 1: Response to emerging infectious diseases in Korea		
SARS[a] (Topic 3)	SARS, patient, hospital bed, infection, isolation, disinfection, suspected patient, outbreak, designation, health personnel, presumed patient, National Institute of health, constitute, dedication, countermeasure	18.2
Influenza A (H1N1; Topic 2)	Influenza A virus(H1N1), infection, hub-hospital, spread, outbreak, prevention, medical staff, treatment, mask, test, suspected case, confirmed case, countermeasure, the dead, Tamiflu	21.7
MERS[b] (Topic 3)	MERS, confirmed case, infection, addition, medical staff, hospital, emergency room, treatment, exposed, isolation, Central MERS Control Countermeasure Headquarters, closed, S-hospital[c], K-hospital[c], P-hospital[c]	21.5
COVID-19 (Topic 2)	COVID-19, medical staff, mask, protective clothing, world, treatment, China, pandemic, virus, outbreak, hospital room, situation, the dead, hospitalization, pneumonia	20.8
Theme 2: Demand for nurses		
SARS (Topic 1)	The aged, nursing care, society, nurse, facility, family, role, profession, psychiatric, system, push ahead, policy, dementia, the disabled, welfare	31.2
Influenza A (H1N1; Topic 1)	Nurse, nursing college, education, service, push ahead, nursing department, the aged, establishment, system, manpower, expansion, employment, quota, the disabled, insufficiency	26.3
MERS (Topic 1)	Expansion, comprehensive nursing service, push ahead, countermeasure, hospital, provide, nursing staff, system, infection disease, caregiving, nurse's aide, measure, reinforce, improvement, nationwide	23.6
COVID-19 (Topic 3)	Securing, support, treatment, medical worker, insufficiency, deploy, dispatch, nurse, critical patient, Daegu, living treatment center, response, allowance, recruitment, dedicated hospital	17.7
Theme 3: Vulnerability in the working environment		
SARS (Topic 2)	Son and daughter, woman, child, risk, exercise, person, workplace, night duty, trust, health, nurse, research, thought, parent, prevention	19.7
Influenza A (H1N1; Topic 4)	Doctor, nurse, patient, medical, hospital bed, OECD[d], level, medical service, facility, medical expenses, Korea, region, insufficiency, intensive care unit, increase	18.7
MERS (Topic 5)	MERS, isolation, patient, suspicion, contact, action, duty, hospital, Korea Centers for Disease Control and Prevention, nurse, infection, test result, infected person, addition, emergency room	16
COVID-19 (Topic 4)	Confirmed case, infection, nurse, hospital, group infection, duty, contact, convalescent hospital, screening center, rest, emergency room, working environment, site, manpower, protective clothing	17.5
Theme 4: Roles and responsibilities of nursing professionals		
SARS (Topic 5)	Nurse, hospital, suspicion, blood, doctor, use, uncover, sexual harassment, nurse's aide, investigation, contravention, drunk driving, illegality, measurement, injection	15.4
Influenza A (H1N1; Topic 3)	Hospital, nurse, doctor, treatment, family, death, incident, contravene, criminal investigation, grandmother, death with dignity, obstetrics and gynecology, adjudge, false, infringement	21.1
MERS (Topic 2)	Hospital, medical staff, nurse, doctor, treatment, family, thought, person, heart, intensive care unit, child, oneself, protective clothing, gratitude, cheer	21.8
COVID-19 (Topic 1)	Doctor, government, Korean Medical Association, expansion, policy, push ahead, national assembly, strike, region, improvement, medical resident, Korean Nursing Association, the public, medical school quota, found	23.9
Theme 5: Government response and public opinion		
MERS (Topic 4)	MERS, situation, virus, the public, outbreak, symptom, response, fear, transmission, government, information, preventive measures, open, anxiety, the Middle East	17

Themes and topics	Top keywords contributing to the topic model	Token, %
COVID-19 (Topic 5)	Nurse, president, the public, gratitude, heart, commitment, encouragement, message, social network service, cheer, consolation, controversy, criticism, expression, labor	13
Theme 6: Vaccination for emerging infectious diseases		
Influenza A (H1N1; Topic 5)	Vaccination, student, influenza A virus(H1N1), school, vaccination, public health center, start, doctor, nurse, constitute, school nurse, plan, securing, school parent, insufficiency	12.2
COVID-19 (Topic 6)	Vaccine, inoculation, AstraZeneca, injection, Korea Disease Control and Prevention Agency, start, public health center, side effect, Pfizer, effect, nurse, nursing facility, targets, worker, launch	7.1
Theme 7: Response to emerging infectious diseases in other countries		
SARS (Topic 4)	SARS, China, the dead, Taiwan, spread, outbreak, WHO[e], Hong Kong, Singapore, infected person, death, Canada, Vietnam, transmission	15.5

[a] SARS: severe acute respiratory syndrome.

[b] MERS: Middle East respiratory syndrome.

[c] Abbreviations for specific hospital names.

[d] OECD: Organisation for Economic Co-operation and Development.

[e] WHO: World Health Organization.

Trends in Nursing-Related Topics in Media Coverage

According to the topic modeling analysis, we classified a total of 7 topics from all news articles, and the topics commonly derived from the 4 infectious disease periods were “response to emerging infectious diseases in Korea,” “demand for nurses,” “vulnerability in the working environment,” and “roles and responsibilities of nursing professionals.” Although the overarching concepts remained consistent, the keywords comprising each topic varied across the duration of infectious diseases. Notably, the keywords that comprised the topics for each epidemic period differed. We observed that the topic “vaccination for emerging infectious diseases” appeared only in the influenza A and COVID-19 pandemics, and “government response and public opinion” appeared only in the MERS and COVID-19 periods. Moreover, the topic “response to emerging infectious diseases in other countries” appeared only during SARS (Table 2). Table 1 shows the topics and word clusters identified by our model.

Table 2. Trends in nursing-related topics in media coverage, reported as tokens for each outbreak.

Trends	SARS[a], %	Influenza A (H1N1), %	MERS[b], %	COVID-19, %
Response to emerging infectious diseases in Korea	18.2	21.7	21.5	20.8
Demand for nurses	31.2	26.3	23.6	17.7
Vulnerability in the working environment	19.7	18.7	16	17.5
Roles and responsibilities of nursing professionals	15.4	21.1	21.8	23.9
Government response and public opinion	—[c]	—	17.1	13
Vaccination for emerging infectious diseases	—	12.2	—	7.1
Response to emerging infectious diseases in other countries	15.5	—	—	—

[a] SARS: severe acute respiratory syndrome.

[b] MERS: Middle East respiratory syndrome.

[c] Not applicable.

Theme 1. Response to Emerging Infectious Diseases in Korea

During the 4 periods, the “response to emerging infectious diseases in Korea” topic accounted for approximately 20%, ranking third during the SARS and MERS outbreaks and second during the influenza A and COVID-19 outbreaks. The keyword patterns were similar between SARS and MERS and between influenza A and COVID-19, reflecting trends observed in the keyword frequency analysis.

Keywords for each period include the name of the respective virus and terms describing the outbreak situation, such as “infection,” “suspected patient,” “outbreak,” and “presumed patient” during the SARS period; “infection,” “spread,” “outbreak,” “confirmed patient,” and “deaths” during the influenza A period; “confirmed case,” “infection,” “additional,” “quarantine,” and “closure” during the MERS period; and

“China,” “pandemic,” “outbreak,” “situation,” “world,” and “the dead” during the COVID-19 period, indicating the evolving scale and severity of infectious disease outbreaks in Korea over time.

In addition, keywords related to infectious disease response and preparedness such as “isolation,” “disinfection,” “designation,” “National Institute of Health,” “constitute,” and “dedication” during SARS changed to “hub-hospital,” “prevention,” “treatment,” “mask,” “countermeasure,” “test,” and “Tamiflu” during the influenza A and MERS periods and keywords related to preparation such as specific hospital names and “hospital,” “emergency room,” “Central MERS Control Countermeasure Headquarters,” “treatment,” “isolation,” and “closed” during the COVID-19 period. Keywords related to preparation, such as “mask,” “protective clothing,” “treatment,” and “hospitalization” diminished in number, with response-focused keywords emerging in their place.

Theme 2. Demand for Nurses

The topic “demand for nurses” ranked first and represented the largest proportion during the SARS to MERS periods; during COVID-19, the topics of “roles and responsibilities of nursing professionals” and “response to emerging infectious diseases in Korea” represented higher proportions, with “demand for nurses” dropping to third place. When examining the keyword flow of topics by period, the keywords “system,” “policy,” “countermeasure,” and “response” were associated with “the aged,” “society,” “facility,” “psychiatric,” “dementia,” “the disabled,” and “welfare” during the SARS period; “service,” “the aged,” and “the disabled” during the influenza A period; “comprehensive nursing service,” “hospital,” and “infection disease” during the MERS period; and “Daegu,” “living treatment center,” and “dedicated hospital” during the COVID-19 period, showing that the areas in which nurses were needed and the reasons for a nursing demand changed as a result of health and welfare systems and policies for infectious disease responses.

In addition, “nursing care,” “family,” “profession,” and “role” during the SARS period; “manpower” during the influenza A period; “nursing staff,” “caregiving,” and “nurse’s aide” during the MERS period; and “treatment,” “medical worker,” and “critical patient” during the COVID-19 period represent keywords related to the perception of the role of nurses. The keywords “nursing college,” “nursing department,” “expansion,” and “quota” during the influenza A period; “nurse’s aide” during the MERS period; and “support,” “deploy,” “dispatch,” “allowance,” and “recruitment” during the COVID-19 period represented changes in ways to meet the nursing demand by period.

Theme 3. Vulnerability in the Working Environment

The topic “vulnerability of the working environment” had the second highest weight during the SARS period, but its weight decreased during the influenza A and MERS periods. Its weight increased to 4th place during the COVID-19 period. During the SARS period, keywords such as “son and daughter,” “woman,” “parent,” “workplace,” “thought,” “risk,” “exercise,” “night duty,” “health,” and “prevention” were topics about women’s conflicts between work and childcare and health risks related to work, while during the influenza A period, keywords such as “patient,” “hospital bed,” “OECD,” “level,” “facility,” “insufficiency,” “intensive care unit,” and “region” showed a quantitative shortage of nurses and medical facilities. During the MERS period, news reports focused on topics such as “contact,” “infection,” “addition,” “duty,” and “emergency room” regarding nurse infections, while during the COVID-19 period, news reports focused on topics such as “group infection,” “duty,” “contact,” “working environment,” “site,” and “rest.”

Theme 4. Roles and Responsibilities of Nursing Professionals

The topic “roles and responsibilities of nursing professionals,” which was the lowest during the SARS period, showed an upward trend from the influenza A period to the COVID-19 period, and during the COVID-19 period, it had the largest change among the 6 derived topics.

The keywords of the “roles and responsibilities of nursing professionals” topic changed mainly based on specific incidents reflecting the situation at the time. During the SARS period, specific incidents related to keywords such as “blood,” “drunk driving,” “hospital,” and “sexual harassment” included a case in which a nurse switched the blood of a drunk driving colleague with her own blood and submitted it, a case in which a doctor sexually harassed a nurse, and an illegal medical practice by a nurse or nursing assistant. During the influenza A period, keywords related to the incidents and accidents at the time, such as a case in which a newborn was switched due to a nurse’s mistake, included “adjudge” and “obstetrics and gynecology” and related to bioethics, such as euthanasia, included “grandmother” and “death with dignity.” During the MERS period, keywords such as “family,” “thought,” “heart,” “gratitude,” and “cheer” showed support for doctors and nurses who were infected during the MERS response process, which occurred mainly in specific hospitals. During the COVID-19 period, keywords such as “medical school quota,” “expansion,” “Korean Medical Association,” and “Korean Nurses Association” mainly focused on issues such as doctors’ strikes and nurses who stayed by patients’ sides during the infectious disease outbreak.

Theme 5-7. Topics Derived Only During a Particular Infectious Disease Period

The fifth topic “government response and public opinion” was derived during the MERS and COVID-19 periods, and it had the second lowest proportion in both periods. During the MERS period, from the keywords “virus,” “the public,” “outbreak,” “anxiety,” “information,” “fear,” and “open,” we could infer that the public experienced anxiety due to a lack of information on new infectious diseases from the government’s disclosure of information. During the COVID-19 period, keywords such as “nurse,” “president,” “gratitude,” “encouragement,” “controversy,” and “criticism” emphasized the controversy over the president’s thank you message to nurses during the doctors’ strike and the public’s support for the struggles of medical staff, such as the “Thanks to Challenge,” as social issues.

"Vaccination for emerging infectious diseases" was derived only during the influenza A and COVID-19 periods, and this topic had the lowest proportion of all topics in both periods. During the influenza A period, the keywords "student," "school," and "school parent" indicate that children and adolescents were the main targets of vaccination, while during the COVID-19 period, the keywords "nursing facility," "targets," and "worker" show that the targets of vaccination changed to older adults and health care workers. "AstraZeneca," "side effects," "Pfizer," and "effect," which appeared only during the COVID-19 period, indicate the situation of and public interest in the development of COVID-19 vaccines, which were not available at the beginning of the outbreak, and vaccination.

The main keywords of the "response to emerging infectious diseases in other countries" topic, which was derived only during the SARS period, consisted of the names of countries that experienced significant damage from new infectious diseases, such as "China," "Taiwan," and "Hong Kong," and keywords about the situation of the outbreak, such as "death," "outbreak," "addition," and "transmission."

Discussion

Principal Findings

This study was based on the basic premise that, in public health crises such as new infectious diseases, media reports related to the field, including nursing and nurses, become more active, reflecting high social interest in major issues. This study attempted to explore the problems that were highlighted during 4 new infectious diseases that occurred in Korea and considerations regarding social awareness and improvement to policies in relation to major nursing issues. Therefore, this study compared and analyzed the main agenda and potential meaning of news articles on nursing and nurses for each new infectious disease period using the big data analysis techniques of topic modeling and keyword network analysis. In addition, we discussed the understanding of current nursing phenomena and the trends and implications of public opinion surrounding nursing based on the results.

As a result of topic modeling analysis, 7 topics were classified from all news articles. The commonly derived topics from the 4 infectious disease periods included "demand for nurses," "response to emerging infectious diseases in Korea," "vulnerability in the working environment," and "roles and responsibilities of nursing professionals." Topics with the same higher concept had changes in the associated keywords according to the infectious disease period. The following sections discuss the potential meaning in each topic by referring to the visualization results of the related keyword network map, focusing on the topics commonly derived during the new infectious disease periods.

Response to Emerging Infectious Diseases in Korea

First, the keywords for the topics during the SARS, MERS, influenza A, and COVID-19 periods showed similar patterns when analyzing the relevance of the topic by period. Depending on the duration of the infectious disease, the keywords "outbreak," "spread," "addition," and "pandemic" were used to express the seriousness of the domestic infectious disease situation. The potential meaning of the words related to the keywords "National Institute of Health," "dedication," and "designation" during SARS and "hub-hospital" during influenza A was the government's practical policy proposal to prepare for new infectious diseases. These policies included the designation of regional hub hospitals and securing state-designated inpatient treatment beds that were proposed during SARS and influenza A and were implemented and partially improved after the end of those infectious disease periods. However, the number of public medical resources, such as treatment beds and health professionals, that could be practically used was significantly insufficient for the number of infected patients during the MERS and COVID-19 periods [24].

In addition, keywords such as "medical staff," "infection," "closed," "hospital room," and "hospitalization" during MERS and COVID-19 highlighted vulnerabilities in emergency room environments, including inadequate infection control measures and gaps in infection control in nursing hospitals and nursing facilities where older adult patients, who are vulnerable to infection, are mainly hospitalized. The emergency room is where infected patients primarily access the hospital. Problems including the lack of facilities such as separate isolation zones and negative pressure isolation rooms for the treatment of infected patients revealed during the MERS period, the lack of education on infection control, and the response of emergency medical workers increase the risk of medical staff becoming infected and are viciously linked to patient infection and the lack of response personnel [25]. In addition, in emergency rooms, there is a greater need for systematic infection control by professional personnel such as nurses because older adult patients in nursing hospitals and facilities are more likely to contract infectious diseases due to low immunity and group living and the risk of death from infection is high [26,27]. This is a matter to be considered first in terms of policy for infection control in environments where patients and medical personnel are at a high risk of infection. After MERS, the enforcement regulations outlined in the Emergency Medical Service Act were revised to improve emergency room facilities and manpower standards [28], and through the revision of the enforcement regulations of the Medical Act during the COVID-19 period, the standards and fees for mandatory infection control rooms for infection control were expanded [29]. However, the effectiveness of the policy is still limited because it does not resolve gaps in infection control outside the scope of the policy, such as the lack of infection control experts and nursing hospitals and facilities with fewer than 100 beds. Although policies such as the revision of the Emergency Medical Service Act aimed to address these gaps, their effectiveness remains limited due to insufficient implementation and resource allocation. Similar challenges in emergency room preparedness were noted in countries such as Italy and the United States during the COVID-19 pandemic, underscoring the global nature of these vulnerabilities.

Demand for Nurses

In the "demand for nurses" topic, the need to increase the number of nurses has been steadily highlighted over the past

20 years, and the keyword “insufficiency,” which appeared only during the influenza A and COVID-19 periods shows that the shortage of nurses during the COVID-19 period was also a problem that was raised during the influenza A period. In the topic analysis results, the focus on nursing demand changed from health and welfare services to responding to infectious diseases, and in the change of keywords over time, the area of public interest also changed according to the reason for the demand in nurses and the field of activity. The areas where nurses’ activities are performed are diverse and include medical institutions, communities, schools, and businesses; nurses are also recognized as particularly indispensable for the national welfare service, especially in a society where the number of older adult and single-person households is increasing [30,31]. Among the government’s systems related to the demand for nurses, the comprehensive nursing service (currently, the integrated nursing care service), which was introduced in 2013 to reduce the burden of care and expenses for people in a nuclear family and an aging society, accelerated the expansion of the service, as domestic nursing and door-to-door cultural issues were raised in terms of hospital infection control during the MERS period in 2015. Nurses with specialized knowledge and skills were recognized as a major element of the service, which proceeded to actively use nurses [32,33]. Many previous studies on integrated nursing care services have shown positive effects of these services for improving patient safety, reducing hospital infections, and increasing patient satisfaction with medical services [32,34,35]. Accordingly, there seems to be no difference in public opinion on the need for a change in the demographic and social structure, the response to infectious disasters at the health care level, a quantitative increase in nurse demand and quantity, and policies to increase the number nurses.

In terms of meeting the demand for nursing during periods of new infectious diseases, the focus was on recruiting nurses and distributing manpower to prevent the infection of the vulnerable, such as older adults, and school quarantine during the SARS and influenza A periods. Changes for the issue of nursing supply were mainly made to comprehensive nursing services as a preventive measure against hospital infections. During the COVID-19 period, there was a shortage of nursing personnel who could be immediately placed in response to infectious diseases in medical institutions in nonmetropolitan areas such as Daegu, life treatment centers, and hospitals dedicated to infectious diseases. The government filled the demand for nurses with temporary measures through the dispatch of nursing personnel recruited through the input of nursing officers and allowances. However, the urgent recruitment of the personnel needed in the field of infectious diseases was ineffective due to insufficient nursing experience, education, and training [36,37]. It is believed that the gap in the nursing field continues because there is a real need for nurses with the experience and capacity to care for infectious disease patients and intensive care patients, while the government’s policies for and public opinion about increasing the nursing supply focused purely on the number of nurses rather than the quality and capacity of nursing.

The fact that the policy to increase the number of nurses recognized nursing as a job that not only resolves the imbalance of local medical personnel or meets medical needs but also resolves employment difficulties and guarantees employment after graduation is similar to the ideas expressed in the original news article [38] titled “Graduation is Employment, Nursing and Explosion of Popularity” and research showing that high school students who are about to pursue careers perceive nursing as an economically stable and good job compared with other occupations [39]. In addition, in research analyzing the meaning of work for new nurses in 2021, more than 50% of the nurses had family and relatives who recommended they pursue nursing, or they applied to the nursing department because they were able to easily get a job after graduation [40].

Since 2008, the year just before the influenza A outbreak, Korea implemented a policy to establish a nursing department and increase the number of admissions. As a result of this policy, during the swine flu period, there were 414,000 licensed nurses in 2019, and 20,000 new nurses entered the workforce every year [30]. However, there are only 215,000 nurses working in medical and health institutions, which represents about one-half of the licensed nurses [41]. According to data from a survey on the status of hospital nursing staff placements every year since 2010, the turnover of new nurses increased from 30.5% in 2011 to 47.7% in 2020, and the number one reason for resignation within a year after the announcement was “work maladjustment” [42,43]. The burden of work on new nurses was correlated with the reality shock caused by the conflict between the hospital organizational culture and nursing professionals [44]. Nurses are aware of the risk of infection at infectious disease sites, but they have a sense of calling and must perform patient care at the forefront [45,46]. In patient care, the experience of a conflict between infection risk situations and professional beliefs increased nurses’ intentions to leave more than the general situation of care [10,47]. Based on these preceding studies, there is likely a difference between the actual clinical field and the public’s perception of nursing in terms of the purpose of employment identified in the results of this study. Although economic feasibility cannot be excluded from policies to resolve employment difficulties, it is necessary to consider nursing professionals in policy and media in order to prevent the departure of new nurses, increase the number of experienced nurses, and change the public’s perception of nurses.

Vulnerability in the Working Environment

In the “demand for nurses” topic during the influenza A and COVID-19 outbreaks, a critical issue was the shortage of skilled nurses—particularly in nonmetropolitan areas—in addition to skills to handle special equipment such as ventilators and extracorporeal membrane oxygen supply required for intensive care nursing as well as clinical experience [36]. Given the strong link between nursing demand, supply, and working conditions, the topics “demand for nurses” and “vulnerability in the working environment” are closely interconnected [37] and can be interpreted together. In this study, hot topics included conflicts between women’s work-family balance during the SARS period, health risks due to night work, domestic medical facilities and manpower at a lower level than in OECD countries during the influenza A period, and the shortage of nurses in nonmetropolitan areas. The changes in keywords such as high secondary infections of nurses related to new infections, emergency rooms, and nursing hospitals; work weighting; and society's expectations for sacrifice and commitment should be considered as major variables to resolve the shortage of nurses.

According to a survey by the Hospital Nurses Association, the average resignation rate of domestic nurses over the past decade has been over 13%. When the reasons for resignation in 2010 and 2020 were compared, the second highest reason of "marriage, childbirth, and childcare" in 2010 fell significantly from 14.8% of responses in 2010 to 6.3% of responses in 2020. "Work maladjustment," which was ranked third (13.1%) in 2010, was ranked first (17.1%) in 2020, and "transition to other jobs," which was ranked second (5.9%) in 2010, increased to 12.2% in 2020 [42,43]. This is likely the result of government policy and efforts to improve the conflict that comes from the balance between work and family for female nurses, which appeared in the topic during SARS. However, high-risk working environments for infection and the lack of experienced nurses during MERS and COVID-19, during which there was a focus on the quantitative increase of nurses following the influenza period; the policy to increase the number of nurses; and the poor treatment suggest that more active interventions and improvement are needed at the organizational level.

Inadequate facilities in which to rest and the social stigma of being perceived as infection carriers—both personally and for family members—accelerate burnout and increase turnover intentions among nurses, and secondary infections among nurses create a vicious cycle of labor shortages and increased patient infections [48,49]. The government expanded the admission quota for nursing colleges starting in 2008 in an effort to increase the absolute total number of nurses, but it did not have a significant impact on resolving the shortage of nurses in local small and medium-sized hospitals [50]. In addition, in March 2018, the Ministry of Health and Welfare announced policies aimed at improving nurses' working conditions, addressing human rights violations such as workplace bullying and sexual harassment, expanding nursing personnel, strengthening professionalism, improving the quality of nursing services, and creating a policy foundation for nursing personnel [51]. In this regard, the system for calculating inpatient nursing management fees was revised to base payments on patient-to-nurse ratios according to the actual situation in local hospitals in April 2018. The system was improved so that additional profits from hospitalization fees generated by changes based on the number of patients were used to improve the treatment of nurses. In addition, the night-only nurse management fee and the night-care fee for general wards established in October 2019 were applied to general hospitals and hospitals outside of Seoul so that additional allowances for night work could be paid. However, since both systems are merely recommendations, there were practical limitations for nurses to feel that their treatment was improved [52]. In February 2019, a Task Force Team was promoted to oversee nursing policies such as managing the supply and demand of nursing personnel and improving the working environment. In May 2021, a nursing policy department was established in the Ministry of Health and Welfare with social consensus on the need for a dedicated department during the COVID-19 period [42]. As such, the government has continuously implemented policies to address the shortage of nurses, and evaluation of the effectiveness of some policies is expected to take some time. This is an opportune time to reconsider whether fragmented and temporary approaches to nursing challenges have exacerbated systemic issues, such as the revision of the medical law that tried to reorganize nursing assistants from the "demand for nurses" topic during the MERS period or the "vulnerability in the working environment" topic, which is related to a high risk of infectious diseases and the need for high-intensity work. In addition, active monitoring is needed to determine how closely a series of policies has approached the practical problems of the nursing field, and feedback mechanisms during a changing infectious disease situation are needed. Although these policies mark progress, their voluntary nature limits their effectiveness in achieving widespread improvements.

Roles and Responsibilities of Nursing Professionals

The keyword changes in the topic of "roles and responsibilities of nursing professionals" mainly included incidents and accidents related to the morality of nurses, such as drunk driving and illegal medical practices during the SARS and influenza A periods. Given the focus on keywords related to infectious diseases during the MERS and COVID-19 periods, the conflict over work ethics in the nursing of infected patients versus the nurse's own safety and social expectations was considered a major agenda.

Since the image of nurses formed in the media has been considered to influence potential future nurses' career choices and intentions and some stories about nurses can negatively influence not only nursing supply but also potential improvements in the working environment and treatment of nurses, many previous studies have proposed monitoring of the images of nursing expressed in the media and internal management of the nursing community [3,53].

The analysis of correlations between keywords during the MERS period highlighted social expectations for nurses in special situations such as public health disasters as well as conflicts between nurses' protective instincts and work ethics and their families and colleagues. In this regard, the original news articles reviewed in this study described the critical condition of medical personnel who had secondary MERS infections and experiences with social isolation, such as nurses' children being ostracized by their neighbors due to the stigma of working with MERS-infected patients [54]. The ambivalence of social expectations for nurses responding to infectious diseases is reflected by the description of medical personnel as heroes through the "Thanks to the Challenge" campaign and the praise for volunteer nurses in Daegu and Gyeongbuk. The precedent of the government's legal response to the collective resignation of Taiwanese medical staff due to the infection and death of medical personnel is another example and was captured in the topic "response to emerging infectious diseases in other countries" during SARS. The ambivalence was also present in an overseas example [3] that described a nurse volunteering at an infectious disease site during the Ebola outbreak as a virus eradication hero versus an infection spreader and villain. Since social stigma and isolation during an infectious disease outbreak add to the mental and social stress of nurses and influence nursing and turnover intentions, the following are needed:

accurate information delivery by the media, a stable medical system, and an organizational protection and support system [48,49].

Meanwhile, in the context of the COVID-19 period, the government attempted to implement a policy to expand the number of medical schools to respond to infectious diseases, and the related medical strike led to the aggravation of nurses' work [55]. Previous studies have highlighted the 2-sided phenomenon of social expectations, which consists of systemic problems such as poor hospital infection control, inadequate government response, and increased public fear through negative media reports rather than the attitude of medical personnel, including nurses, when responding to infectious diseases [6,48]. This is similar to the relationships between the topics of "government response and public opinion" and "roles and responsibilities of nursing professionals," which include indiscriminate disclosure of information about infectious diseases in the media during the MERS period and the controversy over the president's thank you message to nurses during a doctor strike during the COVID-19 period. This suggests that active interest and the adoption of a leading role by the government are needed to avoid delays in policy and social consensus as a result of insufficient government responses and issue-oriented media reports.

Limitations

In this study, limitations were related to the search used to identify issues in the nursing field as well as the data collection period, which was restricted to the periods around outbreaks of new infectious diseases and limited the ability to compare the situation before and after an outbreak. Therefore, future research should adopt a multifaceted approach that incorporates time series analysis, dynamic topic modeling, and expanded data collection periods to capture broader trends. In addition, as this study started during the COVID-19 outbreak and we were unable to predict when the outbreak would end, we collected data only until May 31, 2021, due to limited time available for data analysis. Therefore, any changes in the situation after that data were not reflected in the research results, and we recommend collecting and analyzing data after that date.

Conclusion

This study analyzed the relevance of nursing-related agendas in news articles during 4 major EID outbreaks. Our analysis revealed that nurse shortages, poor working conditions, and inadequate treatment of nurses were recurring issues across the epidemics and were not unique to the COVID-19 pandemic. Although society and nursing professionals agree on the essential role of nurses in the health sector, significant differences persist in perceptions of their roles and responsibilities. The results suggest that retaining experienced nurses to strengthen crisis response capabilities should be closely linked to the working environment. Addressing nurse shortages requires not only increasing their numbers but also implementing substantial policy support to ensure safe working environments, better treatment conditions, and heightened societal recognition of their roles. Medical organizations, professional associations, governments, and the media should develop policies that incorporate nurses' perspectives. These efforts can foster positive social consensus on nursing and support a stable health care system during health crises. Our findings are expected to be used as a basis for establishing strategies to improve the professional role, environment, and treatment of nurses during health crisis situations. Finally, these findings emphasize the need for holistic reforms to strengthen nursing's role in mitigating future health crises. We recommend further research on the effects and changes of nursing-related policies.

Acknowledgments

This research was supported by the Government-wide R&D Fund for Infectious Disease Research (GFID), Republic of Korea (grant number RS-2022-KH125928 [HG22C0051]).

Authors' Contributions

EKY and JO collaboratively developed the study design and co-authored the initial manuscript draft. EKY supervised the study and validated the findings, while JO collected and preprocessed the data and conducted the analysis. Both authors interpreted the results and contributed to the writing and revision of the final manuscript.

Conflicts of Interest

None declared.

References

1. McArthur DB. Emerging infectious diseases. Nurs Clin North Am 2019 Jun;54(2):297-311 [FREE Full text] [doi: 10.1016/j.cnur.2019.02.006] [Medline: 31027668]
2. Infectious Disease Information. Korea Disease Control and Prevention Agency. URL: https://dportal.kdca.go.kr/pot/cv/trend/dmstc/selectMntrgSttus.do [accessed 2024-11-16]
3. McGillis Hall L, Kashin J. Public understanding of the role of nurses During Ebola. J Nurs Scholarsh 2016 Jan;48(1):91-97. [doi: 10.1111/jnu.12182] [Medline: 26642005]
4. Nayna Schwerdtle P, Connell CJ, Lee S, Plummer V, Russo PL, Endacott R, et al. Nurse expertise: a critical resource in the COVID-19 pandemic response. Ann Glob Health 2020 May 11;86(1):49 [FREE Full text] [doi: 10.5334/aogh.2898] [Medline: 32435602]

5. Chitty KK. Professional Nursing: Concepts and Challenges. Philadelphia, PA: Saunders; 2005.
6. Choi EK. Ethical responses to public health emergencies: the 2015 MERS outbreak in South Korea. Korean J Med Ethics 2016 Sep 30;19(3):358-374 [FREE Full text] [doi: 10.35301/ksme.2016.19.3.358]
7. Donelan K, Buerhaus P, DesRoches C, Dittus R, Dutwin D. Public perceptions of nursing careers: the influence of the media and nursing shortages. Nurs Econ 2008;26(3):143-50, 165. [Medline: 18616051]
8. Ehrenstein BP, Hanses F, Salzberger B. Influenza pandemic and professional duty: family or patients first? A survey of hospital employees. BMC Public Health 2006 Dec 28;6:311 [FREE Full text] [doi: 10.1186/1471-2458-6-311] [Medline: 17192198]
9. The COVID-19 Effect: World's nurses facing mass trauma, an immediate danger to the profession and future of our health systems. International Council of Nurses. 2021 Jan 13. URL: https://www.icn.ch/news/covid-19-effect-worlds-nurses-facing-mass-trauma-immediate-danger-profession-and-future-our [accessed 2024-05-09]
10. Labrague LJ, de Los Santos JAA. Fear of COVID-19, psychological distress, work satisfaction and turnover intention among frontline nurses. J Nurs Manag 2021 Apr;29(3):395-403 [FREE Full text] [doi: 10.1111/jonm.13168] [Medline: 32985046]
11. Ruderman C, Tracy CS, Bensimon CM, Bernstein M, Hawryluck L, Shaul RZ, et al. On pandemics and the duty to care: whose duty? who cares? BMC Med Ethics 2006 Apr 20;7:E5 [FREE Full text] [doi: 10.1186/1472-6939-7-5] [Medline: 16626488]
12. Sokol DK. Virulent epidemics and scope of healthcare workers' duty of care. Emerg Infect Dis 2006 Aug;12(8):1238-1241 [FREE Full text] [doi: 10.3201/eid1208.060360] [Medline: 16965703]
13. Brody H, Avery EN. Medicine's duty to treat pandemic illness: solidarity and vulnerability. Hastings Cent Rep 2009;39(1):40-48. [doi: 10.1353/hcr.0.0104] [Medline: 19213194]
14. Martin SD. Nurses' ability and willingness to work during pandemic flu. J Nurs Manag 2011 Jan;19(1):98-108. [doi: 10.1111/j.1365-2834.2010.01190.x] [Medline: 21223410]
15. Gollust SE, Fowler EF, Niederdeppe J. Television news coverage of public health issues and implications for public health policy and practice. Annu Rev Public Health 2019 Apr 01;40:167-185. [doi: 10.1146/annurev-publhealth-040218-044017] [Medline: 30633711]
16. Grün B, Hornik K. topicmodels: an R package for fitting topic models. Journal of Statistical Software 2011;40(13):1-30 [FREE Full text] [doi: 10.18637/jss.v040.i13]
17. Kang J, Kim S, Roh S. [A topic modeling analysis for online news article comments on nurses' workplace bullying]. J Korean Acad Nurs 2019 Dec;49(6):736-747. [doi: 10.4040/jkan.2019.49.6.736] [Medline: 31932568]
18. Kim BK, Kim MK. Keywords network analysis of temple stay using big-data. Tourism Studies 2015;39(5):27-40 [FREE Full text] [doi: 10.17086/JTS.2015.39.5.27.40]
19. Yoon YM, Kim SK, Kim HK, Kim EJ, Jeong Y. Comparison of topics related to nurse on the internet portals and social media before and during the COVID-19 era using topic modeling. Journal of Muscle and Joint Health 2020;27(3):255-267 [FREE Full text] [doi: 10.5953/JMJH.2020.27.3.255]
20. Bigkinds. URL: https://www.bigkinds.or.kr/ [accessed 2024-05-09]
21. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. Journal of Machine Learning Research 2003;3:993-1022 [FREE Full text]
22. Steyvers M, Griffiths T. Probabilistic topic models. In: Landauer TK, McNamara DS, Dennis S, Kintsch W, editors. Handbook of Latent Semantic Analysis. New York, NY: Psychology Press; 2007:427-448.
23. Dotsika F, Watkins A. Identifying potentially disruptive trends by means of keyword network analysis. Technological Forecasting and Social Change 2017 Jun;119:114-127. [doi: 10.1016/j.techfore.2017.03.020]
24. Yoon KJ. Infectious diseases and public health care as seen through the response to COVID-19. Health and Welfare Issues & Focus 2020;377:1-11 [FREE Full text]
25. Yoon S. Deriving the effective improvement elements of the emergency room in large general hospitals. Journal of Critical Care Nursing 2015;8(2):66-79 [FREE Full text]
26. Son C, Yoon M, Kim S, Jo Y. Current Status and Challenges of COVID-19 Infections in Seoul City's Nursing Homes. The Seoul Institute. 2021 May 04. URL: https://www.si.re.kr/node/64880 [accessed 2024-12-27]
27. Choi MJ, Joung YS, Lee HY, Jeong HS. Analysis of needs for infection control education in long-term care hospitals for the epidemic of COVID-19: Borich priority formula and the locus for focus model. Health and Social Welfare Review 2021;41(1):8-21 [FREE Full text] [doi: 10.15709/HSWR.2021.41.1.8]
28. Kim Y. Healthcare reform after MERS outbreak: progress to date and next steps. J Korean Med Assoc 2016;59(9):668. [doi: 10.5124/jkma.2016.59.9.668]
29. The 27th Health Insurance Policy Review Committee in 2021. Korean Ministry of Health and Welfare. 2021 Dec 23. URL: https://www.mohw.go.kr/board.es?mid=a20401000000&bid=0032&tag=&act=view&list_no=369057 [accessed 2024-12-28]
30. Kim J, Kwon H, Park E, Choi H, Lee S. Nursing workforce supply and demand: review and prospect in South Korea. Health Economics and Policy Research 2021;27(4):27-49 [FREE Full text]
31. Beggs-Yeager C, Sharts-Hopko N, McDermott-Levy R. The role of nurses in surveillance to enhance global health security: A delphi study. Nurs Outlook 2021;69(6):1021-1029. [doi: 10.1016/j.outlook.2021.05.011] [Medline: 34332765]

32. Kim J, Kim S, Park E, Jeong S, Lee E. Policy issues and new direction for comprehensive nursing service in the National Health Insurance. J Korean Acad Nurs Adm 2017;23(3):312. [doi: 10.11111/jkana.2017.23.3.312]
33. Jun B. Middle East respiratory syndrome outbreak and infectious disease control in Korea. J Korean Med Assoc 2015;58(7):590. [doi: 10.5124/jkma.2015.58.7.590]
34. Park JH, Lee MH. Effects of nursing and care-giving integrated service on nursing work performance, nurses' job satisfaction and patient safety. Journal of Home Health Care Nursing 2017;24(1):14-22 [FREE Full text] [doi: 10.22705/jkashcn.2017.24.1.014]
35. Son J, Park BK, Lee CH, Ahn KH, Kim JN, Park MH, et al. Incidence rate and characteristics of falls in patients using comprehensive nursing care service. Health Policy and Management 2019;29(2):172-183. [doi: 10.4332/KJHPA.2019.29.2.172]
36. Park S. Corona emergency, but only 4 government-supported nurses. Kangwon News. 2020 Nov 27. URL: https://www.kwnews.co.kr/page/view/2020112600000000114 [accessed 2024-05-09]
37. Shin KR. COVID-19 counter measures and action plan for improvement of nursing treatment. HIRA Research 2021;1(1):103-107 [FREE Full text] [doi: 10.52937/hira.21.1.1.103]
38. Choi C. Graduation leads to employment. Nursing department popularity explodes. Mail News. URL: https://www.imaeil.com/page/view/2009092909391865942 [accessed 2024-05-09]
39. Yeun EJ, Han SY, Lee IJ. Subjectivity on nursing images among high school students: an application of Q-methodology. Korean Society for Subjectivity Studies 2010;21:179-194 [FREE Full text]
40. Kim M, Seo H, Lee J. Relationships between meaning of work, organizational commitment, professional self-image, and retention intention among new nurses. Journal of Korean Clinical Nursing Research 2022;28(1):54-65. [doi: 10.22650/JKCNR.2022.28.1.54]
41. Nursing Statistics. Korean Nursing Association. 2022. URL: https://www.koreanurse.or.kr/resources/statistics.php [accessed 2024-05-09]
42. Analysis of trends in hospital nursing staff deployment status and working conditions survey from 2010 to 2019. Reference of Hospital Nurses Association. URL: https://khna.or.kr/home/pds/utilities.php?bo_table=board1&wr_id=8105 [accessed 2024-05-09]
43. 2021 Hospital Nursing Staffing Status Survey. Reference of Hospital Nurses Association. URL: https://khna.or.kr/home/pds/utilities.php?bo_table=board1&wr_id=8103&page=1 [accessed 2024-05-09]
44. Kim WS. New nurses on reality shock and organizational commitment convergence study. Journal of Digital Convergence 2020;18(11):377-386. [doi: 10.14400/JDC.2020.18.11.377]
45. Article 5 of the Korean Nurses' Ethical Guidelines (Provision of Ethical Nursing). Korean Nursing Association. URL: http://www.koreanurse.or.kr/about_KNA/ethics.php [accessed 2024-11-30]
46. Hsin DH, Macer DR. Heroes of SARS: professional roles and ethics of health care workers. J Infect 2004 Oct;49(3):210-215 [FREE Full text] [doi: 10.1016/j.jinf.2004.06.005] [Medline: 15337337]
47. Bruyneel A, Dello S, Dauvergne JE, Kohnen D, Sermeus W. Prevalence and risk factors for burnout, missed nursing care, and intention-to-leave the job among intensive care unit and general ward nurses: A cross-sectional study across six European countries in the COVID-19 era. Intensive Crit Care Nurs 2025 Feb;86:103885. [doi: 10.1016/j.iccn.2024.103885] [Medline: 39522308]
48. Kim JY. Nurses' experience of Middle East Respiratory Syndrome patients care. Journal of the Korea Academia-Industrial cooperation Society 2017;18(10):185-196 [FREE Full text] [doi: 10.5762/KAIS.2017.18.10.185]
49. Jin D, Lee G. Experiences of nurses at a general hospital in Seoul which is temporarily closed due to COVID-19. J Korean Acad Soc Nurs Educ 2020 Nov;26(4):412-422. [doi: 10.5977/jkasne.2020.26.4.412]
50. Bae H, Kim J. Impact of increasing an enrollment quota in nursing schools on nurse staffing level in local hospitals. Korean Data Anal Soc 2017;19(6):3301-3316 [FREE Full text] [doi: 10.37727/jkdas.2017.19.6.3301]
51. Ministry of Health and Welfare announces “Measures to Improve Nurses’ Working Environment and Treatment”. Korean Ministry of Health and Welfare. 2018 Mar 20. URL: https://www.mohw.go.kr/board.es?mid=a10503010100&bid=0027&act=view&list_no=344262&tag=&nPage=1 [accessed 2024-11-30]
52. Cho S, Hong KJ, Lee J. Estimation of revenue growth generated by changing nurse staffing grades and fees and introducing night shift nursing fees. J Korean Acad Nurs Adm 2021;27(3):191. [doi: 10.11111/jkana.2021.27.3.191]
53. Hall LM, Angus J, Peter E, O'Brien-Pallas L, Wynn F, Donner G. Media portrayal of nurses' perspectives and concerns in the SARS crisis in Toronto. J Nurs Scholarsh 2003 Apr 23;35(3):211-216 [FREE Full text] [doi: 10.1111/j.1547-5069.2003.00211.x] [Medline: 14562487]
54. MERS Fight Medical Staff Bullying Growing...Even Children Suffer. Kim IY. 2015 Jun 16. URL: https://www.yna.co.kr/view/AKR20150616167200063?input=1195m [accessed 2024-05-09]
55. Lee J. Nurses collapse trying to fill doctors' absence. The Hankyoreh Newspaper. 2020 Aug 25. URL: https://www.hani.co.kr/arti/society/society_general/959334.html [accessed 2024-05-09]

Abbreviations

EID: emerging infectious disease
LDA: latent Dirichlet allocation
MERS: Middle East respiratory syndrome
OECD: Organisation for Economic Co-operation and Development
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by A Mavragani; submitted 11.05.24; peer-reviewed by H Lee, PH Liao, CY Han, S Pesälä; comments to author 05.11.24; revised version received 07.12.24; accepted 21.12.24; published 06.01.25.

Please cite as:
Kim J, Yun EK
Topic Modeling of Nursing Issues in the Media During 4 Emerging Infectious Disease Epidemics in South Korea: Descriptive Analysis
J Med Internet Res 2025;27:e60446
URL: https://www.jmir.org/2025/1/e60446
doi:10.2196/60446
PMID:

Original Paper

The Willingness of Doctors to Adopt Artificial Intelligence–Driven Clinical Decision Support Systems at Different Hospitals in China: Fuzzy Set Qualitative Comparative Analysis of Survey Data

Zhongguang Yu[1,2*], PhD; Ning Hu[3*], MSc; Qiuyi Zhao[4], MSc; Xiang Hu[5], PhD; Cunbo Jia[6], PhD; Chunyu Zhang[7], PhD; Bing Liu[6], MSc; Yanping Li[1], PhD

[1]Economics and Management School, Wuhan University, Wuhan, China
[2]Respiratory Centre, China-Japan Friendship Hospital, Beijing, China
[3]School of Management, Beijing University of Chinese Medicine, Beijing, China
[4]The Fourth Affiliated Hospital of Soochow University (Suzhou Dushu Lake Hospital), Suzhou, China
[5]Business School, Hubei University, Wuhan, China
[6]Hospital Office, China-Japan Friendship Hospital, Beijing, China
[7]Department of Human Resources, China-Japan Friendship Hospital, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Yanping Li, PhD
Economics and Management School
Wuhan University
299 Bayi Road, Wuchang District
Wuhan, 430072
China
Phone: 86 68753084
Fax: 86 68754150
Email: ypli@whu.edu.cn

Abstract

Background: Artificial intelligence–driven clinical decision support systems (AI-CDSSs) are pivotal tools for doctors to improve diagnostic and treatment processes, as well as improve the efficiency and quality of health care services. However, not all doctors trust artificial intelligence (AI) technology, and many remain skeptical and unwilling to adopt these systems.

Objective: This study aimed to explore in depth the factors influencing doctors' willingness to adopt AI-CDSSs and assess the causal relationships among these factors to gain a better understanding for promoting the clinical application and widespread implementation of these systems.

Methods: Based on the unified theory of acceptance and use of technology (UTAUT) and the technology-organization-environment (TOE) framework, we have proposed and designed a framework for doctors' willingness to adopt AI-CDSSs. We conducted a nationwide questionnaire survey in China and performed fuzzy set qualitative comparative analysis to explore the willingness of doctors to adopt AI-CDSSs in different types of medical institutions and assess the factors influencing their willingness.

Results: The survey was administered to doctors working in tertiary hospitals and primary/secondary hospitals across China. We received 450 valid responses out of 578 questionnaires distributed, indicating a robust response rate of 77.9%. Our analysis of the influencing factors and adoption pathways revealed that doctors in tertiary hospitals exhibited 6 distinct pathways for AI-CDSS adoption, which were centered on technology-driven pathways, individual-driven pathways, and technology-individual dual-driven pathways. Doctors in primary/secondary hospitals demonstrated 3 adoption pathways, which were centered on technology-individual and organization-individual dual-driven pathways. There were commonalities in the factors influencing adoption across different medical institutions, such as the positive perception of AI technology's utility and individual readiness to try new technologies. There were also variations in the influence of facilitating conditions among doctors at different medical institutions, especially primary/secondary hospitals.

Conclusions: From the perspective of the 6 pathways for doctors at tertiary hospitals and the 3 pathways for doctors at primary/secondary hospitals, performance expectancy and personal innovativeness were 2 indispensable and core conditions in the pathways to achieving favorable willingness to adopt AI-CDSSs.

(J Med Internet Res 2025;27:e62768) doi:10.2196/62768

KEYWORDS

artificial intelligence; clinical decision support systems; willingness; technology adoption; fuzzy set qualitative comparative analysis; fsQCA; pathways

Introduction

Background

Artificial intelligence–driven clinical decision support systems (AI-CDSSs), aimed at enhancing medical decision-making, have been in use since the 1980s [1]. These systems leverage clinical knowledge, patient information, and health data to improve health care services by allowing intelligent detection, management, and improvement of patient health conditions [2,3]. According to statistics, the total number of outpatient visits to medical and health institutions in China reached 5.5 billion in 2023, with doctors at public hospitals averaging 7.1 outpatient visits and 2.3 hospital visits per day and with the burden on primary health care institutional doctors being even greater [4]. According to "World Health Statistics 2023: Monitoring Health for the SDGs, sustainable development goals" released by the World Health Organization (WHO), China has a low doctor density and uneven distribution of health care resources [5]. To relieve the pressure on doctors and improve the efficiency of medical services, China has actively explored the application of AI-CDSSs, with the National Health Commission of China issuing a national policy on the promotion of clinical decision support systems (CDSSs) in 2023.

With the evolution of artificial intelligence (AI) technology, current AI-CDSSs can generate specific assessments and recommendations through logical reasoning, analyze patient data from electronic medical records, and offer decision support to health care providers [6]. AI-CDSSs are widely used in clinical diagnoses, drug treatments, preventive measures, and patient management, and are recognized for their roles in enhancing health care decision-making and doctor performance [7]. For instance, scholars, such as Hanson et al [8], have conducted research from the perspective of patient management, evaluating the application of AI-CDSSs in the nursing field, and have confirmed that certain systems are able to improve the accuracy and comprehensiveness of nursing treatment. Alsharqi et al [9], among others, have evaluated the application effects of AI-CDSSs in the automatic image selection field of echocardiography. They found that AI-CDSSs can effectively identify, distinguish, and explain images through machine learning models. Islam et al [10] studied how AI-CDSSs could help patients continuously observe different parameters for controlling insulin levels, automatically analyzing the personal data of diabetic patients. The application of AI-CDSSs in the field of anesthesiology is also extensive. According to existing research, AI-CDSSs can improve the preoperative use of antibiotics and beta-blockers, reduce the use of inhaled anesthetics, and assist in completing anesthesia records and billing work [11-13].

AI-CDSSs are primarily targeted at physicians, whose acceptance is pivotal for their implementation. Sambasivan et al [14] extended the unified theory of acceptance and use of technology (UTAUT) model and explored the factors influencing doctors' willingness to adopt new CDSSs in a developing country context, considering user involvement in decision-making, perceived threat to autonomy, effort expectancy, and performance expectancy. Their study involved doctors from 12 hospitals representing 10 different specialty areas. However, physicians' attitudes toward these systems varied, with some exhibiting favorable inclinations and others exhibiting reluctance, primarily attributed to apprehensions about technological readiness, data confidentiality, absence of personalized interaction, and skepticism toward emerging technologies.

Thus, although some doctors have expressed positive attitudes toward adopting AI-CDSSs, others still harbor reservations, citing concerns over technology immaturity, data privacy and security, absence of human touch, and a distrust of new technologies. For instance, Wagner et al [15] found that only 54.9% of family doctors were willing to use AI for medical diagnosis, while O'Leary et al [16] reported that 82% of health care professionals saw the usefulness of AI in diagnosing rare diseases. In their 2020 study, Park et al [17] discovered that over 75% of American radiology students believed AI would play a crucial role in future medicine, while almost half of them expressed decreased enthusiasm for radiology due to AI adoption. Scheetz et al [18] found that 71% of the health care professionals surveyed in Australia and New Zealand believed AI would improve medicine, while 85.7% felt it would impact health care manpower. In a South Korean study, only 5.9% of doctors were very familiar with AI [19]. Similarly, in a UK study, medical students indicated they were not adequately prepared for AI. Scholars have also expressed concerns about humans being replaced by AI in health care [20].

Thus, despite the potential benefits, skepticism persists among scholars and doctors regarding the integration of AI in health care, citing concerns about technology replacing human roles. For instance, Poon and Sung [21] highlighted doctors' skepticism toward AI technology in clinical practice that impeded the progress of AI applications owing to a lack of trust.

Investigating doctors' willingness to adopt AI-CDSSs and understanding the factors influencing their acceptance can have a significant impact on the comprehensive integration of AI-CDSSs into clinical applications. Sambasivan et al [14] employed structural equation modeling to explore doctors' willingness to use AI-CDSSs in developing countries, revealing that concerns about potential threats to professional autonomy could dampen doctors' willingness to embrace these systems. Conversely, active involvement in the planning, design, and implementation of AI-CDSSs was associated with increased acceptance and readiness to use these technologies. Similarly, Laka et al [22] utilized logistic regression analysis to explore the adoption willingness of AI-CDSSs, finding that doctors in local primary care facilities, as opposed to those in larger hospitals, identified factors, such as time constraints, perceived threats to professional autonomy, and considerations of patient preferences, as significant barriers to adopting AI-CDSSs.

Current research has focused mainly on examining the impact of individual factors or certain factors affecting the allocation of medical resources. Thus, the literature lacks in-depth analysis of the interactive mechanisms and synergistic effects of factors influencing doctors' intentions to adopt AI-CDSSs. There is also a lack of comprehensive analysis that combines the multiple factors influencing doctors' intentions to adopt AI-CDSSs for assessing their causal relationships. To explore the driving mechanisms of multiple conditional linkages on doctors' adoption willingness, our study focuses on doctors at different medical institutions in China, not only providing important assistance for the application of AI-CDSSs in hospitals at all levels in China, but also offering experiential references for other developing countries in the use and adoption of AI-CDSSs. Using fuzzy set qualitative comparative analysis (fsQCA), we clarify the synergistic effects of multiple factors influencing doctors' intentions to adopt AI-CDSSs, thereby providing theoretical support for promoting their application in the medical field.

Framework for AI-CDSS Adoption Willingness

The integration of the UTAUT was proposed in 2003 by Venkatesh et al [23] to explain the relevant factors influencing an individual's willingness to accept or use new technology. The UTAUT consists of 4 key factors: performance expectancy, effort expectancy, social influence, and facilitating conditions, which influence behavior through willingness [24]. The UTAUT also considers the moderating effects of sex, age, experience, and voluntary use. The theory was established in the context of the organizational implementation of new technology, with the influencing factors having clear utilitarian characteristics. With the emergence of AI technology, an increasing number of scholars are using the UTAUT to study individual AI technology adoption issues. The technology-organization-environment (TOE) framework proposed by Tornatzky et al [25] suggests that technological, organizational, and external environmental factors also have a certain impact on an organization's adoption and implementation of new technology.

There is an interactive relationship between medical institutions and doctors. An organization's attitude toward innovation will affect employees' acceptance of new technology [26]. In addition, the adoption willingness and behavior of AI technology in an organization may be affected by the adoption willingness of employees [27]. Therefore, based on the UTAUT model and TOE framework, this study constructed a multi-layer dynamic impact model for AI technology adoption among physicians.

Technical Factors Influencing the Adoption of AI-CDSSs by Doctors

Performance expectancy is one of the key constructs in the UTAUT model used to explain and predict individual technology acceptance behavior [28]. In a hospital setting, it can capture the extent that doctors believe that using new technology will help improve their job performance [29]. Previous studies have shown that performance expectancy is crucial for doctors' adoption and acceptance of AI-CDSSs [30], similar to the perceived usefulness in the technology acceptance model (TAM). Compared with other technology application scenarios, doctors place more emphasis on the impact of technology on their job performance when adopting new technology. Currently, AI technology has the ability to assist in eliminating redundant work steps, providing decision support, and improving job performance [31,32]. However, there are also issues, such as communication barriers between doctors and AI technology, which can impact work efficiency. The effectiveness of AI technology in the workplace is yet to be widely validated [33]. Therefore, performance expectancy still plays an important role in doctors' willingness to adopt AI technology.

Perceived risk refers to the degree of insecurity that doctors perceive when they are using technology to execute tasks and exchange data [34]. AI technology requires big data to achieve powerful learning, which means that AI technology may involve the input of data from various parties, such as individuals and vendors. When there is a risk of information leakage, doctors using AI technology may face legal, moral, and ethical issues. This can have a significant impact on their willingness to adopt AI technology.

Organizational Factors Influencing the Adoption of AI-CDSSs by Doctors

Social influence refers to the influence doctors feel from their social environment regarding a specific behavior. It is a key factor in the UTAUT model that affects an individual's willingness to adopt new technology [30]. In an organizational context, doctors are frequently influenced by colleagues and leaders, and they enhance their sense of belonging by conforming to these groups. When faced with emerging technologies, such as AI, there may not be enough information for an informed decision, making doctors more susceptible to peer influence. However, leaders also influence doctors' willingness to adopt AI technology [35] and have the power to authorize subordinates, determine job promotions, provide rewards, and administer punishments. Thus, doctors align with leaders to receive recognition.

The UTAUT construct "facilitating conditions" refers to the extent that doctors perceive that the necessary infrastructure and resources in the organization support their use of new technology [30]. Thus, this would also influence a doctor's willingness to adopt AI technology. The promotion of new technology requires organizations to provide various resources such as knowledge, funds, and technology. Simpler and more convenient external support conditions are associated with a greater likelihood of doctors adopting AI technology. Research has shown that facilitating conditions positively influence an individual's willingness to adopt AI technology [36,37].

Individual Factors Influencing the Adoption of AI-CDSSs by Doctors

Technology anxiety refers to an individual's emotional anxiety or fear of the performance of new technology. For example, when individuals believe that the technology may threaten their sense of self, they may experience technology anxiety, which reduces their willingness to adopt it [28]. The existing and potential capabilities of AI technology to replace human abilities are constantly increasing, causing individuals to experience stronger feelings of anxiety compared with other technologies. Therefore, our study included technology anxiety as a factor in our framework.

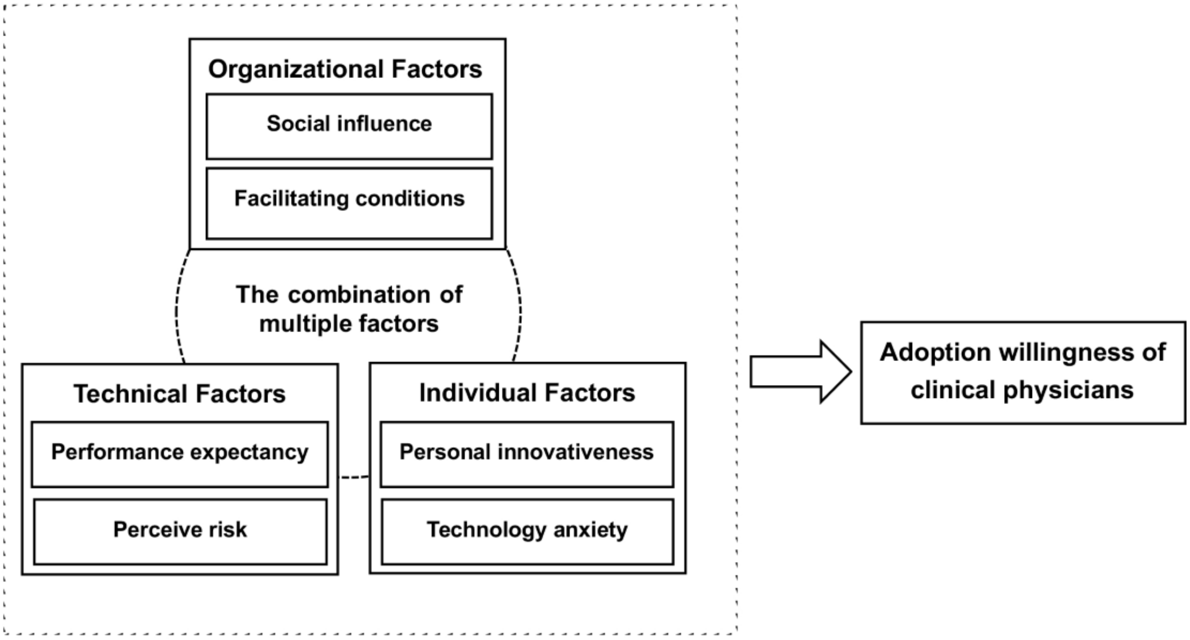
Personal innovativeness reflects an individual's willingness to try something new. Innovation diffusion theory suggests that owing to differences in innovation capabilities, individuals' willingness and behaviors vary in this respect. Some scholars have proven that in consumer scenarios, personal innovativeness positively influences individuals' willingness to adopt self-service technologies [38]. As AI technology is a revolutionary innovation, personal innovativeness is needed to drive doctors toward a greater willingness to adopt it. Therefore, we assume that personal innovativeness significantly impacts doctors' willingness to adopt AI technology.

Study Model

Our study applies the UTAUT model for the basis of our research framework to analyze the factors influencing doctors' adoption of AI-CDSSs, adding other factors as well. Our model considers 3 perspectives: technology, organizations, and individuals. The technical factors included are performance expectancy and perceived risk, the organizational factors are social influence and facilitating conditions, and the individual factors are technology anxiety and personal innovativeness. We have incorporated performance expectancy, social influence, and facilitating conditions from the UTAUT model, and based on the research by Chen et al [37], personal innovativeness and perceived risk have been introduced into the model and confirmed, along with key factors in the UTAUT model that significantly influence doctors' acceptance of AI technology. Huang et al [39] demonstrated how technology anxiety reflected an individual's emotional anxiety or fear regarding the performance of AI technology. When individuals perceived the technology as threatening their sense of self, they experienced technology anxiety, which reduced their willingness to adopt the technology. Therefore, we have included technology anxiety in the model. Considering that AI technology often does not require users to learn how to operate it, as it possesses anthropomorphic characteristics that are different from nonintelligent technologies [8], we did not include effort expectancy. The factors ultimately in our analysis framework are performance expectancy, perceived risk, facilitating conditions, social influence, technology anxiety, and personal innovativeness.

Thus, we assume that doctors' willingness to adopt AI technology will be influenced by these various factors. Although some studies have investigated the individual effects of these elements on doctors' AI adoption willingness, providing a foundation for our understanding of the factors influencing the willingness to adopt AI technology, research struggles to answer how these factors interact to influence the willingness to adopt AI technology under multiple situational conditions. Additionally, research has not identified the deep-rooted causal relationships that are affecting doctors' willingness to adopt AI technology. To fill this gap in the literature, based on the UTAUT model and incorporating the TOE framework, we have explored the complex causal mechanisms of how environmental, technological, and individual factors influence doctors' willingness to adopt AI technology from a configurational perspective, proposing a theoretical model (Figure 1).

Figure 1. Analysis framework of the factors influencing Chinese doctors' willingness to adopt artificial intelligence–driven clinical decision support systems.



Methods

Method Selection

We used fsQCA to explore the complex causal mechanisms influencing clinical doctors' willingness to adopt AI-CDSSs, primarily for the following reasons. First, using this method uncovers the nonlinear relationships between various influencing factors and the doctors' willingness to adopt AI-CDSSs. The fsQCA method also explores combinations of influencing factors instead of individual factors [40]. Second, as our research question was "Which factors can lead clinical doctors to have a higher willingness to adopt AI-CDSSs?", the use of this method can reveal multiple equivalent paths that influence doctors' willingness to adopt AI-CDSSs. Third, compared with other qualitative comparative analysis methods, fsQCA is more suitable for handling continuous variables. Fourth, fsQCA can be applied to different sample sizes ranging from very small (<50 cases) to very large (thousands of cases) [41].

Data Collection

An online questionnaire was developed and shared with doctors in China who met specific criteria related to their work experience. We distributed 578 questionnaires through the Wenjuanxing platform [42]. After eliminating any invalid submissions, we had 450 valid responses, resulting in a questionnaire response rate of 77.9%. Detailed demographic information on the participants is presented in Table 1.

According to "Hospital Classification Management Measures" issued by the National Health Department, hospitals in China are classified into 3 levels. Tertiary hospitals provide medical and health services across regions, provinces, and cities, and nationwide. Secondary hospitals provide comprehensive medical and health services to multiple communities and undertake teaching and research tasks in regional hospitals. Primary hospitals are grassroot hospitals and health centers that provide preventive, medical, health, and rehabilitation services to their local communities [43]. In this study, we divided hospitals into only 2 categories for simplicity: tertiary hospitals and primary/secondary hospitals. Among the respondents, there were 332 responses from clinical doctors working in tertiary medical institutions and 118 responses from clinical doctors in primary/secondary medical institutions, as mentioned above.

Table 1. Basic respondent demographic statistics.

Category	Tertiary hospitals (N=332), n (%)	Primary/secondary hospitals (N=118), n (%)
Sex		
Male	171 (51.5)	53 (44.9)
Female	161 (48.5)	65 (55.1)
Age (years)		
Under 25	27 (8.1)	12 (10.2)
25 to 34	109 (32.8)	31 (26.3)
35 to 44	142 (42.8)	35 (29.7)
45 to 54	51 (15.4)	34 (28.8)
Above 54	3 (0.9)	6 (5.1)
Education		
Bachelor’s degree	89 (26.8)	85 (72.0)
Master’s degree	176 (53.0)	13 (11.0)
Doctoral degree	66 (19.9)	1 (0.9)
Others	1 (0.3)	19 (16.1)
Major title		
Resident physician	100 (30.1)	46 (39.0)
Attending physician	113 (34.0)	39 (33.1)
Associate chief physician	81 (24.4)	19 (16.1)
Chief physician	38 (11.5)	14 (11.9)
Duration of employment (years)		
1 or less	28 (8.4)	15 (12.7)
2 to 5	86 (25.9)	25 (21.2)
6 to 10	71 (21.4)	17 (14.4)
11 to 15	66 (19.9)	23 (19.5)
16 to 20	41 (12.4)	5 (4.2)
21 to 25	19 (5.7)	11 (9.3)
25 or more	21 (6.3)	22 (18.6)

Variable Measurement and Calibration

To ensure the reliability and validity of our scales, we based them on mature scales developed by scholars in the field of technology adoption and appropriately adjusted them for our research question to create measurement items for our main variables [23]. The main variables were performance expectancy, perceived risk, facilitating conditions, social influence, technology anxiety, personal innovativeness, and adoption willingness. Participant responses were on a 5-point Likert scale assessing the extent that they agreed with the content described in the items. We calculated the average score of the corresponding items within each scale to measure the variable. We use SPSS (version 25.0) to analyze the reliability and validity of the scales. All variables had a Cronbach α greater than .8, composite reliability greater than 0.8, Kaiser-Meyer-Olkin (KMO) values greater than 0.7, and average variance extracted values greater than 0.5, indicating that the scales had good reliability and validity. To ensure the overall reliability of our results, we removed items with factor loadings less than 0.6, and the conditions were still met after these items were deleted (Table 2).

To meet the Boolean logic requirements for qualitative comparative analysis, variables need to be transformed into sets and cases need to be assigned to the sets before conducting fsQCA, a process known as data calibration. In this process, we need to establish calibration points for “full membership,” “crossing point,” and “full nonmembership.” We adopted a scholar’s calibration method for the Likert scale questionnaire data, coding “completely agrees (5)” as “full membership,” “neutral (3)” as “crossing point,” and “completely disagrees (1)” as “full nonmembership.” By setting these 3 thresholds, we converted the original data into fuzzy scores ranging from 0 to 1, using the calibrate (x, n1, n2, n3) function in the fsQCA software.

Table 2. Reliability and validity analysis.

Variable and item	Factor loading	Cronbach α	KMO[a]	CR[b]	AVE[c]
Performance expectancy		.831	0.797	0.916	0.611
1	0.913				
2	0.925				
3	0.884				
4	0.543				
Perceived risk		.879	0.848	0.934	0.568
1	0.567				
2	0.674				
3	0.885				
4	0.875				
5	0.838				
Facilitating conditions		.857	0.821	0.929	0.640
1	0.775				
2	0.893				
3	0.768				
4	0.825				
Social influence		.883	0.811	0.940	0.864
1	0.842				
2	0.913				
3	0.810				
4	0.686				
Technology anxiety		.921	0.900	0.890	0.572
1	0.839				
2	0.833				
3	0.859				
4	0.620				
5	0.723				
Personal innovativeness		.865	0.831	0.896	0.650
1	0.786				
2	0.824				
3	0.912				
4	0.880				
Adoption willingness		.887	0.747	0.931	0.768
1	0.961				
2	0.953				
3	0.946				

[a] KMO: Kaiser-Meyer-Olkin.

[b] CR: composite reliability.

[c] AVE: average variance extracted.

Ethical Considerations

This study was approved by the Clinical Research Ethics Committee of China-Japan Friendship Hospital (number: 2024-KY-254). All participants provided informed consent before the investigation began. Furthermore, information on the research participants was kept confidential, and personal private information was not disclosed.

Results

Necessary Condition

According to the fsQCA method, before conducting the configuration analysis, the first step is to perform a necessity analysis on the individual condition variables, with the results reflected through consistency and coverage. Consistency represents the degree that the condition variables are a subset of the outcome variables. Identifying a necessary condition generally requires a consistency score higher than 0.9. Coverage represents the extent that the condition variables explain the outcome. This is only meaningful for conditions that pass the consistency test, with no acceptable threshold. The results of the necessity test for the individual conditions are shown in Table 3.

In Table 3, we can see that the consistency of personal innovativeness in influencing doctors' willingness to adopt AI-CDSSs at primary/secondary hospitals was higher than 0.9 and coverage was as high as 0.933. This indicates that personal innovativeness is a necessary condition influencing doctors' willingness to adopt AI-CDSSs in these hospitals. The consistency of the other variables was less than 0.9, indicating that they are not sufficient to constitute the necessary conditions affecting doctors' willingness to adopt AI-CDSSs, with these variables having relatively weak independent explanatory powers. Thus, we need to further analyze the combining effects of these condition variables and their impacts on our outcome variable.

Table 3. Necessary condition analysis.

Condition variable	Tertiary hospitals		Primary/secondary hospitals	
	Consistency	Coverage	Consistency	Coverage
Performance expectancy	0.882	0.917	0.872	0.832
~[a]Performance expectancy	0.376	0.852	0.423	0.784
Perceived risk	0.471	0.952	0.570	0.890
~Perceived risk	0.775	0.853	0.722	0.763
Facilitating conditions	0.653	0.955	0.757	0.924
~Facilitating conditions	0.611	0.849	0.567	0.739
Social influence	0.787	0.944	0.854	0.934
~Social influence	0.487	0.855	0.479	0.711
Technology anxiety	0.707	0.939	0.661	0.908
~Technology anxiety	0.555	0.854	0.658	0.766
Personal innovativeness	0.861[b]	0.949	0.930[b]	0.933
~Personal innovativeness	0.414	0.834	0.414	0.702

[a] "~" indicates the negation of the condition.

[b] Consistency exceeds 0.8.

Adequacy Analysis of Configuration

As mentioned, we conducted fsQCA separately for tertiary and primary/secondary hospitals. According to the principles of fsQCA, we included 6 condition variables. We retained 85% of the case numbers to set the frequency threshold, with case number thresholds of 5 and 2 for tertiary hospitals and primary/secondary hospitals, respectively. The consistency threshold for each configuration was higher than 0.8, and the proportional reduction in inconsistency threshold was greater than 0.75. A configuration with consistency below the threshold was assigned a value of 0 (Tables 4 and 5).

After standard analysis of the improved truth table, we found 3 types of solutions: complex, intermediate, and parsimonious. Among them, we obtained the intermediate solution through a counterfactual analysis, assuming that the emergence of personal innovativeness may increase doctors' willingness to adopt AI-CDSSs, whereas the other individual conditions may contribute to doctors' willingness to adopt AI-CDSSs. We identified the core conditions for each configuration by comparing the nested relationships between the intermediate and parsimonious solutions. The conditions appearing in both the parsimonious and intermediate solutions were considered core conditions for that configuration, whereas those appearing only in the intermediate solution were considered marginal conditions (Table 6).

In Table 6, we can see that there are 3 pathways leading to positive doctors' willingness to adopt AI-CDSSs in tertiary hospitals, which have been presented below.

Table 4. Truth table for doctors at tertiary hospitals.

Conditional variable						Number	Outcome	Raw consistency	PRI[a] consistency	SYM[b] consistency
A	B	C	D	E	F		Y			
1	1	1	1	1	1	32	1	0.997789	0.994404	0.994404
1	1	0	1	1	1	7	1	0.997023	0.989236	0.989237
1	0	1	1	1	1	38	1	0.995619	0.988805	0.988805
1	1	1	1	0	1	5	1	0.996540	0.984514	0.987172
1	0	0	1	1	1	22	1	0.994604	0.984282	0.984282
1	0	1	1	0	1	44	1	0.992934	0.981120	0.981120
1	0	0	0	1	1	19	1	0.992382	0.976033	0.978558
1	0	0	1	0	1	10	1	0.990879	0.962754	0.963750
1	0	0	0	1	0	8	1	0.988872	0.942290	0.942291
0	0	1	1	0	1	5	1	0.988860	0.929057	0.932779
1	0	0	0	0	1	13	1	0.982544	0.921498	0.931327
1	0	0	1	0	0	8	1	0.984218	0.907492	0.907493
0	0	0	0	1	1	8	1	0.981872	0.904679	0.909656
0	0	0	1	0	1	6	1	0.985961	0.902044	0.902046
1	0	1	1	0	0	10	1	0.971584	0.850015	0.866264
0	0	0	0	0	1	10	1	0.973338	0.826399	0.828025
1	0	0	0	0	0	11	1	0.964521	0.796661	0.797960
0	0	0	0	0	0	24	0	0.895842	0.463338	0.467589

[a] PRI: proportional reduction in inconsistency.

[b] SYM: symmetric.

Table 5. Truth table for doctors at primary/secondary hospitals.

Conditional variable						Number	Outcome	Raw consistency	PRI[a] consistency	SYM[b] consistency
A	B	C	D	E	F		Y			
1	0	1	1	1	1	7	1	0.995976	0.986577	0.986577
1	1	1	1	1	1	15	1	0.991665	0.977517	0.977518
1	0	0	1	1	1	2	1	0.992201	0.963961	0.963961
1	1	0	1	1	1	3	1	0.988830	0.953054	0.953054
1	1	0	1	0	1	3	1	0.992045	0.952049	0.952050
1	0	1	1	0	1	17	1	0.982767	0.947533	0.953350
0	0	1	1	0	1	2	1	0.989427	0.927492	0.927492
1	1	1	1	0	1	4	1	0.984086	0.922251	0.948591
1	0	0	1	0	1	3	1	0.986501	0.920241	0.920242
1	0	1	0	0	1	3	1	0.975536	0.830054	0.830053
1	0	0	0	0	1	5	1	0.969439	0.797216	0.797216
0	0	0	1	0	1	2	0	0.968835	0.732559	0.732558
0	0	0	0	0	1	6	0	0.946550	0.654310	0.654310
1	0	1	1	0	0	4	0	0.965158	0.631557	0.631558
0	0	0	0	1	0	2	0	0.912641	0.349602	0.349602
0	0	0	0	0	0	16	0	0.842963	0.315754	0.327492
1	1	0	0	0	0	3	0	0.937445	0.308522	0.308522
0	1	0	0	1	0	2	0	0.937238	0.290340	0.348758
1	0	0	0	0	0	10	0	0.861165	0.251138	0.251138

[a] PRI: proportional reduction in inconsistency.

[b] SYM: symmetric.

Table 6. Adoption willingness of doctors in different medical institutions.

Conditions	Tertiary hospitals[a]						Primary/secondary hospitals[b]		
	Pathway S1		Pathway S2		Pathway S3		Pathway N1	Pathway N2	
	S1a	S1b	S2a	S2b	S3a	S3b		N2a	N2b
Performance expectancy	PC[c]	PC	—[d]	—	PC	PC	PC	PC	—
Perceived risk	AM[e]	AM	AM	AM	—	—	—	AM	AM
Facilitating conditions	AM	—	AM	—	PM[f]	—	—	—	PM
Social influence	AM	PM	AM	PM	PM	PM	PM		PM
Technology anxiety	—	AM	—	AM	—	PM	—	AM	AM
Personal innovativeness	—	—	PC	PC	PC	PC	PC	PC	PC
Consistency	0.963	0.970	0.964	0.985	0.992	0.994	0.972	0.961	0.983
Raw coverage	0.416	0.451	0.413	0.444	0.591	0.549	0.769	0.514	0.479
Unique coverage	0.018	0.015	0.017	0.011	0.022	0.040	0.279	0.023	0.017

[a] The solution consistency value was 0.804 and solution coverage value was 0.953.

[b] The solution consistency value was 0.810 and solution coverage value was 0.961.

[c] PC: presence of a core causal condition.

[d] The condition can or cannot exist in the configuration.

[e] AM: absence of a marginal causal condition.

[f] PM: presence of a marginal causal condition.

Configuration for Doctors at Tertiary Hospitals

Technology Driven

The core condition for both pathways S1a and S1b was performance expectancy, which played a dominant role in the pathways. The findings indicate that these doctors believe that AI-CDSSs are helpful in clinical work, can improve work efficiency, and can enhance work quality. Pathway S1a indicated that under high performance expectancy, even without perceived technical risks for AI-CDSSs and without organizational factors, such as convenience of AI use and social influence, doctors have favorable adoption willingness for AI-CDSSs. Pathway S1b indicated that under high performance expectancy, without technology anxiety and perceived risk but with social influence, such as influence from surrounding groups, doctors still have favorable adoption willingness for AI-CDSSs.

Individual Driven

The core condition for pathways S2a and S2b was personal innovativeness, which played a primary role in these pathways. The findings indicate that these doctors are willing to try new AI technology, as they typically favor innovativeness that enables continuous learning of new medical technologies and treatment methods. Pathway S2a showed that when doctors reflect strong personal innovativeness, without perceived risks, convenience factors, and social influence, they have positive adoption willingness for AI-CDSSs. Pathway S2b indicated that when doctors have personal innovativeness without perceived risks and technology anxiety for AI-CDSSs, but with a certain degree of social influence, they still tend to have positive adoption willingness for AI-CDSSs.

Technology-Individual Dual Driven

The core conditions for pathways S3a and S3b were performance expectancy and personal innovativeness, indicating that doctors with both high performance expectancy and high personal innovativeness develop strong adoption willingness. Pathway S3a demonstrated that doctors with high performance expectancy and high personal innovativeness need support from certain convenience factors as well as a certain degree of social influence to engender strong adoption willingness for AI-CDSSs. Pathway S3b showed that when doctors have high performance expectancy, high personal innovativeness, and some degree of technology anxiety and social influence, they still tend to have high adoption willingness for AI-CDSSs.

Configuration for Doctors at Primary and Secondary Hospitals

Technology-Individual Dual Driven

The core conditions for pathways N1 and N2a were performance expectancy and personal innovativeness, indicating that doctors with both high performance expectancy and high personal innovativeness have strong adoption willingness for AI-CDSSs. Pathway N1 showed that doctors with high performance expectancy and high personal innovativeness have strong adoption willingness when influenced by leaders, colleagues, and other people regarding the use of AI. Pathway N2a demonstrated that doctors with high performance expectancy, high personal innovativeness, and no perceived risks or technology anxiety for AI technology have strong adoption willingness.

Organization-Individual Dual Driven

The core conditions for pathway N2b were convenience factors and personal innovativeness, indicating that doctors who receive support from their workplace and from technology and have high personal innovativeness develop strong adoption willingness. In other words, with high convenience factors, personal innovativeness, and low social influence, even without perceived risks and technology anxiety, doctors have strong adoption willingness for AI-CDSSs.

Comparative Analysis

Comparing doctors at tertiary hospitals with those at primary/secondary hospitals, we can point out the similarities and differences. The pathways reflect what drives the doctors at these different medical institutions to stronger adoption willingness for AI-CDSSs.

Similarities

AI technology and personal factors play dominant roles in influencing the adoption of AI-CDSSs by doctors at all the hospitals analyzed. There are not only single-dimensional factors that affect doctors' adoption willingness, but also combinations of factors, as in the technology-individual dual model, that have an impact on adoption. At the technological level, doctors believed that the application of AI-CDSSs in clinical diagnosis and treatment processes can provide efficient diagnostic support and improve the quality of clinical services. Although issues, such as overdiagnosis, may exist with AI-CDSSs, doctors considered the overall technology of AI-CDSSs to be safe and reliable. At the individual level, doctors demonstrated strong acceptance and openness to AI technology, showing no anxiety regarding the emergence of new technologies. In particular, they appeared willing to try new AI technologies.

Differences

Convenience factors had a greater impact on the adoption willingness of doctors at primary/secondary hospitals than on the adoption willingness of doctors at tertiary hospitals. According to pathways S1a and S2a, a lack of convenience factors did not affect the strong adoption willingness of doctors at tertiary hospitals. This finding indicates that even when the marginal condition of the convenience factor is missing, these doctors still have a positive adoption willingness for AI-CDSSs. In contrast, looking at pathway N2b, convenience factors are a necessary condition for doctors at primary/secondary hospitals, and these doctors will only favor adoption of AI-CDSSs when convenience factors are present. Thus, convenience factors are the objective material factors influencing the willingness of these doctors to adopt AI-CDSSs. As tertiary hospitals are generally regional hospitals that have comprehensive hospital facilities and advanced information systems, such convenience factors are present; thus, they do not significantly influence the adoption of AI-CDSSs among doctors at these facilities. Unlike tertiary hospitals, primary/secondary hospitals are often county hospitals or primary health care institutions, with some located in remote rural areas and having fewer hardware and software resources to support their doctors. Therefore, the impact of convenience factors on the adoption willingness of doctors at these medical institutions is greater. The survey results imply that only when there is sufficient external support for doctors at these primary and secondary institutions will they actively adopt AI-CDSSs.

Robustness Test

To test the robustness of our results, we adjusted the consistency threshold from 0.8 to 0.85 and 0.72 [44]. There were no substantial changes observed in the configuration of the pathways or parameters. The results indicated that the adjusted structure remained consistent with the original structure and the pathways were the same as those before the adjustment. Therefore, the results remained robust.

Discussion

Principal Findings

We constructed a theoretical framework based on the UTAUT and TOE framework using configurational thinking and fsQCA to configure 6 conditional elements. We explored the multiple concurrent factors and causal complex mechanisms that influence the willingness to adopt AI-CDSSs among clinical doctors at different medical institutions in China from technological, organizational, and individual perspectives. Our results provide a theoretical basis for the further integration of AI-CDSSs into clinical applications. The following are our key results.

We found that the paths driving high AI-CDSS adoption willingness among clinical doctors in tertiary hospitals fell into 6 categories, which were summarized in 3 configurations: technology driven, individual driven, and technology-individual dual driven. The paths driving high AI-CDSS adoption willingness among clinical doctors at primary/secondary medical institutions fell into 3 categories, which were summarized in 2 configurations: technology-individual and organization-individual dual driven.

Comparing tertiary hospitals with primary/secondary medical institutions, we observed some commonalities and some differences in the paths driving doctors to a positive willingness to adopt AI-CDSSs. In terms of commonalities, AI technology and individual factors play dominant roles in doctors’ adoption willingness. The doctors indicated their beliefs that AI-CDSSs can provide efficient diagnostic support and improve the quality of medical services. Moreover, they indicated that they are willing to try new technologies. In terms of differences, convenience factors had a greater impact on doctors at primary/secondary medical institutions. These doctors would actively adopt AI-CDSSs only with sufficient external support.

By studying doctors from different levels of medical institutions in China and their adoption paths of AI-CDSSs, we found that resource availability may be an important factor influencing the adoption willingness of doctors at different medical institutions. For example, tertiary hospitals have greater access to resources, such as funding and technical support, compared with primary/secondary hospitals. Additionally, differences in organizational culture and management styles and values may impact doctors’ attitudes toward AI-CDSSs. The characteristics of patient populations served by different levels of medical institutions may also influence doctors’ perceptions of AI-CDSSs, as patient needs and complexities can vary across hospital settings.

Finally, our research results can serve as a reference for other developing countries for promoting the application of AI-CDSSs or AI technologies in clinical treatment, contributing to enhancing the medical service capabilities of medical institutions.

Theoretical Contributions

Our study contributes to the literature in several ways. First, we built a comprehensive analytical framework for doctors’ willingness to adopt AI technology. This is based on the UTAUT and TOE framework combined with specific characteristics and application scenarios for AI technology, with a focus on technical, organizational, and individual factors. Previous studies have focused mainly on the causal relationships between individual variables and the willingness to adopt AI technology. However, our study introduces multiple factors that can influence doctors’ willingness to adopt AI technology, namely, performance expectancy, perceived risks, convenience factors, social influence, technology anxiety, and personal innovativeness. Building on the UTAUT model, we incorporated additional factors, specifically personal innovativeness, technology anxiety, and perceived risks, into the analytical framework. As such, we were able to enrich the theoretical research regarding the factors influencing doctors’ willingness to adopt AI technology.

Second, we explored the synergistic effects influencing doctors’ willingness to adopt AI-CDSSs from a configuration perspective, expanding the application of the UTAUT model in explaining causal complexity. Although the UTAUT model is widely used to explain individual adoption of new technologies within organizations, existing studies on AI adoption have largely overlooked the complexity of causal relationships. Owing to limitations in research methods, existing technology adoption models have been unable to test and explain the impact of multiple conditions on doctors’ willingness to adopt AI-CDSSs. In our study, we empirically investigated the synergistic effects of 6 specific factors related to technology, organization, and individual aspects on doctors’ willingness to adopt AI-CDSSs from a configuration perspective. By addressing the aforementioned issues, we expand the application of the UTAUT model in explaining causal complexity.

Finally, we used the fsQCA method to analyze the configurations of doctors’ willingness to adopt AI-CDSSs at different medical institutions. Our results indicate that across different medical institutions, performance expectancy and personal innovativeness are the 2 important conditions for doctors to engender strong adoption willingness for AI-CDSSs, whereas perceived risks hinder adoption. Social influence can either promote or hinder doctors’ willingness to adopt AI-CDSSs, and convenience factors have a greater impact on doctors’ adoption willingness at primary/secondary medical institutions. In summary, our research extends the literature on doctors’ willingness to adopt AI-CDSSs and provides theoretical support for future practical applications.

Practical Implications

From the perspective of the 6 configurations among the doctors at tertiary hospitals and the 3 configurations among the doctors at primary/secondary medical institutions, performance expectancy and personal innovativeness were the 2 indispensable and core conditions in the pathways to achieving strong willingness to adopt AI-CDSSs. Thus, AI product providers and health care managers should look closely at these factors when designing and implementing such systems. The organizational factor of facilitating conditions for doctors at primary/secondary medical institutions also appeared to be a necessary condition influencing the adoption willingness of doctors at these institutions. As such, we recommend the following measures for the AI-CDSS process.

Importance of Performance Expectancy in Adoption

For AI product providers, improvements in the quality and applicability of AI products can be achieved by focusing on the designing of AI-CDSSs that meet the specific needs of different clinical doctors. This means involving these doctors in the design and development process of AI-CDSSs to ensure that these systems address practical needs. In addition, to ensure that AI-CDSSs comply with data privacy and security standards, security measures should be included to increase doctors' trust in the systems. The systems should also be evaluated regularly for effectiveness and impact, and should be continuously improved based on feedback to ensure they are meeting user needs and expectations.

Importance of Personal Innovativeness

Health care institution managers should provide the appropriate training and support for doctors before introducing AI-CDSSs. This can be done through various activities, such as videos and practical exercises, among others. Such training can help doctors become comfortable with using AI-CDSSs and increase their effectiveness, boosting doctors' confidence in the systems and their willingness to adopt the systems. In addition, doctors should be actively encouraged to be innovative, cultivate innovation awareness, and improve their acceptance of new technologies.

Importance of Addressing Technology Anxiety

Individual doctors need to actively participate in training to understand the basic functions and uses of AI-CDSSs. Some doctors may be skeptical of AI-CDSSs, fearing that these systems will replace their work or reduce work quality. Appropriate training can help change this negative mindset, encouraging doctors to recognize that these systems are meant to assist and enhance the efficiency and accuracy of clinical work, and not to replace doctors.

Limitations

This study has some limitations that can serve as starting points for future research. First, our research scope was limited to medical institutions in mainland China, thereby lacking comparisons with doctors in foreign medical institutions regarding the adoption paths of AI-CDSSs. Second, differences between medical institutions at different levels, such as funding, staffing, and equipment configuration, may influence doctors' perceptions, acceptance, and experience with AI-CDSSs, leading to variations in questionnaire responses. In addition, we used fsQCA to explore the driving mechanism underlying doctors' adoption willingness based on the interactive matching of multiple conditions, which does not verify the impact of individual variables or a few factors on adoption willingness. To address the limitations of the fsQCA method, future research could consider including the use of structural equation modeling (SEM) to design a more complex model structure by considering the relationships between latent variables and the correlations between multiple observed indicators, thus providing a more comprehensive data analysis and explanation. Finally, the sample size of the questionnaire in primary/secondary hospitals was less than that in tertiary hospitals, but the hospitals included in this study were representative. Because AI has more value in areas with weak medical resources, our future research will focus on the willingness of primary hospitals to use AI. We will refine research plans, expand the sample size of the survey, and improve the compliance of the hospitals included.

Conclusion

Our study built a comprehensive analytical framework for doctors' willingness to adopt AI-CDSSs, considering specific characteristics and application scenarios for AI technology, with a focus on technical, organizational, and individual factors. We used fsQCA to explore doctors' willingness to adopt AI-CDSSs in different types of medical institutions in China along with the factors influencing their willingness. From the perspectives of the 6 pathways of the doctors at tertiary hospitals and the 3 pathways of the doctors at primary/secondary hospitals, performance expectancy and personal innovativeness were 2 indispensable and core conditions in the pathways to achieving favorable willingness to adopt AI-CDSSs. The comparative analysis revealed both similarities and differences between doctors at tertiary hospitals and those at primary/secondary hospitals in terms of their adoption of AI-CDSSs. While technical and individual factors were found to be influential in driving adoption willingness across all hospitals, the impact of facilitating conditions differed between the different levels of medical institutions. Facilitating conditions were identified as a significant driver for adoption among doctors at primary/secondary hospitals, underscoring the importance of external support and resources in facilitating the adoption of AI-CDSSs in these settings.

In conclusion, the results of our research provide valuable insights into the factors influencing doctors' willingness to adopt AI-CDSSs in different health care settings. By addressing performance expectancy, personal innovativeness, and organizational support, health care organizations can promote a more favorable environment for the implementation and utilization of AI technologies, ultimately enhancing clinical decision-making systems and improving patient care outcomes. Continued research and implementation of these strategies can further advance the integration of AI-CDSSs in health care and pave the way for the widespread application of AI technology in clinical practice.

By exploring the positive practice and policy promotion of AI-CDSS application in China, our research can provide a positive reference for the governments of developing countries with similar conditions and uneven distribution of medical resources. At the technical level, the effectiveness and safety of AI technology should be ensured, so that AI can meet the needs of clinical practice. At the organizational level, medical institutions should organize technical training, so that doctors can understand and learn the combination of AI technology and clinical practice, and build a new model of human-machine collaboration. For areas with relatively weak medical resources, the government and medical institutions should increase the infrastructure construction required for AI applications and provide adequate technical support to help doctors solve technical problems.

Acknowledgments

We would like to thank the interviewees who participated in the study. This work was supported by the National Natural Science Fund for Young Scholars of China (72104255) and Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (2021-I2M-1-046).

Data Availability

The raw data supporting the study results are available from the authors upon reasonable request.

Authors' Contributions

ZY, NH, and YL played significant roles in study design, recruitment, data coding, and paper writing. QZ was responsible for investigation, data collection, and editing. XH and CJ were responsible for methodology and writing (review and editing). CZ and BL were responsible for methodology, writing (review and editing), and supervision.

Conflicts of Interest

None declared.

References

1. 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]
2. Osheroff J, Teich J, Levick D, Saldana L, Velasco F, Sittig D, et al. Improving Outcomes with Clinical Decision Support: An Implementer's Guide, Second Edition. New York, NY: HIMSS Publishing; 2012.
3. Shankar P, Anderson N. Advances in sharing multi-sourced health data on decision support science 2016-2017. Yearb Med Inform 2018 Aug 29;27(1):16-24 [FREE Full text] [doi: 10.1055/s-0038-1641215] [Medline: 30157504]
4. China's health industry development Statistical Bulletin for 2023. National Health Commission of the People's Republic of China. 2023. URL: http://www.nhc.gov.cn/guihuaxxs/s3585u/202408/6c037610b3a54f6c8535c515844fae96.shtml [accessed 2024-11-04]
5. World health statistics 2023: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2023:119.
6. Lu Y, Melnick ER, Krumholz HM. Clinical decision support in cardiovascular medicine. BMJ 2022 May 25;377:e059818. [doi: 10.1136/bmj-2020-059818] [Medline: 35613721]
7. Ash J, Sittig D, Guappone A, Dykstra R, Richardson J, Wright A, et al. Recommended practices for computerized clinical decision support and knowledge management in community settings: a qualitative study. BMC Med Inform Decis Mak 2012;12:6 [FREE Full text] [doi: 10.1186/1472-6947-12-6] [Medline: 22333210]
8. Hanson L, Zimmerman S, Song M, Lin F, Rosemond C, Carey T, et al. Effect of the goals of care intervention for advanced dementia: a randomized clinical trial. JAMA Intern Med 2017;177(1):24-31 [FREE Full text] [doi: 10.1001/jamainternmed.2016.7031] [Medline: 27893884]
9. Alsharqi M, Upton R, Mumith A, Leeson P. Artificial intelligence: a new clinical support tool for stress echocardiography. Expert Rev Med Devices 2018;15(8):513-515. [doi: 10.1080/17434440.2018.1497482] [Medline: 29992841]
10. Islam R, Sultana A, Tuhin MN, Saikat MSH, Islam MR. Clinical decision support system for diabetic patients by predicting type 2 diabetes using machine learning algorithms. J Healthc Eng 2023 May 30:6992441 [FREE Full text] [doi: 10.1155/2023/6992441] [Medline: 37287539]
11. Nair BG, Peterson GN, Newman S, Wu W, Kolios-Morris V, Schwid HA. Improving documentation of a beta-blocker quality measure through an anesthesia information management system and real-time notification of documentation errors. Jt Comm J Qual Patient Saf 2012;38(6):283-288. [doi: 10.1016/s1553-7250(12)38036-7] [Medline: 22737780]
12. Nair BG, Peterson GN, Neradilek MB, Newman SF, Huang EY, Schwid HA. Reducing wastage of inhalation anesthetics using real-time decision support to notify of excessive fresh gas flow. Anesthesiology 2013 Apr;118(4):874-884 [FREE Full text] [doi: 10.1097/ALN.0b013e3182829de0] [Medline: 23442753]

13. Freundlich RE, Barnet CS, Mathis MR, Shanks AM, Tremper KK, Kheterpal S. A randomized trial of automated electronic alerts demonstrating improved reimbursable anesthesia time documentation. J Clin Anesth 2013 Mar;25(2):110-114. [doi: 10.1016/j.jclinane.2012.06.020] [Medline: 23333782]
14. Sambasivan M, Esmaeilzadeh P, Kumar N, Nezakati H. Intention to adopt clinical decision support systems in a developing country: effect of physician's perceived professional autonomy, involvement and belief: a cross-sectional study. BMC Med Inform Decis Mak 2012 Dec 05;12:142. [doi: 10.1186/1472-6947-12-142] [Medline: 23216866]
15. Wagner G, Raymond L, Paré G. Understanding Prospective Physicians' Intention to Use Artificial Intelligence in Their Future Medical Practice: Configurational Analysis. JMIR Med Educ 2023 Mar 22;9:e45631. [doi: 10.2196/45631] [Medline: 36947121]
16. O'Leary P, Carroll N, Richardson I. The Practitioner's Perspective on Clinical Pathway Support Systems. 2015 Presented at: 2014 IEEE International Conference on Healthcare Informatics; 15-17 September 2014; Verona, Italy. [doi: 10.1109/ICHI.2014.33]
17. Park CJ, Yi PH, Siegel EL. Medical student perspectives on the impact of artificial intelligence on the practice of medicine. Curr Probl Diagn Radiol 2021;50(5):614-619. [doi: 10.1067/j.cpradiol.2020.06.011] [Medline: 32680632]
18. Scheetz J, Rothschild P, McGuinness M, Hadoux X, Soyer HP, Janda M, et al. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep 2021 Mar 04;11(1):5193. [doi: 10.1038/s41598-021-84698-5] [Medline: 33664367]
19. Oh S, Kim JH, Choi SW, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. J Med Internet Res 2019;21(3):e12422. [doi: 10.2196/12422] [Medline: 30907742]
20. Sit C, Srinivasan R, Amlani A, Muthuswamy K, Azam A, Monzon L, et al. Attitudes and perceptions of UK medical students towards artificial intelligence and radiology: a multicentre survey. Insights into Imaging 2020;11:14. [doi: 10.1186/s13244-019-0830-7] [Medline: 32025951]
21. Poon AIF, Sung JJY. Opening the black box of AI-Medicine. J Gastroenterol Hepatol 2021;36(3):581-584. [doi: 10.1111/jgh.15384] [Medline: 33709609]
22. Laka M, Milazzo A, Merlin T. Factors that impact the adoption of clinical decision support systems (CDSS) for antibiotic management. Int J Environ Res Public Health 2021 Feb 16;18(4):1901. [doi: 10.3390/ijerph18041901] [Medline: 33669353]
23. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
24. Bendary N, Al-Sahouly I. Exploring the extension of unified theory of acceptance and use of technology, UTAUT2, factors effect on perceived usefulness and ease of use on mobile commerce in Egypt. Journal of Business & Retail Management Research 2018 Jan;12(02):60-71. [doi: 10.24052/jbrmr/v12is02/eteoutoaauotufeopuaeouomcie]
25. Tornatzky LG, Klein KJ. Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings. IEEE Transactions on Engineering Management 1982 Feb;EM-29(1):28-45. [doi: 10.1109/tem.1982.6447463]
26. Li YW. Enterprise IT/IS adopt decision behavior model analysis. Modern Management Science 2006(2):29-42 [FREE Full text]
27. Edwards JS, Duan Y, Robins PC. An analysis of expert systems for business decision making at different levels and in different roles. European Journal of Information Systems 2000;9:36-46. [doi: 10.1057/palgrave.ejis.3000344]
28. Menon D, Shilpa K. "Chatting with ChatGPT": Analyzing the factors influencing users' intention to Use the Open AI's ChatGPT using the UTAUT model. Heliyon 2023 Oct 18;9(11):e20962 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e20962] [Medline: 37928033]
29. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157-178. [doi: 10.2307/41410412]
30. Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello CP, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med 2023 Jun 10;6(1):111. [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
31. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons 2018 Nov;62(1):15-25. [doi: 10.1016/j.bushor.2018.08.004]
32. Xu P, Xu XY. The logic and analytical framework of enterprise management change in the era of artificial intelligence. Management World 2020;36:122-129+238 [FREE Full text]
33. Qiu Y, He Q. Research progress on the impact of artificial intelligence on employment and theoretical analysis framework in the Chinese context. Human Resources Development of China 2020(2):90-103. [doi: 10.16471/j.cnki.11-2822/c.2020.2.007]
34. Pillai R, Sivathanu B. Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations. Benchmarking An International Journal 2020 Aug 14;27(9):2599-2629. [doi: 10.1108/bij-04-2020-0186]
35. Chatterjee S, Bhattacharjee KK. Adoption of artificial intelligence in higher education: a quantitative analysis using structural equation modelling. Educ Inf Technol 2020 Mar 31;25(5):3443-3463. [doi: 10.1007/s10639-020-10159-7]
36. Sohn K, Kwon O. Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products. Telematics and Informatics 2020 Apr;47:101324. [doi: 10.1016/j.tele.2019.101324]
37. Chen LSL, Wu KIF. Antecedents of intention to use CUSS system: moderating effects of self-efficacy. Serv Bus 2014;8(4):615-634. [doi: 10.1007/s11628-013-0210-1]

38. Chen J, Li R, Gan M, Fu ZY, Yuan FT. Public acceptance of driverless buses in china: an empirical analysis based on an extended UTAUT model. Discrete Dynamics in Nature and Society 2020;2020:1-13. [doi: 10.1155/2020/4318182]
39. Tsai TH, Lin WY, Chang YS, Chang PC, Lee MY. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. PLoS One 2020 Jan 13;15(1):e0227270. [doi: 10.1371/journal.pone.0227270] [Medline: 31929560]
40. Vis B, Dul J. Analyzing Relationships of Necessity Not Just in Kind But Also in Degree: Complementing fsQCA With NCA. Sociol Methods Res 2018 Nov;47(4):872-899. [doi: 10.1177/0049124115626179] [Medline: 30443090]
41. Pappas IO, Woodside AG. Fuzzy-set qualitative comparative analysis (fsQCA): guidelines for research practice in information systems and marketing. International Journal of Information Management 2021 Jun;58(3):102310. [doi: 10.1016/j.ijinfomgt.2021.102310]
42. Wenjuanxing platform. URL: https://www.wjx.cn/ [accessed 2024-12-16]
43. Du C. An empirical analysis of hierarchical hospital management system in China. Social Science Front 2018(04):84-94.
44. White L, Lockett A, Currie G, Hayton J. Hybrid Context, Management Practices and Organizational Performance: A Configurational Approach. Journal of Management Studies 2020;58(3):718-748. [doi: 10.1111/joms.12609]

Abbreviations

AI: artificial intelligence
AI-CDSS: artificial intelligence–driven clinical decision support system
CDSS: clinical decision support system
fsQCA: fuzzy set qualitative comparative analysis
TOE: technology-organization-environment
UTAUT: unified theory of acceptance and use of technology

Edited by G Tsafnat; submitted 31.05.24; peer-reviewed by J Walsh, M Zahmatkeshan, M Popovic; comments to author 12.10.24; revised version received 04.11.24; accepted 25.11.24; published 07.01.25.

Please cite as:
Yu Z, Hu N, Zhao Q, Hu X, Jia C, Zhang C, Liu B, Li Y
The Willingness of Doctors to Adopt Artificial Intelligence–Driven Clinical Decision Support Systems at Different Hospitals in China: Fuzzy Set Qualitative Comparative Analysis of Survey Data
J Med Internet Res 2025;27:e62768
URL: https://www.jmir.org/2025/1/e62768
doi:10.2196/62768
PMID:

Original Paper

Noninvasive Oral Hyperspectral Imaging–Driven Digital Diagnosis of Heart Failure With Preserved Ejection Fraction: Model Development and Validation Study

Xiaomeng Yang[1,2,3*], MD; Zeyan Li[1,2,3*], MD; Lei Lei[4*], PhD; Xiaoyu Shi[1,2,3*], MD; Dingming Zhang[5], PhD; Fei Zhou[6,7], MD, PhD; Wenjing Li[6,7], MD; Tianyou Xu[1,2,3], MD; Xinyu Liu[1,2,3], MD; Songyun Wang[1,2,3,8], MD, PhD; Quan Yuan[9,10], PhD; Jian Yang[6,7], MD, PhD; Xinyu Wang[8,11], PhD; Yanfei Zhong[4,8], PhD; Lilei Yu[1,2,3,8], MD, PhD

[1] Cardiovascular Hospital, Renmin Hospital of Wuhan University, Wuhan, China
[2] Hubei Key Laboratory of Autonomic Nervous System Modulation, Wuhan University, Wuhan, China
[3] Cardiac Autonomic Nervous System Research Center, Wuhan University, Wuhan, China
[4] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
[5] College of Geomatics, Xi'an University of Science and Technology, Xi'an, China
[6] Department of Cardiology, The First College of Clinical Medical Science, Yichang Central People's Hospital, Yichang, China
[7] Hubei Key Laboratory of Ischemic Cardiovascular Disease, China Three Gorges University, Yichang, China
[8] Medical Remote Sensing Information Cross-Institute, Wuhan University, Wuhan, China
[9] College of Chemistry and Molecular Sciences, Key Laboratory of Biomedical Polymers of Ministry of Education, Wuhan University, Wuhan, China
[10] Institute of Molecular Medicine, Renmin Hospital of Wuhan University, Wuhan, China
[11] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China
[*] these authors contributed equally

Corresponding Author:
Lilei Yu, MD, PhD
Cardiovascular Hospital
Renmin Hospital of Wuhan University
No. 238 Jiefang Road
Wuhan, 430060
China
Phone: 86 02788041911
Email: lileiyu@whu.edu.cn

Abstract

Background: Oral microenvironmental disorders are associated with an increased risk of heart failure with preserved ejection fraction (HFpEF). Hyperspectral imaging (HSI) technology enables the detection of substances that are visually indistinguishable to the human eye, providing a noninvasive approach with extensive applications in medical diagnostics.

Objective: The objective of this study is to develop and validate a digital, noninvasive oral diagnostic model for patients with HFpEF using HSI combined with various machine learning algorithms.

Methods: Between April 2023 and August 2023, a total of 140 patients were recruited from Renmin Hospital of Wuhan University to serve as the training and internal testing groups for this study. Subsequently, from August 2024 to September 2024, an additional 35 patients were enrolled from Three Gorges University and Yichang Central People’s Hospital to constitute the external testing group. After preprocessing to ensure image quality, spectral and textural features were extracted from the images. We extracted 25 spectral bands from each patient image and obtained 8 corresponding texture features to evaluate the performance of 28 machine learning algorithms for their ability to distinguish control participants from participants with HFpEF. The model demonstrating the optimal performance in both internal and external testing groups was selected to construct the HFpEF diagnostic model. Hyperspectral bands significant for identifying participants with HFpEF were identified for further interpretative analysis. The Shapley Additive Explanations (SHAP) model was used to provide analytical insights into feature importance.

Results: Participants were divided into a training group (n=105), internal testing group (n=35), and external testing group (n=35), with consistent baseline characteristics across groups. Among the 28 algorithms tested, the random forest algorithm demonstrated superior performance with an area under the receiver operating characteristic curve (AUC) of 0.884 and an accuracy of 82.9%

in the internal testing group, as well as an AUC of 0.812 and an accuracy of 85.7% in the external testing group. For model interpretation, we used the top 25 features identified by the random forest algorithm. The SHAP analysis revealed discernible distinctions between control participants and participants with HFpEF, thereby validating the diagnostic model's capacity to accurately identify participants with HFpEF.

Conclusions: This noninvasive and efficient model facilitates the identification of individuals with HFpEF, thereby promoting early detection, diagnosis, and treatment. Our research presents a clinically advanced diagnostic framework for HFpEF, validated using independent data sets and demonstrating significant potential to enhance patient care.

Trial Registration: China Clinical Trial Registry ChiCTR2300078855; https://www.chictr.org.cn/showproj.html?proj=207133

(J Med Internet Res 2025;27:e67256) doi:10.2196/67256

KEYWORDS

heart failure with preserved ejection fraction; HFpEF; hyperspectral imaging; HSI; diagnostic model; digital health; Shapley Additive Explanations; SHAP; machine learning; artificial intelligence; AI; cardiovascular disease; predictive modeling; oral health

Introduction

About one-half of patients with chronic heart failure have heart failure with preserved ejection fraction (HFpEF), which has received wide attention in recent years and poses a serious threat to the management of patient health [1,2]. The pathogenesis of HFpEF is complex, and the pathological mechanism is still unclear, so a swift, noninvasive diagnostic strategy is still lacking [3]. Early diagnosis of HFpEF is often difficult because the left ventricular ejection fraction of patients with HFpEF is within the normal range and the symptoms of heart failure are often nonspecific. The diagnosis of HFpEF depends on clinical symptoms, laboratory tests, echocardiography, and an invasive hemodynamics test [4]. However, descriptions of clinical symptoms are subjective, and different clinical symptoms cannot be measured using a unified standard. Echocardiography and hemodynamics tests need to be performed by experienced clinicians in hospitals, thus limiting early identification and timely diagnosis of HFpEF, which introduces obstacles to the health management of individuals with HFpEF [5].

The health of the oral environment has an important impact on cardiovascular health, and oral microenvironmental disorders have been associated with an increased risk of HFpEF [6]. Poor oral health may cause an inflammatory response that is strongly associated with heart failure [7,8]. The tongue, as an important part of the oral cavity, plays an important role in oral health, and the dorsum of the tongue carries the largest number of microbial species, which is an important part of oral health [9]. Inflammation and oxidative stress in the body caused by alterations in the oral microbiome are associated with the risk of developing heart failure [10]. Observing the oral microenvironment of individuals with HFpEF can reflect their lifestyle, dietary habits, and intestinal health [11]. An automated device or system that allows doctors to perform a quick tongue-based diagnosis would be helpful in clinical practice.

Light, which is a single color before dispersion, becomes a pattern of colors arranged in order of wavelength after dispersion; this is called the optical spectrum or optical band [12]. Hyperspectral imaging (HSI) consists of narrower bands, is a technique that captures and analyzes the details of each band in a region, can therefore detect substances that are visually indistinguishable to humans, and has a wide range of applications in many fields [13]. In recent years, many studies have explored the application of HSI technology in medicine [14]. The emergence of handheld HSI cameras, which allow users to quickly capture spectral images, has boosted the application of spectral imaging in medicine [15]. HSI technology has been used in patient information acquisition, medical image analysis, and disease diagnosis [16,17]. HSI is expected to promote the management of patient health. Notably, numerous innovative heart failure prediction models have been developed using common clinical indicators and advanced machine learning techniques [18-23]. These studies primarily focused on the prognosis of heart failure or enhanced existing diagnostic modalities such as echocardiography [24-26]. There exists a significant need for the development of noninvasive, easily accessible diagnostic tools specifically targeting HFpEF, which is a subtype of heart failure characterized by subtle clinical manifestations and complex pathophysiology.

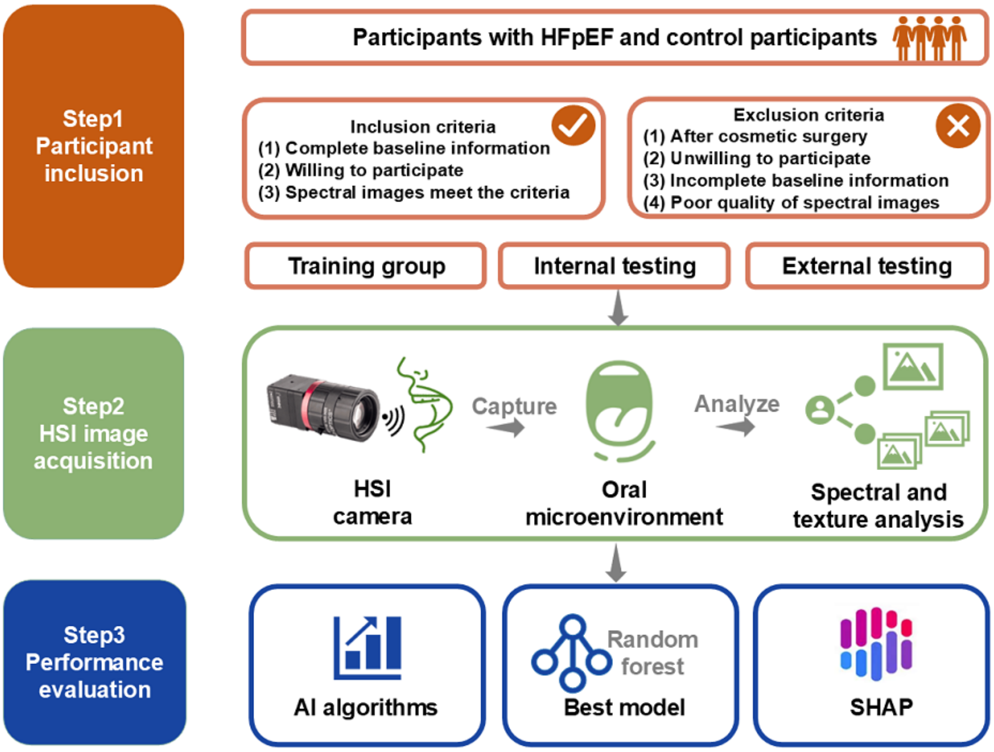
In this study, we collected HSI information of the oral environment, and multiple algorithms were used to select the most characteristic spectral bands of individuals with HFpEF. The best model was selected for internal and external testing, and we used the Shapley Additive Explanations (SHAP) model to additively interpret the best model. The digital HSI HFpEF diagnostic model constructed in this study can help with early detection and management of individuals with HFpEF.

Methods

Study Populations

We included 196 patients who visited Renmin Hospital Wuhan University from April 2023 to August 2023; they comprised the training group and internal testing groups. We included 53 patients who visited Yichang Central People's Hospital from August 2024 to September 2024 as the external testing group (Figure 1). For all participants, HSI and clinical information were collected, and routine tests were conducted. Participants were selected according to the inclusion and exclusion criteria. HFpEF was diagnosed using international standards [4], including (1) typical signs and symptoms of heart failure; (2) left ventricular ejection fraction ≥50%, as assessed by echocardiography performed by a proficient physician; and (3) brain natriuretic peptide >35 ng/L or amino-terminal pro-brain natriuretic peptide >125 ng/L. In addition to these criteria, participants also had to have at least left ventricular hypertrophy or left atrial enlargement with abnormal diastolic function. Individuals were excluded for any of the following reasons: previous cosmetic surgery and facial aesthetics, severe hepatic or renal organ insufficiency, mental or legal incapacity, malignancy-related diseases, other diseases such as psychiatric or neurological disorders, and unable to complete the study activities. Participants were consecutively included in the training and internal testing groups (140 of 196 participants) before randomization using a ratio of 3:1, while participants in the external testing group (35 of 53 participants) were also consecutively included [19]. The manuscript was written in strict accordance with the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [20,21].

Figure 1. Study protocol. AI: artificial intelligence; HFpEF: heart failure with preserved ejection fraction; HSI: hyperspectral imaging; SHAP: Shapley Additive Explanations.



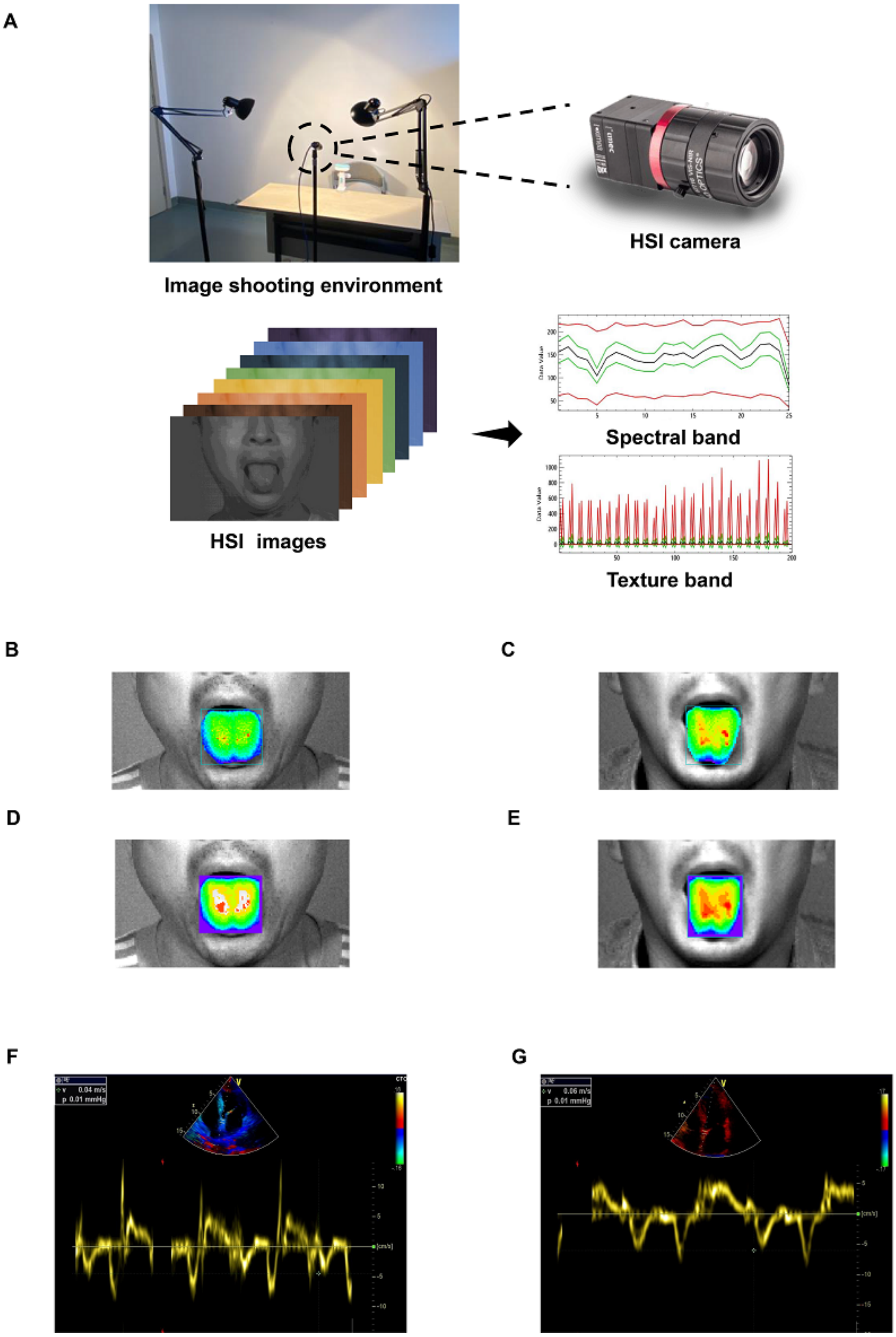
Ethical Considerations

All participants were informed about the study and signed an informed consent document. The study protocol was reviewed and approved by the Ethics Committee of Renmin Hospital of Wuhan University (number WDRM2023-K174) and Yichang Central People's Hospital (number 2024-216-01).

HSI Collection

We prepared a room with good light avoidance conditions and created a dark environment using curtains, blackboards, and other items to avoid external light interference as much as possible to ensure that all participants were in the same light environment for image acquisition. A halogen lamp was chosen as the only light source. The participant's head and face were fixed to ensure that there were no accessories nor hair to obscure facial features. The distance between the HSI camera and the participants was maintained to ensure that the size of the facial features was in the same area of each spectral image and that all participants had the same image size and clarity. We asked participants to expose their tongues for HSI acquisition (Figure 2A). Each participant was provided with a single hyperspectral photomicrograph, and only 1 oral spectral image meeting the criteria was captured.

Figure 2. (A) Hyperspectral imaging (HSI) acquisition environment, (B) processing of a spectral image for a control participant, (C) processing of a spectral image for a participant with heart failure with preserved ejection fraction (HFpEF), (D) processing of a texture image for a control participant, (E) processing of a texture image for a participant with HFpEF, (F) representative HSI and echocardiogram of a control participant, and (G) representative HSI and echocardiogram of a participant with HFpEF.



An HSI camera (MQ0220HG-IM-SM4X4-VIS; XIMEA) captured 25 spectral bands in the 665 nm to 960 nm spectral range. The camera has a pixel resolution of 644 × 484 (black and white) or 640 × 480 (color), pixel size of 7.4 × 7.4 µm, active area size of 4.8 × 3.6 mm, sensor diagonal of 5.9 mm, and dynamic range of 60 dB. The camera has an exposure time of 54 µs to 1 s with a step of 7.56 µs, with a capture time less than 1 s per image (Figure 2A). We checked image quality after image capture to ensure that all images had the same luminance, were in sharp focus, and were of good quality.

HSI Analysis

We preprocessed the obtained images [22,23]. HSI preprocessing methods are of great importance in HSI analysis. Effective preprocessing methods can minimize or even eliminate the influence of extraneous information (eg, sample background, electrical noise, stray light). We first checked the image quality of all oral images, and when poor image quality was found, the participant’s oral images were taken repeatedly in the same environment to ensure that the images could be qualified and used for further analysis. We preprocessed the images using the median filtering method to reduce any noise. HSI was preprocessed using a normalized method to reduce redundant information from the original bands and improve the precision of the HSI [27].

All images were edited using image analysis software (ENVI [Environment for Visualizing Images], version 5.3; NV5 Geospatial Solutions Inc). Backgrounds and clothing were cropped out, and the tongue images were retained. The spectral data of the features were initially extracted from the cropped images to obtain the mean and SD for each spectral photograph. Each spectral image was capable of extracting information for 25 spectral bands. Subsequently, texture analysis was conducted on the image, resulting in the generation of 8 texture features based on the mean and SD values for each band. For each participant, their spectral image could be obtained with 50 original band values and 400 band texture values.

Machine Learning Methods

In order to find the most suitable algorithm for the model, we used as many of the currently known machine learning algorithms as possible, resulting in a total of 28 machine learning algorithms. The single algorithms included linear models, such as linear regression, logistic regression, least absolute shrinkage and selection operator (LASSO) regression, ridge regression, and ElasticNet regression. Other types of single algorithms were also used such as decision tree algorithms, neural network algorithms, and support vector machine (SVM) algorithms. The ensemble models included both a boosting algorithm and bagging algorithm; the boosting algorithms included XGBoost algorithms, and the bagging models included random forest algorithms. For probabilistic models, we used Bayesian algorithms, Bernoulli naïve Bayes algorithms, and Gaussian naïve Bayes algorithms.

We performed 5-fold cross-validation in the training group, which was randomly divided into 5 subsets, each with the same sample size. The model was constructed using 4 subsets, and the model performance was assessed using the remaining data. Next, the model was constructed with another combination of the 4 subsets, and the model performance was assessed using the remaining data. The 5 cycles were performed in sequence, and the results obtained for the model performances were summarized. After constructing the model in the training group, model performance was evaluated in the testing group. We evaluated the importance of features using a random forest variable importance ranking method, where the importance of a feature is typically measured by calculating how often the feature is used in a decision tree: the more often a feature is used in the decision tree or the more it contributes to dividing the data, the higher its importance score. The results of all the decision trees’ assessments of feature importance were summarized to obtain an importance score for each feature in the entire random forest model. Each machine learning model was constructed by incorporating these rankings until the performance of the model could not be improved; the incorporation of features was then stopped, enabling the construction of the model [28].

Evaluation of Model Performance

Based on the previously mentioned diagnostic criteria for HFpEF, we categorized all participants using a bivariate category. During the model construction process, we evaluated the ability of each model to be able to distinguish between participants with HFpEF and control participants. We evaluated the performance of each model by comparing the ability of the model to accurately identify participants with HFpEF. We assessed the performance of all algorithms; the accuracy, sensitivity, specificity, F_1-score, positive predictive value (PPV), and negative predictive value (NPV) of each algorithm were calculated separately. The results are presented in a table. Receiver operating characteristic (ROC) curves of the 5 best performing models for the internal testing and external testing groups were drawn, and the area under the ROC curve (AUC) was calculated. Decision curve analysis (DCA) of the 5 best performing models was implemented to assess the clinical usefulness of the models. Calibration curves of the 5 best performing models were plotted to assess the calibration.

SHAP Model Interpretation

We used the SHAP model to explain the best machine learning algorithm [29]. SHAP is a method of interpreting the output of a machine learning model and assigns weights to the optimal indexes using the Shapley values derived from the analysis; we used it to quantify the contribution of different features to the predicted values [30]. The SHAP value allows visual identification of the impact of different features on the model prediction results. In addition, the SHAP value explains the prediction results for each individual in the training group, helping to understand why the model made a particular prediction. It is also possible to perform an aggregated global interpretation of Shapley values for multiple data points in order to provide a total interpretation of the model and to demonstrate the interconnections between different features. SHAP analysis was implemented using Python software, and the results are presented using visualization methods. We first demonstrated the contribution of the bands selected by the best model and ranking by the contribution of different features. We then ranked the contribution of each individual in the study to the model and showed the SHAP analysis results for 1 participant with HFpEF and 1 control participant using intuitive visualization methods to reveal the contribution of different features.

Statistical Analysis

The data in this study were analyzed using SPSS 23.0 (IBM Corp). Model construction and graphic drawing were completed using R version 3.6.1 (R Foundation for Statistical Computing). We used the Shapiro-Wilk normality test to check the distribution of the data. A 1-way ANOVA was used to compare continuous variables that had a normal distribution, and the results are shown as mean (SD). The Kruskal-Wallis H test was used to compare continuous variables with nonnormal distributions, and the results are shown as median (IQR). The Fisher exact test was used to compare categorical data, and the results are reported as counts and percentages. *P*<.05 was considered statistically significant.

Results

Study Population

Individuals who visited Renmin Hospital of Wuhan University from April 2023 to July 2023 were enrolled. A total of 196 participants were included in this study after excluding individuals who did not meet the inclusion criteria, had incomplete baseline data, did not cooperate with spectral acquisition, or had poor image quality. Data were collected for a total of 140 participants in the training and testing groups (Figure 1). Individuals who visited Yichang Central People’s Hospital from August 2024 to September 2024 were also enrolled. A total of 53 participants were included in this study after excluding individuals who did not meet the inclusion criteria, had incomplete baseline data, did not cooperate with spectral acquisition, or had poor image quality. Data were collected for a total of 35 patients in the external testing groups (Figure 1). The participants were divided into training (n=105), internal testing (n=35), and external testing (n=35) groups. The baseline information, which included basic information, previous medical history, and basic examination and test results, for the training and testing groups was compared (Table 1). There were no significant differences in the baseline data between the training, internal testing, and external testing groups.

Table 1. Participants' baseline characteristics.

Characteristic	Training group (n=105)	Internal testing group (n=35)	External testing group (n=35)	*P* value
Age (years), mean (SD)	62 (11)	64 (11)	59 (15)	.20
Sex, n (%)				.11
Male	64 (61)	28 (80)	24 (68.6)	
Female	41 (39.1)	7 (20)	11 (31.4)	
Current smoker, n (%)				.32
No	39 (37.1)	18 (51.4)	15 (42.9)	
Yes	66 (62.9)	17 (48.6)	20 (57.1)	
Current drinker, n (%)				.94
No	24 (22.9)	8 (22.9)	9 (25.7)	
Yes	81 (77.1)	27 (77.1)	26 (74.3)	
Hypertension				.35
No	53 (50.5)	20 (57.1)	14 (40)	
Yes	52 (49.5)	15 (42.9)	21 (60)	
Diabetes				.61
No	26 (24.8)	9 (25.7)	6 (17.1)	
Yes	79 (75.2)	26 (74.3)	29 (82.9)	
CK-MB[a] (ng/mL), median (IQR)	1.34 (0.59-2.49)	1.62 (0.75-2.34)	1.75 (1.33-2.49)	.22
Cardiac troponin I (ng/mL), median (IQR)	1.45 (0.06-2.90)	1.92 (0.78-4.08)	2.05 (0.31-2.76)	.18
TG[b] (mmol/L), median (IQR)	1.43 (1.10-1.87)	1.57 (1.23-2.01)	1.68 (1.29-2.20)	.14
TC[c] (mmol/L), mean (SD)	3.68 (1.17)	3.59 (1.17)	3.57 (1.04)	.85
HDL-C[d] (mmol/L), median (IQR)	1.30 (1.06-2.00)	1.20 (0.92-1.66)	1.67 (1.25-2.10)	.08
LDL-C[e] (mmol/L), median (IQR)	2.21 (1.46-2.70)	2.11 (1.53-2.34)	2.58 (1.80-3.27)	.052
TSH[f] (μIU/mL), median (IQR)	1.97 (1.32-3.02)	1.78 (1.09-2.71)	1.77 (1.20-2.46)	.34
FT3[g] (pg/mL), median (IQR)	3.25 (2.96-3.58)	3.17 (2.72-3.58)	3.30 (3.09-3.54)	.62
FT4[h] (ng/dL), median (IQR)	1.19 (1.04-1.34)	1.23 (1.07-1.37)	1.19 (1.08-1.29)	.48
BMI (kg/m^2), median (IQR)	24.6 (22.6-26.7)	25.4 (23.7-27.5)	24.7 (22.1-26.8)	.47
NT-proBNP[i] (pg/mL), median (IQR)	95 (60-375)	153 (74-271)	86 (52-114)	.23
HFpEF[j], n (%)				.69
No	74 (70.5)	24 (68.6)	27 (77.1)	
Yes	31 (29.5)	11 (31.4)	8 (22.9)	
LAD[k] (mm), median (IQR)	37.0 (34.0-41.0)	36.0 (31.5-39.0)	36.0 (33.0-39.0)	.38
RAD[l] (mm), median (IQR)	35.0 (33.0-38.0)	34.0 (31.5-36.5)	35.0 (33.5-38.0)	.24
LVEF[m] (%), median (IQR)	55.0 (51.3-60.0)	58.0 (52.0-60.0)	56.0 (51.0-58.0)	.07
LVDD[n] (mm), median (IQR)	49.0 (46.0-52.0)	48.0 (45.5-50.0)	49.0 (45.0-51.5)	.25
E/e′ ratio, median (IQR)	10.7 (8.3-15.3)	11.6 (9.5-16.3)	10.2 (7.9-14.2)	.26

[a] CK-MB: creatine kinase-MB.

[b] TG: triglyceride.

[c] TC: total cholesterol.

[d] HDL-c: high-density lipoprotein cholesterol.

[e] LDL-c: low density lipoprotein cholesterol.

[f] TSH: thyroid-stimulating hormone.

[g] FT3: free triiodothyronine.

[h] FT4: free thyroxine.

[i] NT-proBNP: amino-terminal pro-brain natriuretic peptide.

[j] HFpEF: heart failure with preserved ejection fraction.

[k] LAD: left atrium diameter.

[l] RAD: right atrium diameter.

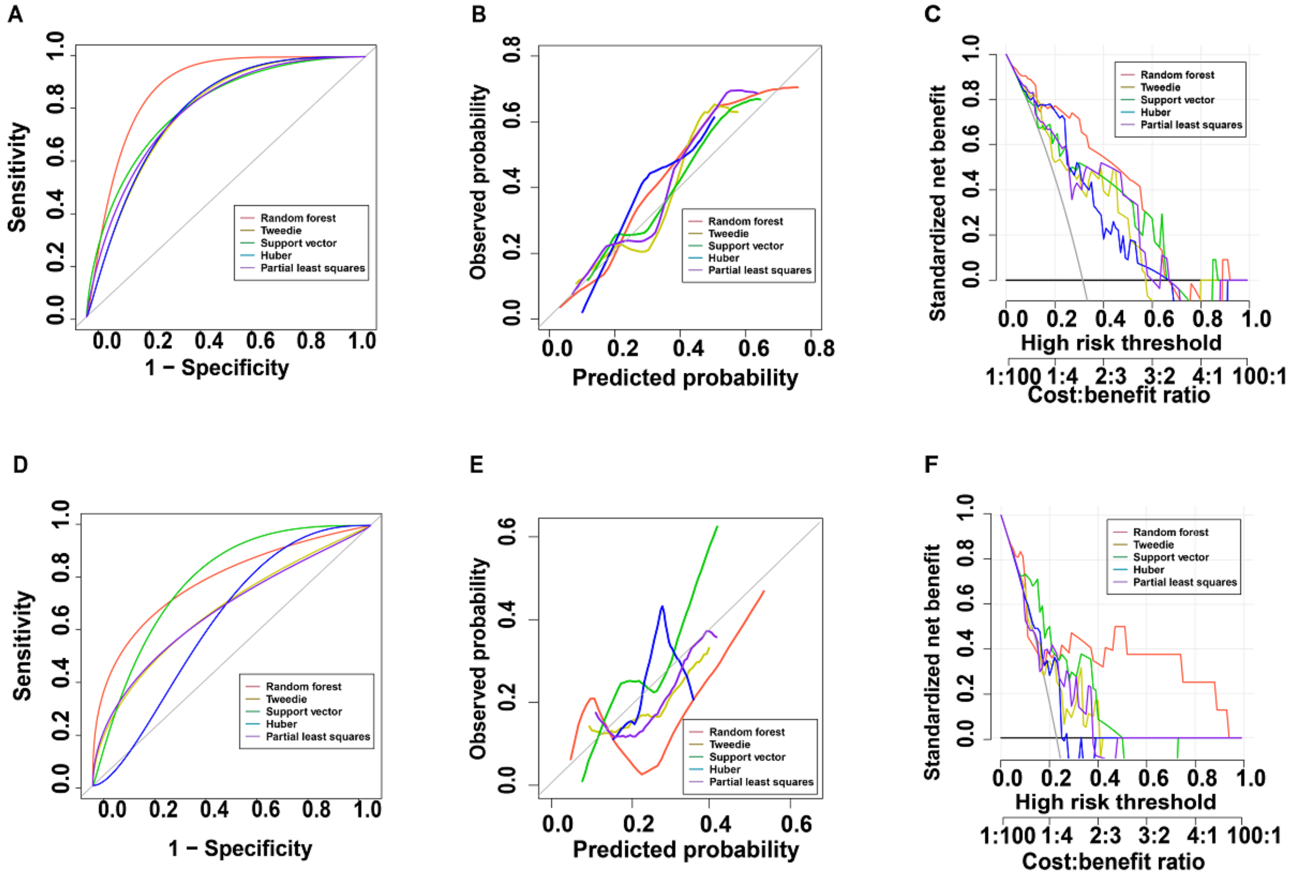
[m] LVEF: left ventricular ejection fraction.

[n] LVDD: left ventricular end diastolic diameter.

Image Processing and Machine Learning Algorithm Comparison

HSI data were collected for the training, internal testing, and external testing groups. Poor image quality for 4 participants in the training group, 1 participant in the internal testing group, and 2 participants in the external testing group made analysis difficult. Therefore, we acquired oral HSI in the same environmental conditions for those participants again and used those images for the next analysis after ensuring that the image quality met the standards. After preprocessing the images, spectral values and texture values of the images were extracted. We present the characteristic tongue HSI and echocardiography images of control participants and participants with HFpEF in Figures 2B-2G. We used 28 algorithms to filter the characteristic spectral bands of images from participants with HFpEF. We used 5-fold cross-validation in the training group, and the model performance was evaluated in the testing group. The accuracy, F_1-score, PPV, NPV, sensitivity, specificity, and AUC of the different algorithms were calculated separately, and the results for the 28 algorithms are listed in Table S1 in Multimedia Appendix 1. Among all the models constructed, the Tweedie, SVM, partial least squares, Huber, and random forest algorithms were the top 5 in terms of performance, showing good ability to identify participants with HFpEF. The ROC curves (Figure 3A), calibration curves (Figure 3B), and DCA (Figure 3C) of the top 5 algorithms were plotted, and the AUCs were calculated (Table S1 in Multimedia Appendix 1). The AUCs in the internal testing group were 0.884 (95% CI 0.769-1.000) for random forest, 0.795 (95% CI 0.633-0.958) for Tweedie, 0.814 (95% CI 0.657-0.972) for SVM, 0.803 (95% CI 0.659-0.947) for Huber, and 0.799 (95% CI 0.640-0.959) for partial least squares.

Figure 3. (A) Receiver operating characteristic (ROC) curve, (B) calibration curve, and (C) decision curve analysis (DCA) for the internal testing group and (D) ROC curve, (E) calibration curve, and (F) DCA for the external testing group.



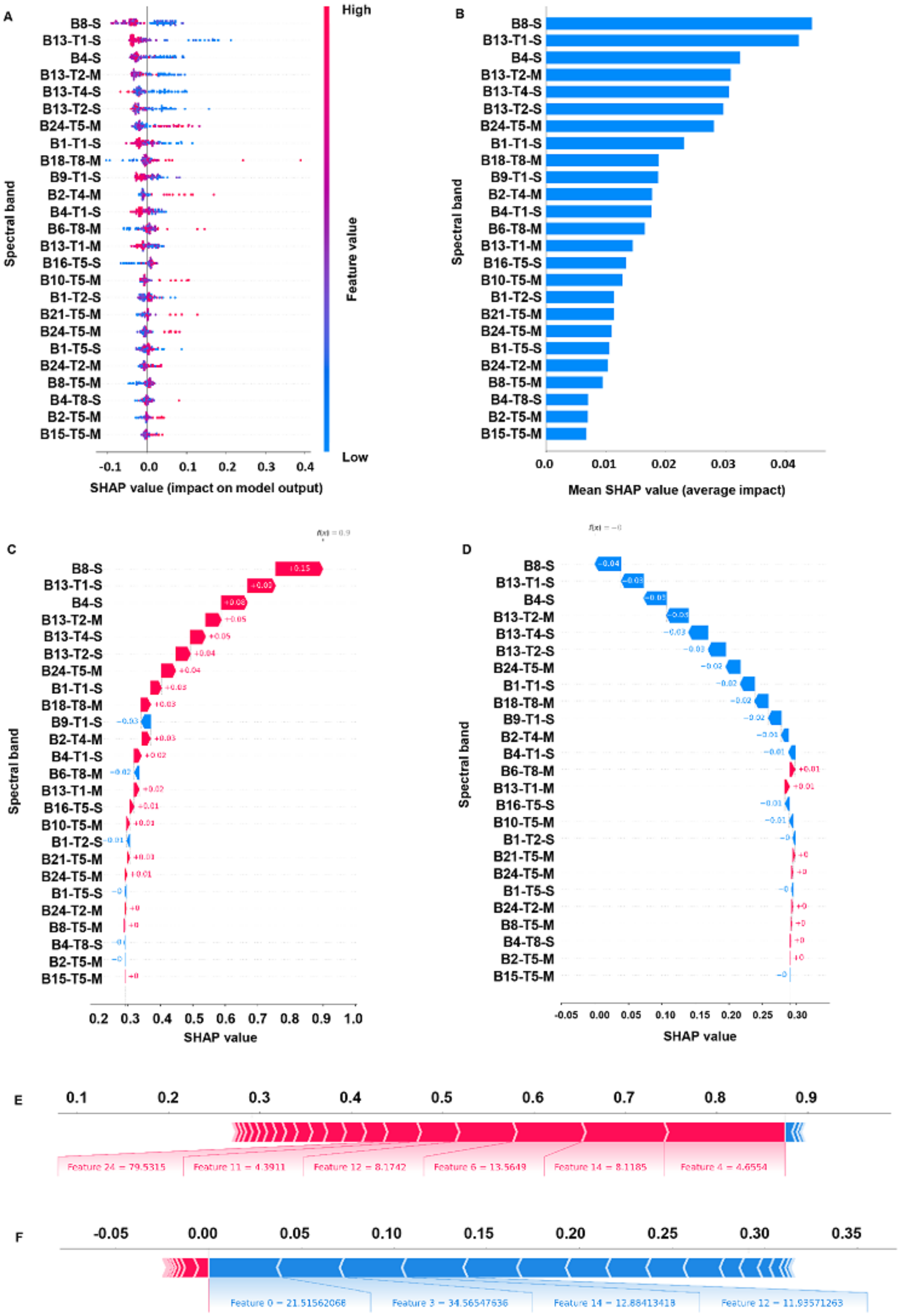
External Validation of the Top 5 Algorithms

The HSI collected in the external testing group were preprocessed in the same way, and the spectral values and texture values were extracted and analyzed. We used the top 5 best performance models in the internal testing group to analyze the external testing data. The accuracy, F_1-score, PPV, NPV, sensitivity, specificity, and AUC of the different algorithms were also calculated separately, and the results for the 5 algorithms are listed in Table S2 in Multimedia Appendix 1. The random forest algorithm still had the best model performance among the 5 algorithms. The ROC curves (Figure 3D), calibration curves (Figure 3E), and DCA (Figure 3F) of the 5 algorithms were plotted, and the AUCs were calculated (Table S2 in Multimedia Appendix 1). The AUCs in the external testing group were 0.812 (95% CI 0.633-0.992) for random forest, 0.676 (95% CI 0.438-0.914) for Tweedie, 0.792 (95% CI 0.632-0.951) for SVM, 0.634 (95% CI 0.428-0.841) for Huber, and 0.671 (95% CI 0.428-0.915) for partial least squares. We extracted the features selected by random forest for further interpretation.

SHAP Interpretation of the Best Algorithms

We used the SHAP model for additive interpretation. The Shapley value was calculated to assign the benefit each characteristic brings to the overall model, showing the contribution of each characteristic to the model's predicted results. We present the characteristics of the top 5 best performing models in order of their contribution to the overall models in Table S3 in Multimedia Appendix 1. The summary plot (Figure 4A) shows the feature importance ranking and distribution via the Shapley value of each spectral band, where the blue bar indicates that the eigenvalue positively affected the model and the red bar indicates that the eigenvalue negatively affected the model. The Shapley value represents the magnitude of each feature's impact on the predicted results, with the point farthest from the centerline indicating a greater influence on the model output. The SHAP feature importance map ranks each characteristic by their contribution (Figure 4B), with features at the top having a greater impact on the model output and those near the bottom having a lesser total impact. SHAP waterfall plots and SHAP bar charts (Figure 4C-4F) were used to visualize the Shapley values of individual samples and their individual results. We visualized the model's ability to recognize HFpEF using a waterfall plot (Figure 4C) and bar chart (Figure 4E) for a control participant and a waterfall plot (Figure 4D) and bar chart (Figure 4F) for a participant with HFpEF. In the waterfall plot, the contribution of each feature is represented by a bar. The length of the bar indicates the magnitude of the feature's influence on the predicted value. A blue bar indicates that the feature increased the predicted value, and a red bar indicates that the feature decreased the predicted value. There was an intuitive difference between the SHAP results for control participants and those for participants with HFpEF. The HFpEF diagnostic model constructed in this study was able to distinguish control participants from participants with HFpEF.

Figure 4. Shapley Additive Explanations (SHAP) of the best algorithm: (A) summary plot, (B) feature importance map, (C) waterfall plot for a control participant, (D) waterfall plot for a participant with heart failure with preserved ejection fraction (HFpEF), (E) bar chart for a control participant, (F) bar chart for a participant with HFpEF.



Discussion

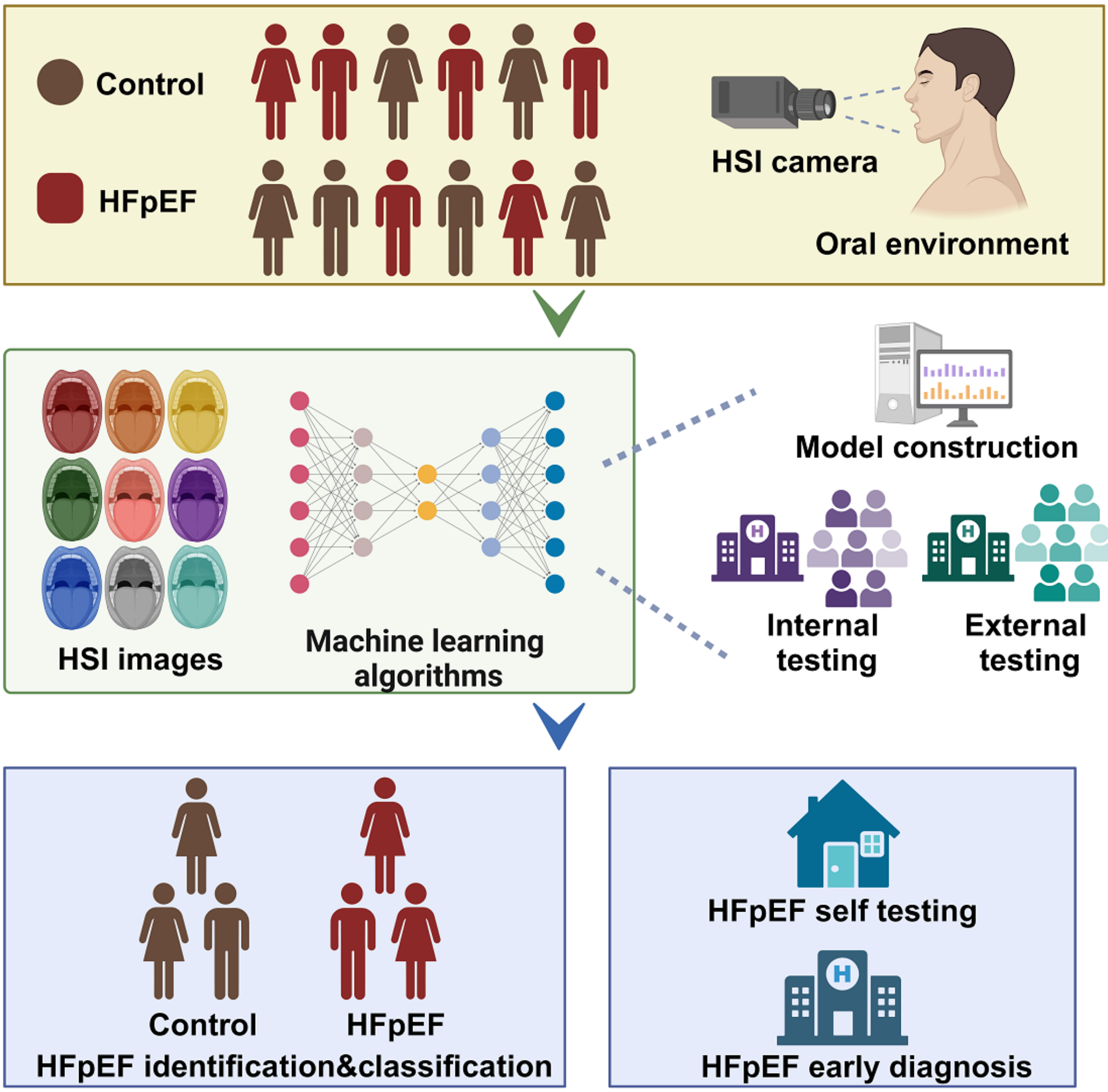
Principal Findings

In this study, we innovatively used HSI to acquire oral images of control participants and participants with HFpEF, extracted the spectral and textural information of the HSI, and characterized the HSI using a variety of machine learning algorithms. The optimal algorithm was selected for the construction of the HFpEF diagnostic model, the model performance was validated in the internal and external testing groups, and the SHAP model was used for additive interpretation.

Digital health care is the future of medicine, introducing considerable convenience to the management of people's health [31]. As artificial intelligence and machine learning continue to advance, enabling automated interpretation and classification of HSI will allow researchers to focus more on deep analysis and decision-making [32]. A large amount of data redundancy is present for HSI data, and traditional image data processing methods struggle to meet processing and analyzing needs; therefore, we generated a series of feature extraction and classification methods for spectral information. The use of multiple artificial intelligence algorithms helps to identify the characteristic bands and textures in large amounts of HSI information [33]. The use of artificial intelligence and machine learning can increase the interpretability of HSI data [34]. To select the most suitable machine learning algorithms for oral spectral images of individuals with HFpEF, we incorporated as many of the current mainstream machine learning models as possible [22]. We analyzed the obtained HSI using 28 machine learning algorithms and filtered the top 5 algorithms for external testing. The random forest algorithm showed good prediction performance in both the internal and external testing groups, so we used the spectral features screened by the random forest algorithm for further analysis. We also performed subgroup analysis based on age and gender in both internal and external testing groups, and the random forest algorithm showed good performance (Figure S1 in Multimedia Appendix 1). Although studies have investigated the use of artificial intelligence for HFpEF diagnosis [19,35,36], these studies primarily focused on constructing echocardiography-based diagnostic models for HFpEF. However, the reliance on specialty physicians to interpret echocardiography exams hinders the early identification of individuals with HFpEF [37]. Moreover, considering the existing diagnostic challenges and limited availability of imaging data during the initial stages of HFpEF, attention should be directed toward noninvasive imaging techniques for diagnosis and management. Therefore, it is imperative to integrate machine learning approaches to develop predictive models for early screening. In our study, we identified spectral bands and textural features that exhibit a close association with the pathophysiological changes observed in HFpEF. This condition is characterized by microvascular dysfunction, inflammation, and oxidative stress leading to alterations in optical properties of oral tissues [3]. HSI effectively captures these subtle changes by detecting variations in light absorption and scattering at different wavelengths. Our model significantly contributes to identifying HFpEF at an early stage, enabling clinicians to promptly initiate appropriate therapies while potentially slowing down disease progression and improving patient health care.

To effectively integrate our diagnostic model into existing health care management for HFpEF, we used an additive interpretation of this digital diagnostic model using the SHAP model. This allowed us to evaluate the contribution of each feature to the prediction, quantify the capability of these features to contribute to the overall model, and demonstrate their impact on the final predictive and diagnostic performance [28]. The SHAP model allows each parameter to be analyzed individually, which can be useful in understanding the decisions made by the model and to improve and rationalize the results of the model [38]. In the macrointerpretation of the SHAP model, the ranking of the contribution of the 25 features screened by random forest to the overall model and the ability of each feature to explain the overall model performance can be clearly seen. In the microinterpretation of the SHAP model, we show the SHAP values for each feature during model evaluation separately for control participants and participants with HFpEF. In the random forest HFpEF diagnostic model, there was a relatively straightforward difference between control participants and participants with HFpEF. The SHAP model demonstrates more intuitively the discriminatory power of the random forest algorithm for participants with HFpEF. Hence, through the provision of intelligent diagnoses, our model effectively bridges the gap between health care professionals and patients, thereby addressing the prevalent issue of frequent misdiagnosis or underdiagnosis and ultimately enhancing the overall quality of care. Furthermore, users can independently access data about oral characteristics at home and transmit the data to health care professionals for expert evaluation through a smartphone app, facilitating continuous monitoring of cardiac health [39]. In addition, in underserved remote areas with limited medical resources, this portable device can be used by mobile medical teams to offer diagnostic screening services to local residents [40], thereby enhancing public health in these regions. Precision medicine and personalized health care represent the future of medical development [41], and our research can also contribute to formulating individualized treatment plans whereby the medical team can gain comprehensive insights into the patient's condition through remotely collected digital data (Figure 5).

Figure 5. Future application scenarios. HFpEF: heart failure with preserved ejection fraction; HSI: hyperspectral imaging.



Limitations

First, in this study, model construction and internal testing were conducted in only 1 center, and external testing was conducted in 1 center. Further improvement of the model for the diagnosis of HFpEF should be conducted in multiple centers. Second, this study used multiple algorithms to select the characteristic hyperspectral bands; however, we did not identify images using deep learning algorithms, which could be used for direct image analysis for HFpEF diagnosis. Third, although we used some means to reduce data redundancy and overfitting, these may still exist. Fourth, different demographic information may affect the results. In future model optimization studies, the applicability of the model in different subgroups and including easily accessible clinical information should be considered. The promotion of HSI technology in HFpEF diagnosis has prospective, broader clinical applications that need to be further explored and developed.

Conclusion

This study demonstrates the innovative use of HSI technology to capture oral images and machine learning algorithms to construct a digital model to diagnose HFpEF. This technology was validated to have excellent performance in both internal and external testing groups. This study offers novel insights into the development of portable devices for rapid identification of HFpEF, thereby facilitating the advancement of digital diagnosis and treatment approaches for HFpEF and ultimately leading to improved patient outcomes and reduced health care costs.

Acknowledgments

We thank all the authors for their joint efforts in the completion of this study and many thanks to the School of Remote Sensing and Information Engineering and State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing of Wuhan University for providing the hyperspectral imaging camera.

This work was supported by grants from the National Ten Thousand Talent Program, National Natural Science Foundation of China (82241057, 82270532); Foundation for Innovative Research Groups of Natural Science Foundation of Hubei Province (2021CFA010); Central University Basic Scientific Research Business Expenses Special Funds (2042024kf0041, 2042022kf1211); Taikang Life Medicine Center PI Program of Wuhan University; and Chutian Talent Program Science and Technology Innovation Team Project, Cross-Innovation Talent Program of Renmin Hospital, Wuhan University (JCRCYG-2022-001).

Data Availability

Individual participant data underlining the results reported in this manuscript after deidentification can be obtained from the corresponding author upon request.

Authors' Contributions

LY, YZ, JY, and XW conceptualized and supervised the study and acquired funding. XY, ZL, LL, XS, FZ, WL, TX, XL, and SW curated the data and performed the formal analysis. XY and ZL wrote the original manuscript draft. LY, YZ, JY, QY, and XW reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional performance evaluations of the algorithms.
[DOCX File , 284 KB - jmir_v27i1e67256_app1.docx]

References

1. Dunlay SM, Roger VL, Redfield MM. Epidemiology of heart failure with preserved ejection fraction. Nat Rev Cardiol 2017 Oct 11;14(10):591-602. [doi: 10.1038/nrcardio.2017.65] [Medline: 28492288]
2. Campbell P, Rutten FH, Lee MM, Hawkins NM, Petrie MC. Heart failure with preserved ejection fraction: everything the clinician needs to know. The Lancet 2024 Mar;403(10431):1083-1092. [doi: 10.1016/s0140-6736(23)02756-3] [Medline: 38367642]
3. Borlaug BA. The pathophysiology of heart failure with preserved ejection fraction. Nat Rev Cardiol 2014 Sep 24;11(9):507-515. [doi: 10.1038/nrcardio.2014.83] [Medline: 24958077]
4. Pieske B, Tschöpe C, de Boer RA, Fraser A, Anker S, Donal E, et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). Eur Heart J 2019 Oct 21;40(40):3297-3317 [FREE Full text] [doi: 10.1093/eurheartj/ehz641] [Medline: 31504452]
5. Guazzi M, Wilhelm M, Halle M, Van Craenenbroeck E, Kemps H, de Boer RA, et al. Exercise testing in heart failure with preserved ejection fraction: an appraisal through diagnosis, pathophysiology and therapy - A clinical consensus statement of the Heart Failure Association and European Association of Preventive Cardiology of the European Society of Cardiology. Eur J Heart Fail 2022 Aug 31;24(8):1327-1345 [FREE Full text] [doi: 10.1002/ejhf.2601] [Medline: 35775383]
6. Molinsky RL, Yuzefpolskaya M, Norby FL, Yu B, Shah AM, Pankow JS, et al. Periodontal status, C-reactive protein, NT-proBNP, and incident heart failure: the ARIC study. JACC Heart Fail 2022 Oct;10(10):731-741 [FREE Full text] [doi: 10.1016/j.jchf.2022.05.008] [Medline: 36175058]
7. Tonelli A, Lumngwena EN, Ntusi NAB. The oral microbiome in the pathophysiology of cardiovascular disease. Nat Rev Cardiol 2023 Jun 09;20(6):386-403. [doi: 10.1038/s41569-022-00825-3] [Medline: 36624275]
8. Carrizales-Sepúlveda EF, Ordaz-Farías A, Vera-Pineda R, Flores-Ramírez R. Periodontal disease, systemic inflammation and the risk of cardiovascular disease. Heart Lung Circ 2018 Nov;27(11):1327-1334. [doi: 10.1016/j.hlc.2018.05.102] [Medline: 29903685]
9. Wilbert SA, Mark Welch JL, Borisy GG. Spatial ecology of the human tongue dorsum microbiome. Cell Rep 2020 Mar 24;30(12):4003-4015.e3 [FREE Full text] [doi: 10.1016/j.celrep.2020.02.097] [Medline: 32209464]
10. Yuzefpolskaya M, Bohn B, Ladanyi A, Khoruts A, Colombo PC, Demmer RT. Oral and gut microbiome alterations in heart failure: Epidemiology, pathogenesis and response to advanced heart failure therapies. J Heart Lung Transplant 2023 Mar;42(3):291-300. [doi: 10.1016/j.healun.2022.12.009] [Medline: 36586790]

11. Beale AL, O'Donnell JA, Nakai ME, Nanayakkara S, Vizi D, Carter K, et al. The gut microbiome of heart failure with preserved ejection fraction. JAHA 2021 Jul 06;10(13):1. [doi: 10.1161/jaha.120.020654]
12. Goetz AFH, Vane G, Solomon JE, Rock BN. Imaging spectrometry for Earth remote sensing. Science 1985 Jun 07;228(4704):1147-1153. [doi: 10.1126/science.228.4704.1147] [Medline: 17735325]
13. Manifold B, Men S, Hu R, Fu D. A versatile deep learning architecture for classification and label-free prediction of hyperspectral images. Nat Mach Intell 2021 Apr 11;3(4):306-315 [FREE Full text] [doi: 10.1038/s42256-021-00309-y] [Medline: 34676358]
14. Shi W, Koo DES, Kitano M, Chiang HJ, Trinh LA, Turcatel G, et al. Pre-processing visualization of hyperspectral fluorescent data with Spectrally Encoded Enhanced Representations. Nat Commun 2020 Feb 05;11(1):726 [FREE Full text] [doi: 10.1038/s41467-020-14486-8] [Medline: 32024828]
15. Cui R, Yu H, Xu T, Xing X, Cao X, Yan K, et al. Deep learning in medical hyperspectral images: a review. Sensors (Basel) 2022 Dec 13;22(24):9790 [FREE Full text] [doi: 10.3390/s22249790] [Medline: 36560157]
16. Lin S, Li Z, Fu B, Chen S, Li X, Wang Y, et al. Feasibility of using deep learning to detect coronary artery disease based on facial photo. Eur Heart J 2020 Dec 07;41(46):4400-4411. [doi: 10.1093/eurheartj/ehaa640] [Medline: 32818267]
17. Zhao Y, Pilvar A, Tank A, Peterson H, Jiang J, Aster JC, et al. Shortwave-infrared meso-patterned imaging enables label-free mapping of tissue water and lipid content. Nat Commun 2020 Oct 23;11(1):5355 [FREE Full text] [doi: 10.1038/s41467-020-19128-7] [Medline: 33097705]
18. Wang J, Wang Y, Duan S, Xu L, Xu Y, Yin W, et al. Multimodal data ‐ driven prognostic model for predicting long ‐ term prognosis in patients with ischemic cardiomyopathy and heart failure with preserved ejection fraction after coronary artery bypass grafting: a multicenter cohort study. JAHA 2024 Nov 27:1. [doi: 10.1161/jaha.124.036970] [Medline: 39604036]
19. Chiou Y, Hung C, Lin S. AI-assisted echocardiographic prescreening of heart failure with preserved ejection fraction on the basis of intrabeat dynamics. JACC Cardiovasc Imaging 2021 Nov;14(11):2091-2104 [FREE Full text] [doi: 10.1016/j.jcmg.2021.05.005] [Medline: 34147456]
20. Klement W, El Emam K. Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies: development and validation. J Med Internet Res 2023 Aug 31;25:e48763 [FREE Full text] [doi: 10.2196/48763] [Medline: 37651179]
21. 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]
22. Leon R, Fabelo H, Ortega S, Cruz-Guerrero IA, Campos-Delgado DU, Szolna A, et al. Hyperspectral imaging benchmark based on machine learning for intraoperative brain tumour detection. NPJ Precis Oncol 2023 Nov 14;7(1):119 [FREE Full text] [doi: 10.1038/s41698-023-00475-9] [Medline: 37964078]
23. Hadoux X, Hui F, Lim JKH, Masters CL, Pébay A, Chevalier S, et al. Non-invasive in vivo hyperspectral imaging of the retina for potential biomarker use in Alzheimer's disease. Nat Commun 2019 Sep 17;10(1):4227 [FREE Full text] [doi: 10.1038/s41467-019-12242-1] [Medline: 31530809]
24. Guo Y, Yu F, Jiang F, Yin S, Jiang M, Li Y, et al. Development and validation of novel interpretable survival prediction models based on drug exposures for severe heart failure during vulnerable period. J Transl Med 2024 Aug 06;22(1):743 [FREE Full text] [doi: 10.1186/s12967-024-05544-6] [Medline: 39107765]
25. Pan X, Wang C, Yu Y, Reljin N, McManus DD, Darling CE, et al. Deep cross-modal feature learning applied to predict acutely decompensated heart failure using in-home collected electrocardiography and transthoracic bioimpedance. Artif Intell Med 2023 Jun;140:102548 [FREE Full text] [doi: 10.1016/j.artmed.2023.102548] [Medline: 37210152]
26. Cen H, Chen S, Feng S, Chen X, Zhu H, Jiang W, et al. Establishing a novel model to assess exercise capacity in chronic heart failure based on stress echocardiography. ESC Heart Fail 2024 Dec;11(6):4335-4347 [FREE Full text] [doi: 10.1002/ehf2.15038] [Medline: 39219247]
27. Amigo JM, Babamoradi H, Elcoroaristizabal S. Hyperspectral image analysis. A tutorial. Anal Chim Acta 2015 Oct 08;896:34-51. [doi: 10.1016/j.aca.2015.09.030] [Medline: 26481986]
28. Hu J, Xu J, Li M, Jiang Z, Mao J, Feng L, et al. Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. EClinicalMedicine 2024 Feb;68:102409 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102409] [Medline: 38273888]
29. Deng H, Eftekhari Z, Carlin C, Veerapong J, Fournier KF, Johnston FM, et al. Development and validation of an explainable machine learning model for major complications after cytoreductive surgery. JAMA Netw Open 2022 May 02;5(5):e2212930 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.12930] [Medline: 35612856]
30. Gao Z, Liu X, Kang Y, Hu P, Zhang X, Yan W, et al. Improving the prognostic evaluation precision of hospital outcomes for heart failure using admission notes and clinical tabular data: multimodal deep learning model. J Med Internet Res 2024 May 02;26:e54363 [FREE Full text] [doi: 10.2196/54363] [Medline: 38696251]
31. Huang T, Xu H, Wang H, Huang H, Xu Y, Li B, et al. Artificial intelligence for medicine: Progress, challenges, and perspectives. The Innovation Medicine 2023;1(2):100030 [FREE Full text] [doi: 10.59717/j.xinn-med.2023.100030]

32. Xue X, Sun H, Yang M, Liu X, Hu H, Deng Y, et al. Advances in the application of artificial intelligence-based spectral data interpretation: a perspective. Anal Chem 2023 Sep 19;95(37):13733-13745. [doi: 10.1021/acs.analchem.3c02540] [Medline: 37688541]
33. Hu W, Huang Y, Wei L, Zhang F, Li H. Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors 2015:1-12 [FREE Full text] [doi: 10.1155/2015/258619]
34. Hong D, He W, Yokoya N, Yao J, Gao L, Zhang L, et al. Interpretable hyperspectral artificial intelligence: when nonconvex modeling meets hyperspectral remote sensing. IEEE Geosci. Remote Sens. Mag 2021 Jun;9(2):52-87. [doi: 10.1109/mgrs.2021.3064051] [Medline: 39604036]
35. Akerman AP, Porumb M, Scott CG, Beqiri A, Chartsias A, Ryu AJ, et al. Automated echocardiographic detection of heart failure with preserved ejection fraction using artificial intelligence. JACC Adv 2023 Aug;2(6):100452 [FREE Full text] [doi: 10.1016/j.jacadv.2023.100452] [Medline: 38939447]
36. Kittleson M, Panjrath G, Amancherla K, Davis L, Deswal A, Dixon D, et al. 2023 ACC Expert Consensus Decision Pathway on Management of Heart Failure With Preserved Ejection Fraction: a report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol 2023 May 09;81(18):1835-1878 [FREE Full text] [doi: 10.1016/j.jacc.2023.03.393] [Medline: 37137593]
37. Shah SJ, Borlaug BA, Kitzman DW, McCulloch AD, Blaxall BC, Agarwal R, et al. Research priorities for heart failure with preserved ejection fraction. Circulation 2020 Mar 24;141(12):1001-1026. [doi: 10.1161/circulationaha.119.041886] [Medline: 32202936]
38. Rodríguez-Pérez R, Bajorath J. Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley values. J Med Chem 2020 Aug 27;63(16):8761-8777. [doi: 10.1021/acs.jmedchem.9b01101] [Medline: 31512867]
39. Kokkonen J, Mustonen P, Heikkilä E, Leskelä RL, Pennanen P, Krühn K, et al. Effectiveness of telemonitoring in reducing hospitalization and associated costs for patients with heart failure in Finland: nonrandomized pre-post telemonitoring study. JMIR Mhealth Uhealth 2024 Feb 07;12:e51841 [FREE Full text] [doi: 10.2196/51841] [Medline: 38324366]
40. Craig W, Ohlmann S. The benefits of using active remote patient management for enhanced heart failure outcomes in rural cardiology practice: single-site retrospective cohort study. J Med Internet Res 2024 Nov 26;26:e49710 [FREE Full text] [doi: 10.2196/49710] [Medline: 39589775]
41. Denny JC, Collins FS. Precision medicine in 2030-seven ways to transform healthcare. Cell 2021 Mar 18;184(6):1415-1419 [FREE Full text] [doi: 10.1016/j.cell.2021.01.015] [Medline: 33740447]

Abbreviations

AUC: area under the curve
DCA: decision curve analysis
ENVI: Environment for Visualizing Images
HFpEF: heart failure with preserved ejection fraction
HSI: hyperspectral imaging
LASSO: least absolute shrinkage and selection operator
NPV: negative predictive value
PPV: positive predictive value
ROC: receiver operating characteristic
SHAP: Shapley Additive Explanations
SVM: support vector machine

Edited by A Coristine; submitted 07.10.24; peer-reviewed by B Wu, GH He; comments to author 13.11.24; revised version received 04.12.24; accepted 20.12.24; published 07.01.25.

Please cite as:
Yang X, Li Z, Lei L, Shi X, Zhang D, Zhou F, Li W, Xu T, Liu X, Wang S, Yuan Q, Yang J, Wang X, Zhong Y, Yu L
Noninvasive Oral Hyperspectral Imaging–Driven Digital Diagnosis of Heart Failure With Preserved Ejection Fraction: Model Development and Validation Study
J Med Internet Res 2025;27:e67256
URL: https://www.jmir.org/2025/1/e67256
doi:10.2196/67256
PMID:

Original Paper

The Virtual Client Experience Survey for Mental Health and Addictions: Revalidation of a Survey to Measure Client and Family Experiences of Virtual Care

Allison Crawford[1], MD, PhD; Anne Kirvan[1], MSW; Marcos Sanches[1], MSc; Amanda Gambin[1], PhD; Denise Canso[1], MSc; Eva Serhal[1], MBA, PhD

Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Allison Crawford, MD, PhD
Centre for Addiction and Mental Health
60 White Squirrel Way
Office 223
Toronto, ON, M6J1H4
Canada
Phone: 1 416 535 8501
Email: allison.crawford@utoronto.ca

Abstract

Background: The onset of the COVID-19 pandemic precipitated a rapid shift to virtual care in health care settings, inclusive of mental health care. Understanding clients' perspectives on virtual mental health care quality will be critical to informing future policies and practices.

Objective: This study aimed to outline the process of redesigning and validating the Virtual Client Experience Survey (VCES), which can be used to evaluate client and family experiences of virtual care, specifically virtual mental health and addiction care.

Methods: The VCES was adapted from a previously validated telepsychiatry survey. All items were reviewed and updated, with particular attention to the need to ensure relevance across mental health care sectors and settings. The survey was then revalidated using the 6 domains of health care quality of the Institute of Medicine (IOM) as a guiding framework. These 6 domains include being safe, effective, patient-centered, efficient, timely, and equitable. The VCES was piloted with a convenience sample of clients and family members accessing outpatient care at the Centre for Addiction and Mental Health (CAMH) in Toronto, Ontario, through video or telephone. A confirmatory factor analysis (CFA) was conducted in MPlus and used to test the factorial structures of the VCES, with minor respecification of the model based on modification indices, factor loadings, reliability, and item-total correlation. The respecifications were checked for alignment with the construct definitions and item interpretation. The reliability of the constructs was estimated by the Cronbach α coefficient.

Results: The survey was completed 181 times. The construct reliability was generally high. Timely was the only subscale with an α lower than 0.7; all others were above 0.8. In all cases, the corrected item-total correlation was higher than 0.3. For the CFA, the model was adjusted after multiple imputations with 20 datasets. The mean chi-square value was 437.5, with df=199 (P<.001). The mean root mean square error of approximation (RMSEA) was 0.08 (SD 0.002), the mean confirmatory fit index (CFI) was 0.987 (SD 0.001), the mean Tucker-Lewis Index (TLI) was 0.985 (SD 0.001), and the mean standardized root mean square residual (SRMR) was 0.04 (SD 0.001).

Conclusions: This study describes the validation of the VCES to evaluate client and family experiences of virtual mental health and addictions care. Given the widespread uptake of virtual care, this survey has broad applicability across settings that provide mental health and addiction care. The VCES can be used to guide targeted quality improvement initiatives across health care quality domains. By effectively addressing challenges as they emerge, it is anticipated that we will continue to move toward hybrid modalities of practice that leverage the strengths and benefits of telephone, video, and in-person care to effectively respond to unique client and family needs and circumstances.

(J Med Internet Res 2025;27:e49844) doi:10.2196/49844

KEYWORDS

virtual care; digital health; mental health; client satisfaction; health care quality; Virtual Client Experience Survey; telehealth; telemedicine; eHealth

Introduction

The onset of the COVID-19 pandemic precipitated a rapid shift to virtual care in health care settings, inclusive of mental health care. Virtual care, often referred to as telehealth or telemedicine, refers to the provision of clinical care using information and communication technologies and can be delivered synchronously (ie, in real-time, at a distance) or asynchronously (ie, separated by time and space) [1,2]. During the early phases of the pandemic, synchronous forms of virtual care, such as video- and telephone-based care, replaced in-person care as the leading modalities of practice in many health care settings [3-6].

Research evaluating virtual mental health care has consistently shown high levels of both client satisfaction [3-7] and provider satisfaction [5,8]. However, recent literature also highlights challenges with virtual care, including inequities in access to and engagement with virtual care [9,10], technology challenges [11], lack of connection between clients and health care providers [4], and concerns regarding privacy and safety [12]. Understanding clients' perspectives on virtual mental health care quality will be critical to inform future policies and practices. In addition, it will be important to understand how these experiences vary across sociodemographic groups, particularly in light of long-standing inequities in health care access and outcomes that risk being reinforced by the widespread uptake of virtual care [10].

To date, there is only 1 known validated measure of client experiences of virtual mental health care, specifically telepsychiatry [6], and there are no known validated measures that can be used to evaluate client experiences across varied types of mental health care. Since the transition to virtual care, a majority of patients across multiple surveys indicate a preference for virtual care to continue to exist as an option [3,4], with many preferring a hybrid approach that combines both in-person and virtual modalities [3]. Validated means to understand client and family experiences of care will be critical to inform evidence-based decision-making regarding virtual care and hybrid modalities of practice moving forward.

This paper outlines the process to redesign and validate the Virtual Client Experience Survey (VCES) that can be used to evaluate client and family experiences of virtual mental health and addictions care.

Methods

Overview

At the outset of the COVID-19 pandemic, the VCES was adapted from a previously validated telepsychiatry survey developed and used within the TeleMental Health program at the Centre for Addiction and Mental Health (CAMH), a mental health and addiction hospital located in Toronto, Ontario [6]. This telepsychiatry survey was updated to respond to the pressing need to evaluate virtual care both within our hospital and externally. Several adaptations were made to extend the use of the tool beyond its original intended use with clients accessing psychiatric consultation and assessment through videoconference [6].

Survey Redesign

The approach to developing and validating the original telepsychiatry survey is outlined in the paper by Serhal et al [6]. Our team of clinicians, researchers, and health care leaders reviewed the telepsychiatry survey and adapted it for use in the current context. All items were reviewed and updated, in particular, with the need to ensure relevance that is accorded with the expansion of virtual care provision beyond physicians and psychiatrists. For example, items on the newly developed VCES ask about the client's experience with the virtual appointment instead of the telepsychiatry appointment. Another item asks whether the health care provider explained the risks and benefits of treatments or interventions rather than the risks and benefits of medications. After consultation with clinicians across multiple programs and services at the hospital, the survey was also updated to ensure applicability not only to clients but to family members as well. This was an important shift given the key role that families play in mental health and addiction care and recovery [13,14] and in supporting access to care through videoconference.

The transition to virtual care during the COVID-19 pandemic also necessitated the examination of virtual care from a digital health equity perspective [10]. Our team consulted with the CAMH Health Equity Office as well as the Provincial System Support Program (PSSP), which administers the Ontario Perception of Care Tool for Mental Health and Addictions (OPOC-MHA) [15], regarding the addition of sociodemographic questions to facilitate health equity analyses. Furthermore, 5 items from the OPOC-MHA were also added to facilitate alignment and comparison between the 2 tools.

The VCES sociodemographic questions ask about gender, age, geographic region, whether the client was born in Canada, racial or ethnic group, illness and disability, and other factors. Clients and families were also asked about their comfort with technology, where they accessed the virtual care appointment from (eg, home, health care organization), the type of device they used to access virtual care (eg, computer, smartphone), and which videoconference platform was used. The telephone was included as an option on the survey, as CAMH, like many health care settings, began providing more services by telephone during the pandemic, particularly for individuals experiencing barriers to video-based care. Finally, as recommended by Serhal et al [6], we added additional survey items to assess access to care, and physical and emotional safety during virtual care appointments. We also added an item about compassionate care, given our research group's interest in compassion as a critical factor in the therapeutic relationship and inpatient and family experience of care [16,17].

The redesigned VCES consists of 22 items to assess the overall quality of care. These items are rated on a 4-point Likert scale (1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree) with additional options including "not applicable" and "prefer not to answer." The final question on the survey is an open-text question to elicit any additional feedback.

Conceptual Framework

Following the adaption of the survey, the survey was revalidated using the 6 domains of health care quality of the Institute of Medicine (IOM) as a guiding framework. These 6 domains include being safe, effective, patient-centered, efficient, timely, and equitable [18]. The IOM domains were selected as a guiding framework to provide a comprehensive evaluation of client satisfaction and experience and to guide targeted quality improvement efforts across the different domains. The heath quality domain definitions [18] have been adapted in Table 1 below for the VCES.

Table 1. Health quality domain definitions.

Health quality domains	Original definition	Adapted definition for the VCES[a] domains
Safe	"Avoiding injuries to patients from the care that is intended to help them [18]."	Virtual care that is physically and psychologically safe and minimizes potential risks.
Effective	"Providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and overuse)" [18]	Effective delivery of virtual care, inclusive of the use of virtual care technologies, to facilitate engagement in virtual care.
Patient-centered	"Providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions" [18]	Virtual care that is collaborative, respectful, and responsive to client and family preferences, needs, and values.
Efficient	"Avoiding waste, in particular, waste of equipment, supplies, ideas, and energy" [18]	Efficiency and ease in accessing virtual care.
Timely	"Reducing waits and sometimes harmful delays for both those who receive and those who give care" [18]	Timely access to virtual care for clients and families, including limited wait times for services.
Equitable	"Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status" [18]	Virtual care that does not vary in quality based on personal characteristics and sociodemographic factors.

[a] VCES: Virtual Client Experience Survey.

Question Validation

Survey Dissemination

The VCES was piloted with a convenience sample of clients (16 years of age and older) and family members accessing a wide range of outpatient mental health and addiction services at CAMH through video or telephone. Survey data was collected using Research Electronic Data Capture (REDCap), a web-based application for securely collecting surveys and managing data [19,20]. The electronic survey took approximately 15 minutes to complete. Survey completion was voluntary, and all survey responses were anonymous. Survey data collection took place in 2021.

Information about the VCES and the process for survey administration was shared with clinicians and administrative staff using multiple channels, including the organizational intranet, email, support and sponsorship from clinical program directors, and program-level presentations by the project team. CAMH clinicians and administrative staff were asked to include information about the VCES, including the REDCap survey links, by email to all CAMH clients and families before their virtual care appointment. A template script for email communication to clients was provided, including a reminder that survey completion was voluntary and that all responses were anonymous. Clients and families were asked to complete a survey once per program or service that they participated in. The survey was made available to clients seeing providers across programs at the hospital, which includes multidisciplinary care providers (eg, physicians, nurses, social workers, occupational therapists, psychologists).

Content Validity

The 22 Likert-scale items on the VCES were independently reviewed by a panel of 5 individuals. including clinicians, researchers, and health care leaders, to determine alignment with the following IOM health quality domains: safe, effective, patient-centered, efficient, and timely [18]. Equity, the sixth IOM domain, is assessed through a comparison of survey response options by sociodemographic group. The panel then met through videoconference to discuss discrepancies until a consensus was reached.

Factorial Structure

To test the factorial structure of the VCES, a confirmatory factor analysis (CFA) was adjusted in Mplus (version 8.2; Muthén & Muthén) [21] using the method of weighted least squares with mean and variance adjusted chi-square test (WLSMV), which is recommended for ordinal outcomes [22]. Multiple imputation using the covariance method [23], with 20 imputed datasets, was used to handle missing values, which were present in 45% of the surveys, with 80% of surveys having 3 or fewer items missing. The evaluation of the model relied on the fit indices RMSEA (root means square error of approximation) [24], CFI (comparative fit index) [25], TLI (Tucker-Lewis index) [26], and SRMR (standardized root mean square residual) [27], as well as the inspection of standardized factor loadings, variance explained, and modification index relative to each item. The initial factorial structure was defined following expert opinion, with minor respecification of the model based on modification indices, factor loadings, reliability, and item-total correlation. The respecifications were checked for alignment with the construct definitions and item interpretation. Cutoffs for fit indices in CFA and reliability are not universally accepted [28]. That said, based on results from Hu and Bentler [29], TLI and CFI higher than 0.95, SRMR lower than 0.08, and RMSEA lower than 0.06 are considered acceptable.

Reliability

The reliability of the constructs was estimated by the Cronbach α coefficient [30], a measure of internal consistency between the items. The corrected item-total correlation (correlation between the item and the total scale score with the item removed) and the α coefficient for the scale with the item removed were also used as sources of information as to how the items fit the factors. A Cronbach α of 0.7 or higher is considered good [28].

Ethical Considerations

Consent to participate was implied. The survey contained an introductory section explaining the purpose of the survey, informing potential respondents that participation was voluntary and anonymous, and informing them that the findings would be presented as grouped or aggregated data for any presentation or publication. There was no compensation offered to clients and families for completing the survey. This project received ethical approval from CAMH Quality Projects Ethics Review (QPER-2021-000).

Results

Descriptive Analysis

In total, the survey was completed 181 times, with 99 surveys completed by clients, 16 by support persons on behalf of clients, and 22 by family members or caregivers. Respondent information is missing on 44 surveys. Sociodemographic characteristics are included in Table 2.

Most respondents used a computer (113/179, 62.4%) or tablet or smartphone (58/179, 32%). Only 5.5% (10/179) reported using a telephone (audio only). The majority of respondents were either comfortable with technology (64/179, 35.8%) or very comfortable (85/179, 47.5%).

Table 2. Participant characteristics.

Characteristics	Participants, n (%)
Role	
Registered CAMH[a] client	99 (72.3)
Supporting a registered CAMH client	16 (11.7)
Support person	22 (16.1)
Sex	
Female	95 (52.5)
Male	76 (42)
Transgender: female to male	3 (1.7)
Transgender: male to female	2 (1.1)
Other	4 (2.2)
Do not know	1 (0.6)
Age group	
13-18 years	7 (3.9)
19-25 years	52 (28.7)
26-34 years	59 (32.6)
35-44 years	17 (9.4)
45-54 years	24 (13.3)
55-64 years	19 (10.5)
65+	3 (1.7)
Born in Canada	
Yes	121 (67.2)
No	59 (32.6)
Prefer not to answer	1 (0.6)
Ethnicity	
Asian: East (eg, Chinese, Japanese, Korean)	8 (4.4)
Asian: South (eg, Indian, Pakistani, Sri Lankan)	18 (9.9)
Asian: South East (eg, Malaysian, Filipino, Vietnamese)	4 (2.2)
Black: African (e.g., Ghanaian, Kenyan, Somali)	9 (5)
Black: Caribbean (eg, Barbadian, Jamaican)	5 (2.8)
Black: North American (eg, Canadian, American)	2 (1.1)
First Nations	2 (1.1)
Indian: Caribbean (eg, Guyanese with origins in India)	1 (0.6)
Latin American (eg, Argentinean, Chilean, Salvadoran)	7 (3.9)
Métis	1 (0.6)
Middle Eastern (eg, Egyptian, Iranian, Lebanese)	9 (5)
White: European (eg, English, Italian, Portuguese, Russian)	29 (16)
White: North American (eg, Canadian, American)	73 (40.3)
Mixed heritage (eg, Black-African and White-North American)	6 (3.3)
Other(s)	3 (1.7)
Prefer not to answer	4 (2.2)
Illness and disability	
Chronic illness	24 (13.3)

Characteristics	Participants, n (%)
Developmental disability	25 (13.8)
Substance use disorder	21 (11.6)
Learning disability	28 (15.5)
Mental illness	116 (65.1)
Physical disability	7 (3.9)
Sensorial disability	3 (1.7)
None	30 (16.6)
Others	6 (3.3)
Prefer not to answer	14 (7.7)

[a] CAMH: Centre for Addiction and Mental Health.

Reliability

The construct reliability, as measured by Cronbach α, was generally high. Timely was the only subscale with an α lower than 0.7; all others were above 0.8 (more details in Table 3). In all cases, the corrected item-total correlation was higher than 0.3. Considering the items in the model, item 9 (effective) and items 2 and 10 (timely) had lower corrected item-total correlation, with all the other items having correlations around 0.7 or above. Within the “timely” factor, the low item-total correlation was likely reflective of the small number of items associated with this domain.

Table 3. Confirmatory factor analysis results and item statistics.

Factors and survey items	Loading[a]	SE[b]	Missing[c], %	R^2 [d]	Correlation[e]	Mean[f] (SD[g])
Efficient (Cronbach α[h]>=0.87)						
1. It was easy to access virtual care at CAMH[i].	0.88	0.03	12	0.78	0.79	3.47 (0.71)
3. It was easy to book my virtual appointment.	1.00	0.02	9	1.00	0.85	3.51 (0.66)
21. The physical location of where I accessed my virtual appointment was convenient for me.	0.88	0.03	8	0.78	0.72	3.54 (0.72)
Timely (Cronbach α=0.66)						
2. The wait time for services was reasonable for me.[j]	0.72	0.06	23	0.52	0.46	3.26 (0.82)
10. I was able to get a virtual appointment sooner than an in-person healthcare appointment.	0.81	0.06	42	0.66	0.46	3.27 (0.76)
Effective (Cronbach α=0.80)						
4. During my virtual appointment, I was able to see the healthcare provider clearly.	0.89	0.03	10	0.79	0.69	3.54 (0.70)
5. During my virtual appointment, I was able to hear the healthcare provider clearly.	0.91	0.02	6	0.84	0.69	3.43 (0.75)
9. I believe virtual care is just as effective as in-person healthcare.	0.55	0.06	15	0.31	0.38	2.99 (0.89)
22. Overall, I am satisfied with my virtual appointment.	0.96	0.02	9	0.92	0.74	3.51 (0.72)
Patient-centered (Cronbach α=0.96)						
6. I am confident that the healthcare provider at CAMH and my other service providers are working as a team.	0.85	0.03	10	0.72	0.77	3.48 (0.73)
7. I feel that there was an adequate amount of time allotted for the virtual appointment.	0.87	0.03	5	0.75	0.79	3.54 (0.68)
8. I felt comfortable during my virtual appointment.	0.92	0.02	13	0.85	0.81	3.50 (0.77)
12. Staff understood and responded to my needs and concerns.[j]	0.95	0.01	16	0.89	0.83	3.52 (0.72)
13. I was treated with respect by program staff.[j]	0.96	0.01	7	0.92	0.89	3.62 (0.71)
14. I received compassionate virtual care.	0.97	0.01	15	0.94	0.91	3.59 (0.73)
16. The healthcare provider spoke with me about my mental health and/or addiction in a way that I could understand.	0.97	0.01	27	0.95	0.89	3.54 (0.73)
17. I was involved as much as I wanted to be in decisions about my treatment services and supports.[j]	0.97	0.01	29	0.94	0.89	3.54 (0.69)
19. I am confident that I will be able to follow the healthcare provider's recommendations.	0.87	0.03	16	0.75	0.75	3.36 (0.76)
Safe (Cronbach α=0.89)						
11. I was assured my personal information was kept confidential.[j]	0.87	0.03	10	0.76	0.80	3.50 (0.74)
15. I felt safe (emotionally and physically) during my virtual appointment.	0.99	0.01	8	0.98	0.82	3.56 (0.74)
18. The healthcare provider explained to me the benefits and risks of any treatments or interventions that were recommended during my virtual appointment.	0.89	0.02	15	0.79	0.79	3.45 (0.76)
20. I understand what to do if I have a mental health and/or addiction emergency following this appointment.	0.82	0.03	11	0.67	0.69	3.44 (0.73)

[a] Confirmatory factor analysis standardized loadings.

[b] Standard error of the confirmatory factor analysis loading.

[c] Proportion of the subject with missing values in the item. These items were imputed using multiple imputation for the confirmatory factor analysis

analysis.

[d] Variance explained by the confirmatory factor analysis model.

[e] Corrected item-total correlation (correlation between the item and the sum score with the item removed).

[f] Item mean (Item values range from 1=strongly disagree to 4=strongly disagree).

[g] Item standard deviation.

[h] Cronbach α (internal consistency).

[i] CAMH is referenced in the survey that was disseminated internally. The external version refers to “this organization.”

[j] Items from the Ontario Perception of Care Tool for Mental Health and Addictions (OPOC-MHA) [15].

Confirmatory Factor Analysis

Confirmatory Factor Analysis loadings are shown in Table 3 above. The model was adjusted after multiple imputations with 20 datasets. We show fit statistics using the mean and SD of these 20 datasets. The mean chi-square value was 437.5, with df=199 (P<.001). The mean RMSEA was 0.08 (SD 0.002), which is considered reasonable. The mean CFI was 0.987 (SD 0.001), the mean TLI was 0.985 (SD 0.001), and the mean SRMR was 0.04 (SD 0.001). These values are considered to be quite good. With the exception of item 9 in the “effective” factor, all the others have standardized loadings higher than 0.7.

The model shown in Table 3 had 2 respecifications: item 19 was moved from “effective” to “patient-centered,” and item 18 was moved from “patient-centered” to “safe.” These changes were based on the modification index and discussions with the project team; however, they did not meaningfully change the fit measures. While item 9 does not fit as well in the “effective factor,” it was maintained as is, as it represented the best conceptual fit.

The sample size for structural equation models is not well defined because of the high number of parameters of interest and broad range of model complexity. Considering our model with 5 factors with an average of 4 indicators per factor and loadings around 0.8, Wolf et al [31] found that a sample size of 120 tends to be sufficient for power above 80%. These results are negatively affected by the presence of missing values; as such, our sample of 181 is expected to have reasonable power for a CFA analysis.

Discussion

Principal Findings

This paper provides a valid and reliable measure to evaluate client and family experiences of virtual mental health and addiction care, focusing on 6 domains of health care quality. This survey has been used internally at CAMH to guide the next steps and recommendations for virtual mental health care and has been shared externally in both English and French to support quality measurement within other organizations. Information about the survey was shared through a webinar, and the survey was subsequently downloaded hundreds of times across Canada and internationally.

Implications for Virtual Mental Health Care

Given the widespread uptake of virtual care, this survey has broad applicability across settings that provide mental health and addiction care. The VCES can be used to guide targeted quality improvement initiatives across the 6 IOM domains of health care quality, including safe, effective, patient-centered, efficient, timely, and equitable care [18]. It also has the potential to be adapted to other health care specialties and contexts.

Considerations for Administering the VCES

From our experience, we found the use of REDCap (Research Electronic Data Capture; Vanderbilt University) to be an effective means to disseminate the survey, collate results, and ensure anonymity. Some training and guidance regarding survey administration processes were required to ensure that the survey was delivered as part of the standard of care. We found it advantageous to delineate a time-limited window for survey dissemination and completion to limit time demands on both clients and clinicians. In terms of interpreting results, we set a cutoff of 3 (agree) and above as a positive result indicating high quality of care. Finally, there was a high degree of interest in receiving feedback on the survey results. We offered aggregated program-level reports once an adequate sample of surveys was completed in order to preserve respondent anonymity; this is particularly important as demographic questions attached to the survey could render individual results identifiable.

Next Steps

While the findings from the VCES provided invaluable and timely feedback, it is important not to limit patient and family feedback to a survey. The open text option on the survey provided a breadth of feedback not necessarily addressed within the existing survey questions. Analysis of this qualitative data may spur future adaptations to the survey and, on a practical level, may steer quality improvement initiatives resulting from survey findings. Furthermore, ongoing patient and family engagement in interpreting survey results and prioritizing and guiding quality improvement will be essential to making meaningful and impactful improvements.

Patient experience of virtual care should also be considered alongside provider experience of virtual care. To complement findings from the VCES and to capture another key perspective within health care, our team has developed and validated the Virtual Provider Experience Survey (VPES). These complementary tools provide key metrics to inform how health care organizations and systems can move toward the “quadruple aim,” a well-recognized framework for health care that seeks to improve population health outcomes, reduce health care costs, and improve patient, family, and provider experiences of care [32]. The consistent use of standardized tools across settings is a valuable way to inform policy and program planning with respect to virtual care [33] and to further our advancement toward the quadruple aim. Further practice-based research is also needed to examine the relationship between the experience of care and health outcomes for clients and families who access virtual mental health and addiction care and the cost-effectiveness of those services.

Limitations

While the VCES was validated with 181 survey responses, it is possible that some clients or family members completed the survey more than once if they were involved in multiple programs or services during the evaluation period. In addition, when administering surveys, there is always the risk of self-selection and response bias. We sought to mitigate this risk by asking clinicians and administrative staff to send the survey link to all clients and families accessing virtual care at CAMH. A response rate can be beneficial to ascertain the extent of potential bias; however, we were unable to determine the response rate, as it is unknown how many clients and families received the link in their email correspondence from clinicians or administrative staff. It is quite likely that those who have a higher degree of familiarity or comfort with virtual care were more likely to complete an online survey sent by email.

Conclusion

We sought to address a notable gap in the literature by redesigning and validating a measure of client and family experiences of virtual mental health care. Client and family perspectives are needed to evaluate the current state of mental health service delivery and highlight areas that need improvement. By effectively addressing challenges as they emerge, it is anticipated that we will continue to move toward hybrid modalities of practice that leverage the strengths and benefits of telephone, video, and in-person care to effectively respond to unique client and family needs and circumstances [11].

Acknowledgments

The authors would like to acknowledge members of the virtual mental health team at CAMH for supporting the development and dissemination of the VCES, as well as CAMH clinicians and administrative staff for their support in sharing the VCES with clients and families. We would also like to acknowledge the clients and families who took the time to share their feedback on their experiences of virtual care. Generative AI was not used for any section of the manuscript. No funds, grants, or other financial support were received for this project.

Data Availability

The data that support the findings presented in this report are available upon request from the corresponding author.

Authors' Contributions

AC contributed to study conceptualization, data analysis, writing, and review. AK contributed to study conceptualization, data collection and analysis, writing, and review. MS contributed to data analysis, writing, and review of the manuscript. AG contributed to data collection and analysis, review, and editing of the manuscript. DC contributed to the study conceptualization, data collection and analysis, review, and editing of the manuscript. ES contributed to the study conceptualization, data analysis, review, and editing of the manuscript.

Conflicts of Interest

None declared.

References

1. Stamenova V, Agarwal P, Kelley L, Fujioka J, Nguyen M, Phung M, et al. Uptake and patient and provider communication modality preferences of virtual visits in primary care: a retrospective cohort study in Canada. BMJ Open 2020;10(7):e037064 [FREE Full text] [doi: 10.1136/bmjopen-2020-037064] [Medline: 32636284]
2. Mechanic O, Persaud Y, Kimball A. Telehealth systems. In: StatPearls. Florida, United States: StatPearls Publishing; 2023.
3. Benudis A, Re'em Y, Kanellopoulos D, Moreno A, Zonana J. Patient and provider experiences of telemental health during the COVID-19 pandemic in a New York city academic medical center. Psychiatry Res 2022;311:114496 [FREE Full text] [doi: 10.1016/j.psychres.2022.114496] [Medline: 35306378]
4. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Patient attitudes toward telepsychiatry during the COVID-19 pandemic: a nationwide, multisite survey. JMIR Ment Health 2020;7(12):e24761 [FREE Full text] [doi: 10.2196/24761] [Medline: 33302254]
5. Schubert NJ, Backman PJ, Bhatla R, Corace KM. Telepsychiatry and patient-provider concordance. Can J Rural Med 2019;24(3):75-82 [FREE Full text] [doi: 10.4103/CJRM.CJRM_9_18] [Medline: 31249155]
6. Serhal E, Kirvan A, Sanches M, Crawford A. Client satisfaction and experience with telepsychiatry: development and validation of a survey using clinical quality domains. J Med Internet Res 2020;22(9):e19198 [FREE Full text] [doi: 10.2196/19198] [Medline: 32755896]
7. Torales J, Vilallba-Arias J, Bogado JA, O'Higgins M, Almirón-Santacruz J, Ruiz Díaz N, et al. Satisfaction with telepsychiatry during the COVID-19 pandemic: patients' and psychiatrists' report from a university hospital. Int J Soc Psychiatry 2023;69(1):156-160 [FREE Full text] [doi: 10.1177/00207640211070762] [Medline: 34991382]

8. Wright J, Dewan S, Hilty D, Dewan NA. Health care providers' perceptions of quality, acceptance, and satisfaction with telebehavioral health services during the COVID-19 pandemic: survey-based study. JMIR Ment Health 2020;7(12):e23245 [FREE Full text] [doi: 10.2196/23245] [Medline: 33180740]
9. Budhwani S, Fujioka J, Thomas-Jacques T, De Vera K, Challa P, De Silva R, et al. Challenges and strategies for promoting health equity in virtual care: findings and policy directions from a scoping review of reviews. J Am Med Inform Assoc 2022;29(5):990-999 [FREE Full text] [doi: 10.1093/jamia/ocac022] [Medline: 35187571]
10. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
11. Budhwani S, Fujioka JK, Chu C, Baranek H, Pus L, Wasserman L, et al. Delivering mental health care virtually during the COVID-19 pandemic: qualitative evaluation of provider experiences in a scaled context. JMIR Form Res 2021;5(9):e30280 [FREE Full text] [doi: 10.2196/30280] [Medline: 34406967]
12. Cowan KE, McKean AJ, Gentry MT, Hilty DM. Barriers to use of telepsychiatry: clinicians as gatekeepers. Mayo Clin Proc 2019;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
13. Kourgiantakis T, Ashcroft R, Mohamud F, Fearing G, Sanders J. Family-focused practices in addictions: a scoping review. Journal of Social Work Practice in the Addictions 2021;21(1):18-53. [doi: 10.1080/1533256x.2020.1870287]
14. Kourgiantakis T, McNeil SR, Hussain A, Logan J, Ashcroft R, Lee E, et al. Social work’s approach to recovery in mental health and addiction policies: a scoping review. Social Work in Mental Health 2022;20(4):377-399. [doi: 10.1080/15332985.2021.2019169]
15. Rush DB, Lentinello E, Khalkhali S, Cvetanova Y, Rotondi N, Furlong A, et al. Development of a client perception of care tool for mental health and addictions: qualitative and quantitative psychometric analysis. Centre for Addiction and Mental Health. 2014. URL: https://kmb.camh.ca/eenet/sites/default/files/wp-content/uploads/2016/04/OPOC-Final-Report_July-2014.pdf [accessed 2024-10-22]
16. Kemp J, Zhang T, Inglis F, Wiljer D, Sockalingam S, Crawford A, et al. Delivery of compassionate mental health care in a digital technology-driven age: scoping review. J Med Internet Res 2020;22(3):e16263 [FREE Full text] [doi: 10.2196/16263] [Medline: 32141833]
17. Wiljer D, Strudwick G, Crawford A. Caring in a digital age: exploring the interface of humans and machines in the provision of compassionate healthcare. In: Without Compassion, There Is No Healthcare: Leading with Care in a Technological Age. Ontario: McGill University Press; 2020:33-57.
18. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the quality chasm: A new health system for the 21st century. National Academies Press. 2001. URL: http://www.ncbi.nlm.nih.gov/books/NBK222274/ [accessed 2020-04-02]
19. 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;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. 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;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
21. Muthén LK, Muthén BO. Mplus User's Guide. Eighth edition. Statmodel. 2017. URL: https://www..com/download/usersguide/MplusUserGuideVer_8.pdf [accessed 2023-03-23]
22. Muthén BO. Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Statmodel. 1997. URL: https://www.statmodel.com/download/Article_075.pdf [accessed 2023-03-23]
23. Asparouhov T, Muthen B. MPlus Web Notes: Multiple imputation with Mplus. Statmodel. URL: https://tinyurl.com/36d7m9zy [accessed 2024-11-18]
24. Steiger JH. Structural model evaluation and modification: an interval estimation approach. Multivariate Behav Res 1990;25(2):173-180. [doi: 10.1207/s15327906mbr2502_4] [Medline: 26794479]
25. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107(2):238-246. [doi: 10.1037/0033-2909.107.2.238] [Medline: 2320703]
26. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 1980;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]
27. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal 2007;14(3):464-504. [doi: 10.1080/10705510701301834]
28. Lance CE, Butts MM, Michels LC. The sources of four commonly reported cutoff criteria. Organizational Research Methods 2006;9(2):202-220. [doi: 10.1177/1094428105284919]
29. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
30. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):297-334. [doi: 10.1007/BF02310555]

31. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas 2013;76(6):913-934 [FREE Full text] [doi: 10.1177/0013164413495237] [Medline: 25705052]
32. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
33. Barsom EZ, van Hees E, Bemelman WA, Schijven MP. Measuring patient satisfaction with video consultation: a systematic review of assessment tools and their measurement properties. Int J Technol Assess Health Care 2020:1-7. [doi: 10.1017/S0266462320000367] [Medline: 32624044]

Abbreviations

CAMH: Centre for Addiction and Mental Health
CFA: confirmatory factor analysis
CFI: confirmatory fit index
IOM: Institute of Medicine
OPOC-MHA: Ontario Perception of Care Tool for Mental Health and Addictions
PSSP: Provincial System Support Program
REDCap: Research Electronic Data Capture
RMSEA: root mean square error of approximation
SE: standard error
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index
VCES: Virtual Client Experience Survey
VPES: Virtual Provider Experience Survey
WLSMV: weighted least squares with mean and variance adjusted chi-square test

---

Edited by T de Azevedo Cardoso; submitted 11.06.23; peer-reviewed by J Sharp, SH Hong, A Makai, J Fitzsimon; comments to author 22.03.24; revised version received 23.05.24; accepted 23.09.24; published 03.01.25.

Please cite as:
Crawford A, Kirvan A, Sanches M, Gambin A, Canso D, Serhal E
The Virtual Client Experience Survey for Mental Health and Addictions: Revalidation of a Survey to Measure Client and Family Experiences of Virtual Care
J Med Internet Res 2025;27:e49844
URL: https://www.jmir.org/2025/1/e49844
doi:10.2196/49844
PMID:39752192

Original Paper

Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives

Bryan Cleal[1*], PhD; Yanbing Chen[2*], PhD; Mandy Wäldchen[3], MSc; Hanne Ballhausen[4], MSc; Drew Cooper[1,3], MSc; Shane O'Donnell[3], PhD; Christine Knoll[3,5], MD; Niklas Krug[5]; Klemens Raile[6], MD; Tebbe Ubben[4,7]; Adrian Tappe[4,7], BSc; Dana Lewis[3,8], BA; Ingrid Willaing[1,9], MPH; Timothy Skinner[10], PhD; Katarina Braune[1,5], MD

[1]Diabetes Management Research, Steno Diabetes Center Copenhagen, Herlev, Denmark
[2]Embry-Riddle Aeronautical University, Daytona Beach, FL, United States
[3]School of Sociology, University College Dublin, Belfield, Ireland
[4]#dedoc° Diabetes Online Community, Berlin, Germany
[5]Charité – Universitätsmedizin Berlin, Berlin Institute of Health at Charité, Berlin, Germany
[6]Department of Pediatrics, Vivantes Klinikum Neukölln, Berlin, Germany
[7]AndroidAPS, Hamilton, New Zealand
[8]OpenAPS, Seattle, WA, United States
[9]Department of Public Health, Copenhagen University, Copenhagen, Denmark
[10]Australian Centre for Behavioural Research in Diabetes, Melbourne, Australia
[*]these authors contributed equally

Corresponding Author:
Bryan Cleal, PhD
Diabetes Management Research
Steno Diabetes Center Copenhagen
Borgmester Ib Juuls Vej 83
Herlev, 2730
Denmark
Phone: 45 39 68 08 00
Email: bryan.richard.cleal@regionh.dk

Abstract

Background: Although commercially developed automated insulin delivery (AID) systems have recently been approved and become available in a limited number of countries, they are not universally available, accessible, or affordable. Therefore, open-source AID systems, cocreated by an online community of people with diabetes and their families behind the hashtag #WeAreNotWaiting, have become increasingly popular.

Objective: This study focused on examining the lived experiences, physical and emotional health implications of people with diabetes following the initiation of open-source AID systems, their perceived challenges, and their sources of support, which have not been explored in the existing literature.

Methods: We collected data from 383 participants across 29 countries through 2 sets of open-ended questions in a web-based survey on their experience of building and using open-source AID systems. Narratives were thematically analyzed, and a coding framework was identified through iterative alignment.

Results: Participants consistently reported improvements in glycemia, physical health, sleep quality, emotional impact on everyday life, and quality of life. Knowledge of open-source AID systems was largely obtained through the #WeAreNotWaiting community, which was also the primary source of practical and emotional support. The acquisition of the components to build an open-source AID system and the technical setup were sometimes problematic.

Conclusions: The #WeAreNotWaiting movement represents a primary example of how informed and connected patients proactively address their unmet needs, provide peer support to each other, and obtain results through impactful, user-driven solutions. Alongside providing evidence on the safety and efficacy of open-source AID systems, this qualitative analysis helps in understanding how patients' experiences and benefits range from psychosocial improvements to a reduction in the burden of managing diabetes.

International Registered Report Identifier (IRRID): RR2-10.2196/15368

(J Med Internet Res 2025;27:e48406) doi:10.2196/48406

KEYWORDS

automated insulin delivery; diabetes technology; type 1 diabetes; insulin pumps; continuous glucose monitoring; peer support; community support; open source; impact; users; diabetes; emotional health; challenges; support; unmet needs; mobile phone

Introduction

Overview

In recent years, much attention has been given to the beneficial impacts that online peer support has on people living with chronic health conditions, yet the exact nature of these impacts may still appear to be somewhat intangible [1]. Through our findings, we provide an example of peer support in which the focus of the interaction is very tangible and where the impacts are profound and wide-ranging. The case in point being type 1 diabetes (T1D), where, in recent years, patients have been taking on the role of innovators in the design and development of technology used for their treatment. People with T1D who have developed and disseminated open-source automated insulin delivery (AID) systems exemplify a potential within online peer-to-peer communities that is only just beginning to be realized, particularly among people with chronic health conditions. To better understand how developments have reached the cutting edge in T1D treatment, we first provide some background about the challenges of everyday diabetes management and how open-source AID technology holds promise to alleviate them.

Background and Challenges in T1D Management

T1D is a lifelong condition caused by the autoimmune-induced loss of insulin-producing cells in the pancreas. Until the discovery of insulin by Banting et al [2] a century ago, T1D inevitably resulted in death by ketoacidosis within months. This changed once pharmaceutically procured insulin was available. Yet, while developments of the last 100 years in pharmaceuticals and technology have improved the physical health and life expectancy of people with T1D from a biomedical perspective, the burden of managing the condition remains a challenge. Life with T1D exists in the center of a continuous data feedback loop, where dosing of exogenous insulin via subcutaneous injections or insulin pumps must be frequently adjusted in accord with glucose levels and predicted trends, carbohydrate intake, physical activity, individual physiology, and behavior, among a variety of other factors.

The #WeAreNotWaiting Movement

Given the complexity of diabetes management, the everyday experience of managing the condition may lead to frustration, a feeling that was also a key driver in the emergence of the movement that has subsequently become known as #WeAreNotWaiting [3]. Initially, this frustration was related to issues with the accessibility of data from continuous glucose monitors (CGMs) in real time. Taking matters into their own hands, individuals reverse engineered commercially available devices to enable uploading of device data to the cloud in real time. In a model of diffusion that has characterized subsequent developments, the source code was published as open-source software, freely available to anyone and linked by the social media hashtag #WeAreNotWaiting. Eventually, a large, global community has united under this banner.

One of the most significant innovations to emerge from the #WeAreNotWaiting movement are open-source AID systems—sometimes referred to as do-it-yourself artificial pancreas systems—such as OpenAPS [4], AndroidAPS [5], iAPS [6], and Loop [7]. These systems link CGM sensors and insulin pumps with predictive control algorithms running on smartphones or microcontrollers. Insulin dosing is automatically adjusted according to predictions based on real-time data from CGMs, individual settings, and user inputs such as meal information. This technology emerged well in advance of the recent availability of commercial AID systems in select countries [8].

Current Evidence and User Perspectives

Evidence from randomized controlled trials and observational studies has demonstrated the safety and efficacy of commercial AID systems [9], as well as a positive impact on the users' lived experience and quality of life [10]. Given the lengthy, complex development and approval processes, only few systems are currently licensed, and their functionality is limited due to regulatory constraints. Even in countries with market availability of commercial AID systems, they are not universally accessible to the user, with reimbursement policies for devices varying considerably between countries [11].

Open-source AID systems work much like commercial AID systems at a basic level, connecting devices and automatically adjusting insulin dosing via predictive algorithms. They differ from commercial systems in terms of device choice, interoperability, transparency, access, customization, and usability. While commercial systems traditionally include manufacturer-designed education for clinicians and users, open-source AID user support and education initially took place via peer support outside of clinical settings [12]. The source code for open-source AID systems is freely available from online portals. In addition, these communities have also created documentation that has been translated into many different languages. Peer support is freely available to help with the setup and use of these systems.

It is estimated that >10,000 individuals worldwide are currently using open-source AID systems, and the uptake continues to increase globally. Evidence from a recent randomized controlled trial [13] and real-world studies based on self- and caregiver-reported outcomes [14,15], in silico [16], user-provided data [17], and observational studies [18-20] point to the safety and effectiveness of these systems, with improvements in clinical parameters such as glycated hemoglobin (HbA_{1c}), time in range, occurrence of hypoglycemia, and glycemic variability. However, until recently, the perspective of the user and their reported outcomes has only been touched upon [21-25], and there remains much to learn about how open-source AID systems impact the lives of those who use them.

Study Objectives

The focus of this study was to establish the physiological, cognitive, and emotional impact of open-source AID system use. We further aimed to identify sources of support and challenges associated with setup, regular use, and maintenance of open-source AID systems. This study was designed and conducted by the OPEN (Outcomes of Patient's Evidence with Novel, Do-it-Yourself Artificial Pancreas Technology) consortium [26], an international, interdisciplinary team of patient innovators, clinicians, and scientists, many of whom also live with T1D and use open-source AID systems.

Methods

The research data were obtained from responses to open-ended questions included in a cross-sectional, web-based survey examining the use of open-source AID systems. The survey, titled DIWHY, was conducted between November 2018 and March 2019 [26].

Survey Design

The survey (Multimedia Appendix 1) was created by the OPEN consortium in collaboration with further open-source AID system users and was piloted with a small group of them before final release. The Checklist for Reporting Results of Internet E-Surveys was used to guide survey development [27]. The survey comprised 39 items in total with 2 sets of open-ended questions, which sought to capture lived experiences with open-source AID systems in the form of narratives. For these questions, participants could provide a free-text response with a maximum length of 1000 words each. The first set of questions assessed sources of information, support, and emotions associated with the preparation and setup process. The second question set addressed the impact of open-source AID system use on everyday life, associated changes, and challenges with respect to the transition to open-source AID systems.

Participants and Recruitment

Participants included in this study were adults (aged >18 y) who are living with diabetes (type 1, type 2, or other) and are current users of open-source AID systems. There were no restrictions in time since diagnosis or commencement of open-source AID systems. In this study, participants were exclusively recruited from the global diabetes online community, leveraging the outreach of the #WeAreNotWaiting movement. This approach enabled us to tap into a highly empowered and informed population of people with diabetes and particularly target those who were users of open-source AID systems. We used several social media channels, including the Facebook (Meta Platforms) groups Looped (with >6000 members) and AndroidAPS users (with >1800 members as of November 2018) and regional subgroups, and posted publicly on Twitter using the hashtags #WeAreNotWaiting and #DIYAPS to engage with the wider diabetes online community. Participation was further promoted through announcements on the OPEN project website. There were no paid promotions or targeted advertising on any platform. Participants were able to choose between 2 language options (English and German).

Data Collection and Analysis

Data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) [28], and the qualitative analysis of the narratives was performed using NVivo (version 12; Lumivero, 2018). The narratives were analyzed with an approach applying the principles of template analysis [29], in which a hierarchical coding structure is recommended to allow researchers to capture the diversity of meaning within broader overarching themes. This approach was deemed necessary due to the framing of the open-ended questions, in which respondents were prompted to reflect on specific aspects of their open-source AID system use experience. Therefore, the initial template and coding were deductively driven, with physical impact, emotional impact, sources of support, and challenges established as the overarching themes. In the subsequent stage of coding, the deductively coded data were analyzed inductively to identify subthemes within the overarching framework. Initially, 3 coders worked (BC, YC, and MW) on a small sample of the deductive data and presented putative inductive themes for evaluation within the author group. These were discussed and refined to ensure that there was equivalence in relation to the levels of abstraction and thus in the hierarchical organization of the template. On the basis of this initial inductive coding, it was agreed that only 1 level of abstraction was necessary and that the overarching themes could be articulated sufficiently by 1 level of subthemes. The resulting comprehensive codebook, which included both deductively and inductively developed codes, descriptions, example quotes, and thematic categories, are detailed in Tables 1-4. All the data were then coded deductively in accordance with the full template by BC and YC, although emergent themes not established in the initial inductive analysis were also proposed. Finally, 2 coders (HB and DC) reviewed the coding template to assess the external validity of the coding process. Where clarity or consistency was questioned, further group discussions were held until all issues were resolved.

Table 1. Comprehensive codebook detailing thematic analysis of open-source automated insulin delivery system user narratives collected from a multinational cohort of 383 participants. This table presents findings related to the physical health benefits and quality of life impact of using open-source automated insulin delivery systems, such as improvements in glycemic control, time in range, and reduced hypoglycemia.

Theme	Description	Example quotes
(A) Physical impact and quality of life impact		
Glycemic outcome improvement (A1)	Refers to improved time in range and HbA_{1c}[a] levels, less glucose variability	• "I purchased the Miaomiao brand Bluetooth transmitter to send my blood glucose levels to my phone, thereby having access to alerts that have undoubtedly saved my life." • "My biggest hope was to control my blood sugar peaks due to gastroparesis and this works 95% of the of time very well. Thanks to this technology, I am a big step closer to my goal." • "He's not T1, but was concerned about my time in range and felt the DIY Loop system was better than approved FDA methods." • "Looping has dramatically improved time in range and how I feel. After 5 weeks Looping, my A_{1C} lowered from 6.8 to 6.3."
Hypo- and hyperglycemia (A2)	Refers to fewer hypo- and hyperglycemia and reduced complications associated	• "Overall, my blood sugar adjustment has improved tremendously. Hypoglycaemia is extremely rare. Times above the limit occur but are limited in duration." • "My values have become immensely better. Hypos with unconsciousness did not occur anymore." • "I've always had problems with hypos. The Freestyle showed me in the morning that I was hypoglycemic, nevertheless, I spent the night—retrospectively—in hypoglycemia. Or even a whole night with levels over 250 mg/dl."
Sleep quality (A3)	Denotes all aspects of improved sleep quality for either caregivers or children, such as increased sleep duration, fewer sleep interruptions, and feeling better rested in the morning	• "Thereby having access to alerts that have undoubtedly saved my life, both while sleeping and while alone in public transportation, among others." • "Only looping for about 2-3 weeks. So far, better quality sleep for my husband and me (no nocturnal hypos which would happen regularly prelooping)." • "I sleep so much better. I no longer fear lows. I understand my body's needs much better." • "It was hard to convince my wife that I was going to turn over control of my diabetes to open source code that I downloaded from the internet. It actually got to a point where I explained to her that I was doing it with or without her approval. When she realized how passionate I was about it, we sat down together and I explained the ins and outs of everything. She sleeps so much better now knowing I'm healthier overall and more safe overnight."
Exercise management (A4)	Refers to the improvements in physical exercise since the use of DIYAPS[b]	• "Better sleep and exercise management, AAPS has very good objectives to work through. Diabetes management is now less hassle and lower risk." • "Exercising and working out is a lot more easier since I no longer have that many low BGs. I overall feel physically and mentally much better since my time in range has increased by 30%."
Quality of life (A5)	Refers to the mentioned improvements in quality of life and describes the degree to which an individual is healthy, comfortable, and able to participate in or enjoy life events	• "But today, at 44, I have a system that already reduces a lot, and ensures very good values and thus a better quality of life." • "Help with stress situations both high and low, preworkout and postworkouts are less drastic, overall better mental health since I don't need to make so many decisions throughout the day about BGs and knowing that I have the ability to have more time in range and a better A_{1C} while still living a pretty normal life." • "A DIY system gives me this flexibility and adaptability. A huge weight has been lifted off my shoulders knowing that I will be safe during the night, and won't be doing undue damage to my organs during stressful situations like exams or during panic episodes." • "Since then, every night was like that and every day I feel more active and fitter...that's what I call Quality of Life."

[a] HbA_{1c}: glycated hemoglobin.

[b] DIYAPS: do-it-yourself artificial pancreas systems.

Table 2. Comprehensive codebook detailing thematic analysis of open-source automated insulin delivery (AID) system user narratives collected from a multinational cohort of 383 participants. This table summarizes user-reported experiences regarding the positive and negative emotional effects of open-source AID systems.

Theme	Description	Example quotes
(B) Emotional impact		
Positive (B1): describes positive emotions of participants related to the experience with open-source AID in daily use, including anticipation, enjoyment, excitement, relief, freedom, and inspiration.		
Anticipation	Describes hopeful emotional states of anticipation and great expectations of participants that lie on the open-source AID system for improved diabetes management and hope for improved quality of life	• “Looking forward to next steps: predicting algorithms.” • “I had high hopes that everything would be different from now and I do not have to do anything for my diabetes anymore, and that the loop regulates everything on its own.” • “Hope and impatience since I was looking for an atomization almost for 20 years.”
Enjoyment	Describes that participants enjoy using the open-source AID system or enjoy the satisfied outcomes brought by open-source AID system use	• “Now I enjoy quiet nights and hope for a future without long term complications.” • “I am doing well and I am happy to use the loop.” • “My CGM alarms are all turned off now and I love the silence.” • “In terms of expectations, I had a vague feeling that closed looping could be a solution to my lack of control during the night but I did not have very clear or measurable expectations. I just wanted to try it, I was curious to see what it was going to bring me. And now, of course, sooooooo glad that I did that!” • “Just awesome. By far the biggest impact has been the reduction in time spent “doing” diabetes. I am rarely woken at night even after intense exercise and little tweaks which previously interrupted my day have largely disappeared. It is a revelation.”
Excitement	Describes exciting emotions of participants related to the experience with open-source AID systems in daily use, including excitement, happiness, and satisfaction with AID system use and associated results	• “I was just so excited to learn about DIY closed looping, and ordered gear immediately.” • “I have been looping for only 4 months now and I am really grateful to all the people that have worked so hard to developed this amazing technology.” • “I was very excited learning about how the system worked. I saw it as a challenge to understand and build it. Understanding the algorithm and building it myself gave me a great understanding of the system and mitigated any fears that I might have had about it’s functioning.” • “To me, loop is a miracle—I can hardly believe it after 35 years to finally have normal blood glucose levels.” • “Just awesome. By far the biggest impact has been the reduction in time spent “doing” diabetes. I am rarely woken at night even after intense exercise and little tweaks which previously interrupted my day have largely disappeared. It is a revelation!” • “[I am] able to read the blood sugar over a watch and no longer have to perform finger pricks, should become true! I still remember the exact moment, I immediately had tears in my eyes.”
Relief	Describes that participants feel relieved since their diabetes-related complications alleviated and quality of life improved after using open-source AID systems	• “I felt well and relieved, within a few days there was a clear improvement in my control.” • “It’s the reduction in disease burden, reduction in lows and highs and the overall improvement in time in range that has changed things so much for me.” • “Since the change, I feel safer and simply more comfortable.” • “My diabetes does not bother me any more, I can accept it and even like it.”
Freedom	Describes the feeling of freedom since participants gained more control of their diabetes by using open-source AID systems	• “The essential feeling of freedom and the feeling of being in control of diabetes.” • “I feel free for the first time in years!”
Inspiration	Describes that participants were encouraged and motivated during their journey of building open-source AID systems	• “No special ‘key events’ other than reading many stories of a diverse group of T1’s who all seemed to overcome all the I-am-new-to-software-building challenges and reading how happy they now we’re that they took this effort. Also, reading questions of people and seeing that they were answered fast and elaborately by more experienced and knowledgeable users made a great impression on me.” • “I was immediately thrilled and the hints that the whole thing wouldn’t be completely legal somehow motivated me even more.” • “I knew that other people could do it so I was convinced I should be able to do it myself as well. I expected to run into issues in the building process, but I wasn’t held back as I knew I would also run into solutions.”

Theme	Description	Example quotes
Negative (B2): describes negative emotions, such as fear, frustration, worry and anxiety of users, mainly due to the concern of not being able to build and maintain open-source AID systems and difficulties encountered.		
Caution	Describes participants' concerns in regard to building and maintaining open-source AID systems	• "I was quite worried about trying it, there were many things I didn't quite understand and found technically challenging" • "I am still a bit afraid to close the loop." • "I have other fears: will my OTG cable have a loose connection again tonight? will I get to an old replacement pump (combo)? We integrate the Dexcom G6 with xdrip or disassembled xdrip the battery - how was it read today? Will the pharmaceutical industry put more obstacles in our way?" • "I have had some fears of system failure, but i do not have them anymore"
Frustration	Describes participants' frustration when experiencing difficulties in sourcing, building, or maintaining open-source AID systems	• "Unfortunately, I first failed because of the technology (availability of the accessories, order from the USA). It was a bit frustrating, but I did not want to give up." • "I also experienced a sadness for how long I had struggled (43years) with inadequate tools to manage my condition." • "My biggest challenge has been the isolation of it all. Online support is always present very timely, but not the same as having someone to be present to help troubleshoot when loop stops. I live in very rural area and sometimes I feel like I'm stranded on an island. Frustrating at times, but I would not trade my cure for anything!!"
Worry and anxiety	Describes that participants worry about the problems that may arise from the use of open-source AID systems	• "as I have no idea what the programme is doing and every time I keep my fingers crossed that I won't see any screen other than in the tutorial because I have no idea how to fix that." • "The biggest hurdle was I was not confident I could build the loop app on my own." • "And I have concerns about the dependence on Bluetooth/Wifi/Internet/ servers." • "I had heard things about DIY closed loop systems in the various cgm related Facebook groups, but had always thought I wouldn't be able to build & maintain one." • "It was something that I considered at the time, but never thought I would be able to do it as I'm not very good with coding/tech." • "suddenly relying on the phone instead of your own mind is a little strange, I often did not trust the loop and delivered boluses on my own or ignored the TBR suggestions." • "A defective libre sensor the other day (already on the 2nd or 3rd day) has given me a lot of restlessness and effort, because the measurements were incorrect every now and then. I was Worried every now and then a single component of the overall system will fail and cause high adjustment effort."

Table 3. Comprehensive codebook detailing thematic analysis of open-source automated insulin delivery (AID) system user narratives collected from a multinational cohort of 383 participants. This table outlines the challenges users face during setup, regular use, and maintenance of open-source AID systems, including technical difficulties and knowledge gaps.

Theme	Description	Example quotes
(C) Challenges		
Accessibility (C1): describes the challenges encountered by the participants regarding accessing the open-source AID systems device, including sourcing hardware, cost, and understanding of the open-source AID systems rationale.		
Sourcing hardware	Describes the problem of sourcing hardware that is compatible with open-source AID systems	• “My main struggle was with losing access to my favorite pump (OmniPod). I switched back to a tube based pump I'm favor of this system, while hoping that the pods are cracked soon. I followed the OpenOmni efforts.” • “but then I was unlikely to do it as I didn’t have the right pump anymore” • “And my health insurance was vague about the info they needed to get the pump and supplies reimbursed. Denied my request several times. Took a lot of communication between health care providers, myself, insurance which was a frustrating process for me.” • “My health care providers initially didn’t want to prescribe the (loopable) pump.”
Cost	Describes the issue of costs associated with building open-source AID systems	• “The main challenge has been self-funding the CGM which is expensive in New Zealand.” • “I started with the pump therapy just before loopin. The health insurance did not supply me with a pump at the time, so I was on my own without diabetic care.” • “The Dana RS pump has the one huge advantage of the open interface to control, otherwise the pump looks like a product from the penultimate decade, compared to other systems. The needles are not great either and the counterclockwise luer lock should be replaced. The should be a standard for pump connections to avoid manufacturer tie-ins and make cost reductions possible (the equipment is still too expensive).”
Understanding	Describes problems with understanding the rationale provided by the instructions for open-source AID system set-up	• “My initial feeling of starting to build my loop was that it is poorly documented and difficult to diagnose setup issues.” • “Understanding the system was both challenging and frustrating. The documentation is poorly explained and the system has its limitations but it is the best we have.” • “In the meantime there is a ton of information about closed loops, which was an intensive learning and reading phase in both the medical literature and the posts and documentation of the community. Since at the beginning the overview was missing, this was also a bit confusing because the meaning of individual components was not yet estimable for the own implementation when reading. What is a Wixel, can it be eaten, and if so, how many carbs do I have to expect?. It had to burst a lot of knots until it was clear which software components to make the (actually quite simple) beginning. But maybe that's just too complicated for me, a certain tendency is not to be denied.” • “In the lead up to building, I felt overwhelmed by all info.” • “Apart from the setup difficulties and my lack of English skills I had no difficulties.”
Setting up (C2): describes the difficulties encountered by the participants when maintaining the use of open-source AID systems, including adjusting and fine-tuning, consumption of time and effort, inconvenience in everyday life and technical issues.		
Adjusting and fine-tuning	Describes the process of determining the factors and further adjusting and fine-tuning	• “Determining the factors with decimal place accuracy, calculating the sensitivity factor was/is in first place effortful.” • “The only difficulty in the change of therapy (apart from the unfamiliar handling of the technique in general) was to adjust the loop for different types of sports, depending on the time of day and physical condition.” • “disillusionment followed immediately - even the updated nightscout / set up is very complicated to operate for non-nerds. It’s often not clear what happens and what is shown on night scout.”
Time and effort	Describes the amount of time and effort required for open-source AID system setup	• “It takes a lot of time and attention to begin with” • “It has been a tricky process, as we have been learning by doing” • “With the technical implementation, it was an uphill battle. It took a while to work but was worth the effort...I do not want to be without loop...”

Theme	Description	Example quotes
Everyday life	Describes inconvenience with using open-source AID systems in daily life regarding exercise and diet, etc	• “I had to reduce the overall carb intake in order to achieve better time in Range, and this had a major impact on my and my family lifestyle” • “The regulation of sports is still a bit difficult” • “I discovered Loop app was terrific, but the phone was too bulky. So I purchased a new pump outright and started AAPS with a tiny phone. I am actually about to try AAPS on a stand-alone watch. The bulk of the gear is super important to me, as I can't be carrying a lot of excess stuff at work” • “I always think about if I need to charge anything (with Enlite it’s just a bit more complicated than with other systems.) I have to keep reminding myself not to go out of the house without my new hand bag - and not without a smartwatch” • “Social judgment from part of the DT1 Community which [accused] me of obsession with the disease.”
Technical challenges	Describes technical issues with the equipment, such as battery charging	• “The issues I’ve had have been of a minor technical issue only, like accidentally shorting out my Miao Miao charger, and having difficulty ordering another. So I changed to Dexcom G5, and have learned to rebattery my transmitters, making it actually cheaper than using Libre. I have also had some troubles with my phone updating it’s OS and becoming useless. I also have some battery issues with my new pump.” • “Biggest challenge was building the app in Android studio, as I have no idea what the programme is doing and every time I keep my fingers crossed that I won't see any screen other than in the tutorial because I have no idea how to fix that.” • “The only difficulties that I am sometimes experiencing are technical problems such as connectivity issues between the Riley link and the pump, and blood sugar fluctuations once the insulin sensitivity changes, which are more noticeable now with a tighter BG management than before, where every day with diabetes was just chaos.”

Table 4. Comprehensive codebook detailing thematic analysis of open-source automated insulin delivery system user narratives collected from a multinational cohort of 383 participants. This table highlights the role of peer support, online communities, and other resources in facilitating the adoption and use of open-source AID systems, emphasizing the sense of community and shared expertise.

Theme	Description	Example quotes
(D) Support		
Family and friends (D1)	Describes that the participants received support from their family or friends	• "My husband supported and encouraged me, which helped but I'd have gone ahead even without that support." • "Friend who had acquired the components and assisted [me] in building. He was already using [AndroidAPS]."
Online support (D2)	Describes the types of support that participants seek on the web, such as social media, blogs, forums, and other people with diabetes	• "First, I found information about AAPS in Freestyle Libre Forum." • "I felt the whole process was very simple with very comprehensive instructions and support through Facebook." • "March 2017 was the first time I started looking for treatment improvements. I acquired all the information and knowledge in through own research on the internet." • "In our area there is a Whatsapp diabetes group that meets in person now and then. A PwD there works in IT and was contacted about 1,5 years ago. With his technical support I have closed my loop." • "I set up a Nightscout server and the AndroidSeries600Upload app for his Medtronic pump in the hospital - and read a lot - and at first set up OpenAPS for myself. As a technophile I was of course immediately on fire, but I still took 2 months to read about the topic online intensively, trawling through forums, Facebook etc., and my anticipation and enthusiasm grew steadily."
DIY[a] community (D3)	Describes that the participants received support from the DIY community, such as help from author DL	• "I got support from the DIY community via face to face meetings and via the online community." • "When I could connect the technical side (IT) with my diabetes, it all started. Then I met Adrian and Dana [authors AT and DL] in person."
Conference and meeting (D4)	Describes the support that the participants received from conferences or meetings	• "However, due to the great help from the looper.de group and the looper meeting, I already wanted much more at this time and have been able to implement this with a lot of reading and informing and with still some technical problems"
Medical professionals (D5)	Describes that the participants received support from medical professionals, such as doctors, nurses, diabetes educators, endocrinologists, etc	• "I actually first heard about DIY options from my doctor, who referred me to another patient who was already using one."
Self-support (D6)	Describes that the participants learn to build DIYAPS[b] all by themselves, with no direct support from others	• "I like the user manual, which it written clearly step by step." • "[N]o direct support or advice from third parties."

[a] DIY: do-it-yourself.

[b] DIYAPS: do-it-yourself artificial pancreas systems.

Ethical Considerations

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Charité–Universitätsmedizin Berlin (EA2/140/18). Informed consent was obtained electronically from all individual participants included in the study. The deidentified datasets are available from the corresponding author upon request. Participation was anonymous and voluntary. No financial or other compensation was provided.

Results

Participant Characteristics

In total, 383 participants (N=722, 53% of participants of the DIWHY survey) responded to the open-ended questions in the survey, and there were a combined 645 responses to the 2 open-ended items. Characteristics and clinical features of the cohort are shown in Table 5.

Table 5. Demographic and self-reported health characteristics of the participants using open-source automated insulin delivery (AID) systems.

Demographics	Values
Gender (n=383), n (%)	
Men	203 (53)
Women	179 (46.7)
Other	1 (0.3)
Age (y), mean (SD)	43 (12)
Type of diabetes (n=383), n (%)	
Type 1	381 (99.5)
Type 2	0 (0)
Other	2 (0.5)
Duration of diabetes (y), mean (SD)	30 (12)
Duration of open-source AID system use (y), mean (SD)	4 (2)
Most recent self-reported HbA_{1c}[a] level, mean (SD)	5.89 (0.62)
Type of open-source AID system used regularly (n=423), n (%)	
OpenAPS	65 (15.4)
AndroidAPS	245 (57.9)
Loop	110 (26)
Other[b]	3 (0.7)
Region (n=383), n (%)	
Europe	282 (73.6)
Germany	184 (48)
United Kingdom	41 (10.7)
Austria	7 (1.8)
Spain	7 (1.8)
Netherlands	4 (1)
Finland	6 (1.6)
Czech Republic	5 (1.3)
Bulgaria	4 (1)
Sweden	2 (0.5)
Other[c]	22 (5.7)
North America	69 (18)
Canada	15 (3.9)
United States	54 (14.1)
Western Pacific	23 (6)
Australia	12 (3.1)
New Zealand	11 (2.9)
Asia	2 (0.5)
South Korea	2 (0.5)
Africa	2 (0.5)
Algeria	1 (0.3)
South Africa	1 (0.3)
I'd rather not say	5 (1.3)

Demographics	Values
Education (n=379), n (%)	
Doctorate or graduate degree	154 (40.6)
Bachelors, professional or associate degree	129 (34)
Trade, technical or vocational training	27 (7.1)
Some college credits	11 (2.9)
Secondary school	34 (9)
Some secondary or primary school	22 (5.8)
No schooling completed or none of the above	2 (0.5)
Occupational status (n=382), n (%)	
Full time	265 (69.4)
Part time	60 (15.7)
Unemployed	4 (1.1)
Retired	20 (5.2)
Student	24 (6.3)
Other or none of the above	9 (2.4)
Professional background (n=305), n (%)	
Medicine	76 (24.9)
Tech	81 (26.6)
Finance	40 (13.1)
Other	108 (35.4)
Household annual income (US $; n=336), n (%)	
<20,000	34 (10.1)
24,000-34,999	27 (8)
35,000-49,999	48 (14.3)
50,000-74,999	87 (25.9)
75,000-99,999	44 (13.1)
>100,000	79 (23.5)
I’d rather not say	17 (5.1)

[a] HbA_{1c}: glycated hemoglobin.

[b] xDrip, Nightscout, offline uploader for Medtronic 600 series, Hackabetes Artificial Pancreas Project, custom, or own developments.

[c] Belgium, Croatia, France, Hungary, Ireland, Italy, Lithuania, Poland, Romania, Russia, Serbia, Slovakia, and Switzerland.

Template Analysis

Emotional and Quality of Life Impact

“Anticipation” and “curiosity” were emotions mentioned by participants in relation to their first encounter with open-source AID technology. This highlights the intuitive appeal of this solution for diabetes management—“I had envisioned this type of solution for many years and was looking out for the emergence of suitable technology” (Man, age 59 years, United Kingdom)—and why, for many people, initial reservations were quelled by the potential improvements offered by AID systems—“One is a little uncertain, but the curiosity for the improvement of control has won!” (Gender unspecified, age 49 years, Germany).

However, as this sentiment indicates, anxiety and caution were also a part of the emotional responses experienced by our participants. This could be as they confronted the prospect of developing their own system—“I was quite worried about trying it, there were many things I didn’t quite understand and found technically challenging” (Woman, age 69 years, United Kingdom)—but was also apparent even once the system had been successfully built—“Originally it felt like a big step to let the algorithm make changes” (Woman, age 42 years, Australia).

Thus, both the challenges—real and anticipated—in setting up the system and the prospect of allowing an algorithm to undertake a life-critical role could have a negative emotional impact.

However, for the most part, initial concerns about the complexity of the technology and ceding control to an algorithm were replaced by a sense of "pride" and "relief"—"I feel very good and proud I did it because it was technically difficult to build it with my pump and CGM" (Man, age 50 years, Germany).

The sense of relief experienced by our participants was 2-fold; relief that the system was built and functioning but also a sense of being partially released from the burden of everyday diabetes management—"The most impressive thing is how little diabetes suddenly plays a role, how simple everything has become, how rarely one suddenly has to wonder about metabolic fluctuations, how well one can sleep, knowing that blood sugar stays in range" (Woman, age 49 years, Germany).

Twisting the concern with automation, some participants also noted that it was precisely because control was given over to an algorithm that improved outcomes could be achieved—"I was happy to hand over control to something which makes fewer irrational decisions and is less emotionally involved in the process" (Woman, age 35 years, United Kingdom).

The relief they experienced did not come without considerable effort and "frustration," and this was also a common emotion in the narratives. Part of this frustration was related to the reliability of technology:

> *While the burden of what to do in reaction to blood glucose has gone down, the tech troubleshooting and figuring out how to fine-tune has increased greatly. Traded one problem for another.* [Woman, age 62 years, United States]

As can be inferred from this comment, frustration was also driven by an expectation that the level of automation would be greater: "A few months into closed looping I am starting to see results, though I was expecting [it] to be easier and thought it was going to handle much more the ups and downs by itself" (Man, age 41 years, Netherlands).

Yet, most participants declared themselves happy to invest the effort when the reward was so tangible and transformative for overall quality of life:

> *...it doesn't just fix everything and that there are still settings to adjust and check, but once these were okay then I've had very few issues. It has allowed me to take a back seat with my diabetes care...It has taken huge amounts of the diabetes burden away from me!* [Woman, age 25 years, United Kingdom]

In fact, the work also served as a source of inspiration, with many participants gaining new insights into important factors influencing glucose fluctuations:

> *Looping has provided me much detailed insight into the inner-workings of my endocrine system and diabetes management. I've learned that my insulin ratios and [basal rates] needed to be greatly adjusted. As I've learned, two bad settings can mask each other and end up appearing to be correct. The learning curve is steep, but very rewarding.* [Woman, age 24 years, France]

Overall, participants indicated that the net gains of open-source AID were extensive and profound. Often, a sense of gratitude was expressed:

> *The community has helped me so much. I can't express my gratitude to all developers, helpers and people in my local community as well who freely give their time and skills to make this possible.* [Woman, age 65 years, Australia]

Source of Support

The community mentioned earlier highlights the particular model of diffusion that has fostered the use of open-source AID systems. To echo a common refrain in this context, *do-it-yourself* does not mean *do-it-alone* [15], so while each user is ultimately responsible for building their own system, the support that they can obtain in doing so is potentially extensive.

For those without preexisting skills in IT, support was at hand, for some, among one's established social network, for example, family and friends:

> *I was very intimidated at first as I have extremely limited coding knowledge. After following along in the group for a while, I began to get more comfortable. My boyfriend also encouraged me and offered to help set it up since he has a bit more tech knowledge.* [Woman, age 22 years, Canada]

Both technical and medical expertise within the individual's network was an important antecedent to the uptake of open-source AID systems:

> *My partner immediately supported me because, as a doctor, it was very clear to him that it had a much better metabolic effect...The support of the Facebook group especially for small logistic things was very important to me.* [Woman, age 37 years, Germany]

This participant's comments also highlighted that direct support from health care professionals (HCPs) may be lacking:

> *The support by my social environment has increased, the support by doctors and the diabetologist's office are completely lost, I consider this to be a risk, I am waiting for the moment when our diabetologists will not only be "not forbidden" but required to inform about the Closed Loop as the gold standard of therapy.* [Woman, age 37 years, Germany]

This is not to say that HCPs were not supportive or positive about AID outcomes, but the support they could provide was more often informal and emotional rather than practical:

> *My two diabetologists know about the loop and are amazed/enthusiastic about my [glucose] levels, but unfortunately cannot support me for legal reasons.* [Woman, age 55 years, Germany]

While an individual's social relations could profoundly impact the building and maintaining an open-source AID system, support was still available even without direct expertise in one's personal network—"I've found the technology almost impossibly difficult to deal with and have had a considerable amount of personal help from other users" (Woman, age 62 years, United Kingdom).

For some users, this occurred at face-to-face meetings (eg, "build events"), where expert users could guide them, but, for the majority, such support was obtained via online forums—"The biggest (and for me only) help with technical problems or 'fine-tuning' the settings is provided by the Looper online community" (Woman, age 27 years, Germany).

The #WeAreNotWaiting community was the main source of support cited by our participants, and this was multifaceted. In its most basic form, the online documentation developed by users for users was an essential resource and frequently praised for its clarity. Beyond this, in the various social media–based groups connected with open-source AID systems, there was also a wealth of information from reading existing threads and others' posts, where frequently asked questions and troubleshooting topics could guide through challenging aspects of the process. Finally, the online forums also served as a real-time support network, where users could expect rapid and reliable responses to whatever issue they might reach out for:

> *I don't want to finish without mentioning the importance that the support groups are having to me. Both in helping understand and setting the system and managing the everyday life... It is completely amazing being able to be connected to so many people who are also looping and give and get support.* [Woman, age 40 years, Spain]

Also notable was how the encounter with this community and its essentially altruistic spirit could inspire new users to be willing to participate and serve within this support network:

> *What one cannot do, the many can manage. The group has helped me. I'm getting involved as well and spread the knowledge so others can benefit from it.* [Man, age 54 years, Germany]

Challenges

Principally, the challenges reported included (1) accessibility and (2) technical setup and maintenance. There were 3 prevailing issues regarding accessibility. The first one was cost, since essential hardware was not readily available via health care services or insurers:

> *CGM is prohibitively expensive in my country. I only started using it two months before looping as part of preparing to loop. I am trying to hang in there paying for it because of the fantastic benefits but it is a big drain on family income.* [Woman, age 51 years, Australia]

In addition, even in circumstances where hardware was potentially available to users via public health care or insurance, access could still be problematic if potential users were not eligible according to insurers, or HCPs had reservations about recommending devices that could be used for open-source AID systems:

> *My healthcare providers initially didn't want to prescribe the (loopable) pump. And my health insurance was vague about the info they needed to get the pump and supplies reimbursed. Denied my request several times. Took a lot of communication between health care providers, myself, insurance which was a frustrating process for me.* [Woman, age 33 years, country of residence undisclosed]

The final aspect of accessibility was the anticipated complexity of the process and the documentation to be followed to set up an open-source AID system—"I had heard things about DIY closed-loop systems in the various CGM-related Facebook groups, but had always thought I wouldn't be able to build & maintain one" [Woman, age 47 years, United Kingdom].

Thus, the obstacle was sometimes more about expectation than experience, with people deterred from the attempt by the expectation that they would not be able to solve technical issues.

However, for others, it was as much the experience as the expectation that could provide an obstacle to access—"There was a lot to learn. I often sat crying in front of the computer" (Woman, age 35 years, Germany).

Technical challenges associated with building open-source AID systems were prevalent in the narratives, with only a handful indicating that the process was straightforward:

> *I tried about a year before I actually started to build a system and it proved too difficult. After a year of being burned out and things not being any better, I tried again and succeeded.* [Woman, age 36 years, United States]

Participants conveyed difficulties with both hardware and software components, for example, with connectivity loss.

Even with all components fully functioning and connected, other technical challenges remained, though these were more related to the technicalities of diabetes management than technology per se. Users of open-source AID systems take on the role of diabetes experts as much as programmers and are required to fine-tune the settings on their devices in accordance with a selection of parameters:

> *[The] first weeks of looping were a bit hard because my ratios were off and it was hard to understand why Loop is making some decisions. Or, what's even more important, which parameter should be tweaked in order to make it behave better.* [Man, age 32 years, Poland]

Determination of these settings is generally undertaken by HCPs in the context of prescribed, commercial devices, ideally in collaboration with the user, but it may remain more or less opaque to the individual with diabetes. So, although many users were highly engaged with diabetes management previously, there was a learning curve involved for most, not least because the level of control for different parameters allowed by open-source AID systems extends significantly beyond those in commercial systems:

> *So I have to work more on my settings. Nothing is (fully) automatic and runs all by itself. For me, as a technician who believes in the possibilities of self-regulating automation, there is still a lot to be done.* [Man, age 51 years, Germany]

The combination of the different challenges involved in building an open-source AID system evoked another issue for some, inasmuch that considerable time was required to resolve the issues emerging from building and maintaining the system—“My husband has suggested several times that I was doing more work with the system than without, because of the frustration & time (whole weekends) involved in getting my loop back up & running” [Woman, age 51 years, United States].

Yet, while time was undoubtedly a factor to be dealt with, the extent to which it was perceived as a challenge, impinging on everyday life, was often weighed against the time spent “doing diabetes” before transitioning to open-source AID systems:

> *My own personal tight control prior to looping was very time-consuming. Post APS I save more than 1hr every day. Imagine my experience of living 1/24th longer life for the rest of my life because of APS.* [Man, age 46 years, New Zealand]

Physical Impact

For the most part, participants reported marked improvements in physical health in accordance with the measures traditionally used to gauge this in T1D, such as HbA_{1c} and time in range, as reported elsewhere [15]. For some, these improvements occurred soon after their switch to open-source AID systems:

> *Benefits from the first 8 weeks: 80-90% of the time-in-range without changing my lifestyle! Previously that value was 40-45%.* [Man, age 37 years, Germany]

The time regarding changes in HbA_{1c} was longer but still commented upon, often as levels within a reference range for people without diabetes:

> *I knew some that tuning was needed but I was patient. Now I have used a DIY system 24/7 approx 23 months for almost two years! Hba_{1C} is 5.2%, I’m happy.* [Man, age 44 years, Finland]

In addition to these clinical outcome improvements, participants also reported changes in their health based on more immediate, everyday experiences. The experience of hypoglycemia was something alluded to extensively—“The blood glucose fluctuations and the hypos have become much less, I feel much safer and I am doing things again that I used to avoid” (Woman, age 45 years, Germany).

The sense of safety can have a profound effect on an individual’s life. Both hyper- and hypoglycemia in their moderate expression can induce physical symptoms that are unpleasant and disruptive but, in their extreme extent, are potentially fatal. Ameliorating the risk of highs and lows, open-source AID systems served to diminish the unpleasant symptoms and, at the same time, reduce the anxiety attached to what might happen in worst-case scenarios. This, aside from its direct benefits, also had follow-on effects on other health-related aspects. Many participants noted benefits of open-source AID systems related to physical activity—“I’m much less afraid of unplanned physical activity because the loop usually regulates it with a few extra carbs” (Woman, age 55 years, Germany).

Again, the point here is not only that glucose levels are within range during exercise but also that the potential fear around exercising was lessened. Fear of exercise and its unpredictable impact on glucose levels represents a clear obstacle removed by open-source AID systems, with potential general health benefits—“I have recently started exercising again after years of sedentary living” (Woman, age 31 years, Australia).

By far, the most frequently mentioned quality of life improvement among our participants was sleep duration and quality:

> *I can sleep and have no alarms from the CGM at night. In the morning I wake up with a value that I do not have to correct. This has a positive effect on the blood sugar during the day.* [Woman, age 38 years, Germany]

Persistently disturbed sleep is by any reckoning something that one would expect to impact health and overall quality of life, but for people living with T1D, it is a given:

> *I SLEEP. That’s the most brilliant, life-changing thing. I’d been sleep-deprived for so long I didn’t even realize what a difference it would make.* [Woman, age 49 years, United States]

In a similar vein to exercise, open-source AID systems had a dual impact in relation to sleep. It served to alleviate symptoms of hypo- and hyperglycemia that could disturb sleep directly or trigger alarms on devices waking people up. Simultaneously, it helped reduce the fear of nocturnal hypoglycemia, which could result in sleep difficulties due to anxiety and adverse aversion strategies, such as intentionally aiming for higher glucose levels before sleep:

> *I have no anxiety about sleeping alone when my wife is working away from home. I actually sleep through the night. Eating out is no longer a major stress since I know that even if I underestimate carbs [it] will fix my errors overnight and I will wake mostly in range.* [Man, age 42 years, United Kingdom]

As might be expected, improved sleep was also associated with further physical and mental health benefits:

> *Waking up that first morning in a normal range, and every morning thereafter was amazing. It’s incredible how much more you can get done in a day when you wake up in a normal blood glucose range.* [Man, age 39 years, United States]

Part of this is obviously about having more energy because of being well-rested. Beyond that, there is the important difference of waking up with glucose in target range and how this resonates through the rest of the day:

> *The almost fully automatic delivery of needed insulin has made life a lot easier and once the factors are set correctly, it is almost possible to live like a ‘healthy’ person. It’s also much easier to start a new day, starting at a value of 90 mg/dl and not 180-200 as before! Working days are much easier than before!* [Man, age 47 years, Germany]

Rather than using considerable time, energy, and resources attempting to reestablish balance in one's glucose levels, open-source AID systems allowed people to concentrate more on the business of the day. So, whereas previously a working day may have felt more like "running the gauntlet," a different experience and outlook on life could be fostered:

> *Since I've used [open-source AID], I was upgraded on my job, I'm mentally faster and sleep like a baby without alarms. I've started several personal projects and [I'm] currently on professional certification. I have plenty of quality time now without hypo or hyper and finally happy.* [Woman, age 40 years, Spain]

Discussion

Principal Findings and Their Implications

The findings we present here concerning the impact of open-source AID systems highlight their immense benefits from the perspective of the user, simultaneously setting the extremely challenging nature of diabetes and the ways it may compromise quality of life into relief. This is the first study to analyze narratives and to examine the emotional and physical health impact of open-source AID systems in adult users. Our findings are in line with our analysis of children and adolescents using open-source AID systems, including their caregivers [30], although there were age-specific findings, for example, navigating diabetes throughout puberty and remote monitoring and control by caregivers that only applied to the child cohort. Our results also align with other, smaller cohort studies examining the user experience with open-source AID systems [22,24,31] with the literature pointing to the importance of setting expectations for both onboarding and the ongoing use of AID systems [32,33]. Studies of users of commercial AID systems found similar results [34,35]. Furthermore, the sense of community and empowerment, often referred to as "paying it forward," was almost exclusively described in open-source AID system users.

In our approach to analyzing the data, we opted to use 4 categories as the basic framework for our template. This was necessary, first and foremost, because these topics were already framed in the wording of the questions. That said, the findings also underline the somewhat fluid nature of the categories, especially with respect to the physical and emotional impact that open-source AID systems can have. So, while we have sought to tease emotional and physical impacts apart for the purposes of our analysis, our findings ultimately serve to highlight how inexorably bound up they are. This is most intuitively illustrated through the example of sleep, where poor quality sleep inevitably impacts emotional well-being, which may, in turn, also impact glucose levels, both directly and indirectly [36].

Peer Support as a Driver of Innovation and Empowerment

Although there is not the same level of symbiosis between the categories "challenges" and "support," the findings did point to a strong link between the two in the sense that many of the challenges associated with initiating and maintaining the use of an open-source AID system were resolved via support from a wider community of users. The sense of community underpinning the development and diffusion of open-source AID systems and peer support as a key resource for practical but also emotional support were predominant topics in other qualitative studies on the lived experience with open-source AID systems [22,31]. For many, the discovery of a peer group that one could identify and engage with was as important and meaningful as building their AID system.

As has previously been noted [37], digital platforms can provide opportunities for peer support and the exchange of experiential knowledge about living with illness. The importance of peer support for people with diabetes in the context of online communities has been clearly highlighted elsewhere [12,22,23,38,39]. Moreover, engagement in these communities has been shown to positively impact HbA_{1c} levels [23,40]; reduce diabetes-related distress [41]; and foster support and connection, advocacy, self-expression, information and education, technical support, and humor as a coping strategy [38,42].

However, it is also evident that the type of peer support upon which the dissemination of open-source AID systems has been based is of a somewhat different order. In part, this reflects something of the nature of T1D, where the prevailing model of care, inasmuch as it requires individuals to be actively engaged in their care, may potentially foster the growth and dissemination of expertise [43]. In this situation, the delineation of expertise into "professional" and "laypeople" seems outmoded and evokes the well-known Shavian aphorism that the former serves the purpose of conspiring against the latter [44].

As AID algorithms are being constantly developed further and new features will be introduced (eg, fully closed-loop systems without bolusing for meals), future research should address the lived experiences of people with diabetes associated with their use in addition to the analysis of clinical outcomes. Their full health impact can only be evaluated if real-world user experiences are included.

Looking back at how innovations in diabetes treatment were perceived by HCPs and the scientific community over the last century, increased autonomy and empowerment of people with diabetes have continuously been regarded with skepticism. Similarly, it was debated if people with diabetes were capable of blood glucose self-monitoring in the 1970s [45], calculating their insulin doses by themselves, or understanding real-time readings of their CGM device [46]—all aspects that are standard of care today. In the IT age, people with diabetes creating their own technological tools might be just another iteration of patient empowerment but accompanied by similar controversy. The "lesson learned" from these controversies should be the urgency to foster collaboration with patients and involve them early as stakeholders—whether in research and development of medical devices or the development of care concepts that will ultimately affect them.

This study is the first large-scale qualitative study assessing the lived experiences of adult open-source AID system users. Moreover, it is a study with a truly multinational scope, and in its stakeholder engagement via the involvement of the #WeAreNotWaiting community, it strives to remain true to the values of the phenomenon it is investigating.

Of course it should be noted that the survey was undertaken in 2019, and thus the participants can be considered early adopters of open-source AID systems. At one level, this means that the size of our sample represents what, at that time, was a significant proportion of all users of open-source AID systems. However, by the same token, given the subsequent dynamic technical development and rapid expansion of open-source AID system users since that time, the experiences captured in our study may not be reflective of later experiences and current use of open-source AID systems. In addition, many position papers and an international consensus paper have been published to provide guidance to HCPs who wish to support people using open-source AID systems, which may have contributed to increased knowledge and a change of attitudes [47,48]. A selection bias may be present with the survey only being available in German and English.

Conclusions

The efforts of the #WeAreNotWaiting community are changing the landscape of available treatment options and the way we look at the role of patients as initiators rather than as passive recipients of a health care product or service. The online communities that support this movement have not only transformed diabetes care with its technology but have also eased the individual burden for those involved due to the tools and peer support it has fostered. The extensive testimony from users of open-source AID systems acquired in this study provides new insights, highlighting factors inspiring people to adopt such solutions, user experiences in transitioning to open-source AID systems, and the transformative impact of AID systems on the everyday life of people with diabetes. These results may contribute to a better understanding of their unmet needs; the impact of AID systems on physical and emotional health; and some of the current challenges to the uptake of AID technology in terms of access, availability, and usability.

Acknowledgments

This study is part of the OPEN project. OPEN is funded by the European Commission’s Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie Action Research and Innovation Staff Exchange (grant 823902), the Deutsche Forschungsgemeinschaft (DFG)-funded Digital Clinician Scientist Program of the Berlin Institute of Health, and the SPOKES Wellcome Trust Translational Partnership Program.

Data Availability

The datasets generated and analyzed during this study are available from the OPEN project on request. Researchers interested in working with the OPEN data are encouraged to submit inquiries for data access to the project coordinator (SO) via email. The authors welcome the opportunity to collaborate and accommodate research needs where possible.

Authors' Contributions

BC, KB, SO, AT, KR, DL, and TS designed the study. KB and KR were responsible for administration and ethics approval. KB collected the data. KB, MW, and TS processed the data. BC, YC, MW, HB, DC, and SO analyzed the data. BC, YC, and KB drafted the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
DIWHY Questionnaire for Adults.
[PDF File (Adobe PDF File), 85 KB - jmir_v27i1e48406_app1.pdf]

References

1. Hossain SN, Jaglal SB, Shepherd J, Perrier L, Tomasone JR, Sweet SN, et al. Web-based peer support interventions for adults living with chronic conditions: scoping review. JMIR Rehabil Assist Technol 2021 May 25;8(2):e14321 [FREE Full text] [doi: 10.2196/14321] [Medline: 34032572]
2. Banting FG, Best CH, Collip JB, Campbell WR, Fletcher AA. Pancreatic extracts in the treatment of diabetes mellitus: preliminary report. Can Med Assoc J 1922 Mar;12(3):141-146 [FREE Full text] [Medline: 20314060]
3. Lewis D. History and perspective on DIY closed looping. J Diabetes Sci Technol 2019 Jul;13(4):790-793 [FREE Full text] [doi: 10.1177/1932296818808307] [Medline: 30348013]
4. Lewis D. #WeAreNotWaiting to reduce the burden of Type 1 diabetes. OpenAPS.org. URL: https://openaps.org/ [accessed 2023-04-21]
5. Welcome to the AAPS documentation. Android APS. URL: https://androidaps.readthedocs.io/ [accessed 2023-04-21]
6. Artificial-Pancreas / iAPS. GitHub. URL: https://github.com/Artificial-Pancreas/iAPS [accessed 2023-04-21]

7. What is the Loop app? LoopDocs. URL: https://loopkit.github.io/loopdocs/ [accessed 2023-04-21]
8. Peyser T, Dassau E, Breton M, Skyler JS. The artificial pancreas: current status and future prospects in the management of diabetes. Ann N Y Acad Sci 2014 Apr;1311:102-123. [doi: 10.1111/nyas.12431] [Medline: 24725149]
9. Fang Z, Liu M, Tao J, Li C, Zou F, Zhang W. Efficacy and safety of closed-loop insulin delivery versus sensor-augmented pump in the treatment of adults with type 1 diabetes: a systematic review and meta-analysis of randomized-controlled trials. J Endocrinol Invest 2022 Mar;45(3):471-481. [doi: 10.1007/s40618-021-01674-6] [Medline: 34535888]
10. Farrington C. Psychosocial impacts of hybrid closed-loop systems in the management of diabetes: a review. Diabet Med 2018 Apr;35(4):436-449. [doi: 10.1111/dme.13567] [Medline: 29247547]
11. Pfiester E, Braune K, Thieffry A, Ballhausen H, Gajewska KA, O'Donnell S. Costs and underuse of insulin and diabetes supplies: findings from the 2020 T1International cross-sectional web-based survey. Diabetes Res Clin Pract 2021 Sep;179:108996 [FREE Full text] [doi: 10.1016/j.diabres.2021.108996] [Medline: 34363862]
12. Crocket H. Peer mentoring in the do-it-yourself artificial pancreas system community. J Diabetes Sci Technol 2020 Nov;14(6):1022-1027 [FREE Full text] [doi: 10.1177/1932296819883876] [Medline: 31646890]
13. Burnside MJ, Lewis DM, Crocket H, Meier R, Williman J, Sanders OJ, et al. Open-source automated insulin delivery in type 1 diabetes. N Engl J Med 2022 Sep 08;387(10):869-881. [doi: 10.1056/NEJMoa2203913] [Medline: 36069869]
14. Braune K, O'Donnell S, Cleal B, Lewis D, Tappe A, Willaing I, et al. Real-world use of do-it-yourself artificial pancreas systems in children and adolescents with type 1 diabetes: online survey and analysis of self-reported clinical outcomes. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e14087 [FREE Full text] [doi: 10.2196/14087] [Medline: 31364599]
15. Braune K, Gajewska KA, Thieffry A, Lewis DM, Froment T, O'Donnell S, et al. Why #WeAreNotWaiting-motivations and self-reported outcomes among users of open-source automated insulin delivery systems: multinational survey. J Med Internet Res 2021 Jun 07;23(6):e25409 [FREE Full text] [doi: 10.2196/25409] [Medline: 34096874]
16. Toffanin C, Kozak M, Sumnik Z, Cobelli C, Petruzelkova L. In silico trials of an open-source android-based artificial pancreas: a new paradigm to test safety and efficacy of do-it-yourself systems. Diabetes Technol Ther 2020 Feb;22(2):112-120. [doi: 10.1089/dia.2019.0375] [Medline: 31769699]
17. Melmer A, Züger T, Lewis DM, Leibrand S, Stettler C, Laimer M. Glycaemic control in individuals with type 1 diabetes using an open source artificial pancreas system (OpenAPS). Diabetes Obes Metab 2019 Oct;21(10):2333-2337. [doi: 10.1111/dom.13810] [Medline: 31183929]
18. Lum JW, Bailey RJ, Barnes-Lomen V, Naranjo D, Hood KK, Lal RA, et al. A real-world prospective study of the safety and effectiveness of the loop open source automated insulin delivery system. Diabetes Technol Ther 2021 May;23(5):367-375 [FREE Full text] [doi: 10.1089/dia.2020.0535] [Medline: 33226840]
19. Gawrecki A, Zozulinska-Ziolkiewicz D, Michalak MA, Adamska A, Michalak M, Frackowiak U, et al. Safety and glycemic outcomes of do-it-yourself AndroidAPS hybrid closed-loop system in adults with type 1 diabetes. PLoS One 2021;16(4):e0248965 [FREE Full text] [doi: 10.1371/journal.pone.0248965] [Medline: 33819289]
20. Petruzelkova L, Jiranova P, Soupal J, Kozak M, Plachy L, Neuman V, et al. Pre-school and school-aged children benefit from the switch from a sensor-augmented pump to an AndroidAPS hybrid closed loop: a retrospective analysis. Pediatr Diabetes 2021 Jun;22(4):594-604. [doi: 10.1111/pedi.13190] [Medline: 33576551]
21. Hng T, Burren D. Appearance of Do-It-Yourself closed-loop systems to manage type 1 diabetes. Intern Med J 2018 Nov;48(11):1400-1404. [doi: 10.1111/imj.14105] [Medline: 30387315]
22. Schipp J, Skinner T, Holloway E, Scibilia R, Langstrup H, Speight J, et al. How adults with type 1 diabetes are navigating the challenges of open-source artificial pancreas systems: a qualitative study. Diabetes Technol Ther 2021 Aug;23(8):546-554. [doi: 10.1089/dia.2020.0652] [Medline: 33720767]
23. Litchman ML, Lewis D, Kelly LA, Gee PM. Twitter analysis of #OpenAPS DIY artificial pancreas technology use suggests improved A1C and quality of life. J Diabetes Sci Technol 2019 Mar;13(2):164-170 [FREE Full text] [doi: 10.1177/1932296818795705] [Medline: 30198751]
24. Marshall DC, Holloway M, Korer M, Woodman J, Brackenridge A, Hussain S. Do-It-Yourself artificial pancreas systems in type 1 diabetes: perspectives of two adult users, a caregiver and three physicians. Diabetes Ther 2019 Oct;10(5):1553-1564. [doi: 10.1007/s13300-019-00679-y] [Medline: 31440989]
25. Vallis M, Holt RI. User-driven open-source artificial pancreas systems and patient-reported outcomes: a missed opportunity? Diabet Med 2022 May;39(5):e14797. [doi: 10.1111/dme.14797] [Medline: 35092089]
26. O'Donnell S, Lewis D, Marchante Fernández M, Wäldchen M, Cleal B, Skinner T, et al. Evidence on user-led innovation in diabetes technology (the OPEN project): protocol for a mixed methods study. JMIR Res Protoc 2019 Nov 19;8(11):e15368 [FREE Full text] [doi: 10.2196/15368] [Medline: 31742563]
27. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
28. 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]
29. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psychol 2015 Apr 03;12(2):202-222 [FREE Full text] [doi: 10.1080/14780887.2014.955224] [Medline: 27499705]

30. Braune K, Krug N, Knoll C, Ballhausen H, Thieffry A, Chen Y, et al. Emotional and physical health impact in children and adolescents and their caregivers using open-source automated insulin delivery: qualitative analysis of lived experiences. J Med Internet Res 2022 Jul 14;24(7):e37120 [FREE Full text] [doi: 10.2196/37120] [Medline: 35834298]
31. Suttiratana SC, Wong JJ, Lanning MS, Dunlap AR, Hanes SJ, Hood KK, et al. Qualitative study of user experiences with loop, an open-source automated insulin delivery system. Diabetes Technol Ther 2022 Jun;24(6):416-423 [FREE Full text] [doi: 10.1089/dia.2021.0485] [Medline: 35099278]
32. Lewis D. Automated Insulin Delivery in Real Life (AID-IRL): real-world user perspectives on commercial AID. J Diabetes Sci Technol 2022 Mar;16(2):500-503 [FREE Full text] [doi: 10.1177/1932296820957360] [Medline: 32935554]
33. Lewis D. Setting expectations for successful artificial pancreas/hybrid closed loop/automated insulin delivery adoption. J Diabetes Sci Technol 2018 Mar;12(2):533-534 [FREE Full text] [doi: 10.1177/1932296817730083] [Medline: 28918653]
34. Hendrieckx C, Poole LA, Sharifi A, Jayawardene D, Loh MM, Horsburgh JC, et al. "It is definitely a game changer": a qualitative study of experiences with in-home overnight closed-loop technology among adults with type 1 diabetes. Diabetes Technol Ther 2017 Jul;19(7):410-416. [doi: 10.1089/dia.2017.0007] [Medline: 28537437]
35. Barnard KD, Wysocki T, Thabit H, Evans ML, Amiel S, Heller S, Angela Consortium. Psychosocial aspects of closed- and open-loop insulin delivery: closing the loop in adults with Type 1 diabetes in the home setting. Diabet Med 2015 May;32(5):601-608. [doi: 10.1111/dme.12706] [Medline: 25615888]
36. Nefs GM, Bazelmans E, Donga E, Tack CJ, de Galan BE. Sweet dreams or bitter nightmare: a narrative review of 25 years of research on the role of sleep in diabetes and the contributions of behavioural science. Diabet Med 2020 Mar;37(3):418-426. [doi: 10.1111/dme.14211] [Medline: 31833083]
37. Kingod N, Cleal B, Wahlberg A, Husted GR. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res 2017 Jan;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]
38. Tenderich A, Tenderich B, Barton T, Richards SE. What are PWDs (people with diabetes) doing online? A netnographic analysis. J Diabetes Sci Technol 2019 Mar;13(2):187-197 [FREE Full text] [doi: 10.1177/1932296818813192] [Medline: 30477335]
39. Reidy C, Klonoff DC, Barnard-Kelly KD. Supporting good intentions with good evidence: how to increase the benefits of diabetes social media. J Diabetes Sci Technol 2019 Sep;13(5):974-978 [FREE Full text] [doi: 10.1177/1932296819850187] [Medline: 31096768]
40. Litchman ML, Edelman LS, Donaldson GW. Effect of diabetes online community engagement on health indicators: cross-sectional study. JMIR Diabetes 2018 Apr 24;3(2):e8 [FREE Full text] [doi: 10.2196/diabetes.8603] [Medline: 30291079]
41. de Vries L, van der Heijden AA, van 't Riet E, Baan CA, Kostense PJ, Rijken M, et al. Peer support to decrease diabetes-related distress in patients with type 2 diabetes mellitus: design of a randomised controlled trial. BMC Endocr Disord 2014 Mar 04;14:21 [FREE Full text] [doi: 10.1186/1472-6823-14-21] [Medline: 24593296]
42. Hilliard ME, McQuaid EL, Nabors L, Hood KK. Resilience in youth and families living with pediatric health and developmental conditions: introduction to the special issue on resilience. J Pediatr Psychol 2015 Oct;40(9):835-839 [FREE Full text] [doi: 10.1093/jpepsy/jsv072] [Medline: 26275974]
43. Gottlieb SD, Cluck J. "Going rogue": re-coding resistance with type 1 diabetes. Digit Cult Soc 2018;4(2):137-156 [FREE Full text] [doi: 10.14361/dcs-2018-0208]
44. Ashton JR. All professions are conspiracies against the laity? (George Bernard-Shaw). J Epidemiol Community Health 2003:161. [doi: 10.1136/jech.57.3.161-a]
45. Bernstein RK. Blood glucose self-monitoring by diabetic patients: refinements of procedural technique. Diabetes Care 1979;2(2):233-236. [doi: 10.2337/diacare.2.2.233] [Medline: 520127]
46. De Block C, Manuel-y-Keenoy B, Van Gaal L. A review of current evidence with continuous glucose monitoring in patients with diabetes. J Diabetes Sci Technol 2008 Jul;2(4):718-727 [FREE Full text] [doi: 10.1177/193229680800200426] [Medline: 19885251]
47. Braune K, Lal RA, Petruželková L, Scheiner G, Winterdijk P, Schmidt S, OPEN International Healthcare Professional Network and OPEN Legal Advisory Group. Open-source automated insulin delivery: international consensus statement and practical guidance for health-care professionals. Lancet Diabetes Endocrinol 2022 Jan;10(1):58-74 [FREE Full text] [doi: 10.1016/S2213-8587(21)00267-9] [Medline: 34785000]
48. Gawrecki A, Klupa T, Araszkiewicz A, Matejko B, Szadkowska A, Wolnik B, et al. Utilization of do-it-yourself artificial pancreas systems in the management of patients with type 1 diabetes: a position statement of the Pump School Education Initiative by Diabetes Poland. Pol Arch Intern Med 2019 Feb 28;129(2):141-142 [FREE Full text] [doi: 10.20452/pamw.4468] [Medline: 30816103]

Abbreviations

AID: automated insulin delivery
CGM: continuous glucose monitor

HbA_{1C}: glycated hemoglobin
HCP: health care professional
OPEN: Outcomes of Patient’s Evidence with Novel, Do-it-Yourself Artificial Pancreas Technology
REDCap: Research Electronic Data Capture
T1D: type 1 diabetes

Edited by A Mavragani; submitted 21.04.23; peer-reviewed by H Crocket, I Degen, M Nagpal; comments to author 20.09.23; revised version received 30.11.23; accepted 30.05.24; published 06.01.25.

Please cite as:
Cleal B, Chen Y, Wäldchen M, Ballhausen H, Cooper D, O'Donnell S, Knoll C, Krug N, Raile K, Ubben T, Tappe A, Lewis D, Willaing I, Skinner T, Braune K
Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives
J Med Internet Res 2025;27:e48406
URL: https://www.jmir.org/2025/1/e48406
doi:10.2196/48406
PMID:

Original Paper

Major Adverse Kidney Events in Hospitalized Older Patients With Acute Kidney Injury: Machine Learning–Based Model Development and Validation Study

Xiao-Qin Luo[1], MD; Ning-Ya Zhang[2], MS; Ying-Hao Deng[3], MD; Hong-Shen Wang[4], MD; Yi-Xin Kang[5], MD; Shao-Bin Duan[5], MD

[1]Department of Geriatrics, The Second Xiangya Hospital of Central South University, Changsha, China

[2]Information Center, The Second Xiangya Hospital of Central South University, Changsha, China

[3]Health Management Center, National Clinical Research Center for Metabolic Diseases, Hunan Provincial Clinical Medicine Research Center for Intelligent Management of Chronic Disease, The Second Xiangya Hospital of Central South University, Changsha, China

[4]Department of Critical Care Medicine, The Second Xiangya Hospital of Central South University, Changsha, China

[5]Department of Nephrology, Hunan Key Laboratory of Kidney Disease and Blood Purification, The Second Xiangya Hospital of Central South University, Changsha, China

Corresponding Author:
Shao-Bin Duan, MD
Department of Nephrology
Hunan Key Laboratory of Kidney Disease and Blood Purification
The Second Xiangya Hospital of Central South University
139 Renmin Road
Changsha, 410011
China
Phone: 86 73185295100
Email: duansb528@csu.edu.cn

Abstract

Background: Acute kidney injury (AKI) is a common complication in hospitalized older patients, associated with increased morbidity, mortality, and health care costs. Major adverse kidney events within 30 days (MAKE30), a composite of death, new renal replacement therapy, or persistent renal dysfunction, has been recommended as a patient-centered endpoint for clinical trials involving AKI.

Objective: This study aimed to develop and validate a machine learning–based model to predict MAKE30 in hospitalized older patients with AKI.

Methods: A total of 4266 older patients (aged ≥ 65 years) with AKI admitted to the Second Xiangya Hospital of Central South University from January 1, 2015, to December 31, 2020, were included and randomly divided into a training set and an internal test set in a ratio of 7:3. An additional cohort of 11,864 eligible patients from the Medical Information Mart for Intensive Care Ⅳ database served as an external test set. The Boruta algorithm was used to select the most important predictor variables from 53 candidate variables. The eXtreme Gradient Boosting algorithm was applied to establish a prediction model for MAKE30. Model discrimination was evaluated by the area under the receiver operating characteristic curve (AUROC). The SHapley Additive exPlanations method was used to interpret model predictions.

Results: The overall incidence of MAKE30 in the 2 study cohorts was 28.3% (95% CI 26.9%-29.7%) and 26.7% (95% CI 25.9%-27.5%), respectively. The prediction model for MAKE30 exhibited adequate predictive performance, with an AUROC of 0.868 (95% CI 0.852-0.881) in the training set and 0.823 (95% CI 0.798-0.846) in the internal test set. Its simplified version achieved an AUROC of 0.744 (95% CI 0.735-0.754) in the external test set. The SHapley Additive exPlanations method showed that the use of vasopressors, mechanical ventilation, blood urea nitrogen level, red blood cell distribution width-coefficient of variation, and serum albumin level were closely associated with MAKE30.

Conclusions: An interpretable eXtreme Gradient Boosting model was developed and validated to predict MAKE30, which provides opportunities for risk stratification, clinical decision-making, and the conduct of clinical trials involving AKI.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200061610; https://tinyurl.com/3smf9nuw

(J Med Internet Res 2025;27:e52786) doi:10.2196/52786

KEYWORDS

major adverse kidney events within 30 days; older; acute kidney injury; machine learning; prediction model

Introduction

Acute kidney injury (AKI) is a clinical syndrome characterized by a rapid decline in renal function [1]. The incidence of AKI has been reported to be about 10%-15% in hospitalized patients and over 50% in critically ill patients [2-4]. AKI is common in older individuals, associated with increased morbidity, mortality, and health care costs [5-7]. The prevalence of multiple comorbidities as well as age-related changes in the kidneys, systemic vasculature, and immune system render older patients more susceptible to kidney damage and less likely to recover [5,8]. Given the already high and increasing incidence of AKI in the older, there is a tremendous need to develop clinical tools for risk stratification of postAKI outcomes in this age group.

Major adverse kidney events within 30 days (MAKE30), a composite of death, new renal replacement therapy (RRT), or persistent renal dysfunction (PRD), has been recognized as a key metric of post-AKI outcomes [9,10]. Patients with no renal recovery during hospitalization show a higher risk of long-term adverse outcomes than those who recovered [11,12]. Furthermore, the initiation of RRT after AKI onset is associated with long-term renal dysfunction and death [13,14]. Previous studies have confirmed MAKE30 as a common, feasible to measure, and clinically meaningful endpoint in clinical trials involving AKI [15-17]. The prediction of MAKE30 can be critical for evaluating patient prognosis, guiding clinical decision-making, and facilitating the conduct of clinical trials.

Clinical prediction models are mathematical tools that are primarily intended to assist physicians in their clinical decision-making [18]. Recently, the rapid development in medical big data and advances in computer science have sparked a growing interest in applying machine learning techniques to develop clinical prediction models [19]. Compared with traditional statistical methods, advanced machine learning algorithms can better integrate large amounts of clinical data, fit complex nonlinear relationships, and analyze high-order interactions. The eXtreme Gradient Boosting (XGBoost) algorithm, an efficient implementation of the gradient boosting framework, represents one of the notable advances in machine learning and is widely used in the medical field [20-23]. XGBoost excels in preventing overfitting during the training process, enhancing predictive performance and robustness in complex data scenarios.

Therefore, the objective of this study was to develop and validate a clinical prediction model for MAKE30 in hospitalized older patients with AKI using the machine learning XGBoost algorithm.

Methods

Study Design

This study identified older patients (aged ≥ 65 years) with AKI who were admitted to the Second Xiangya Hospital of Central South University from January 1, 2015, to December 31, 2020. AKI was defined according to the Kidney Disease: Improving Global Outcomes criteria as an increase in serum creatinine (SCr) by ≥26.5 μmol/L within 48 hours or to ≥1.5 times baseline within 7 days [24]. Baseline SCr was defined as the lowest creatinine in the past 7 days. AKI was determined every time a SCr measurement occurred based on the changes in SCr over the past 48 hours and 7 days. Patients were excluded if they were diagnosed with end-stage renal disease, had a hospital stay of less than 48 hours, had an initial SCr≥353.6 μmol/L at admission, or required RRT before the diagnosis of AKI.

In addition, a cohort containing eligible patients from the Medical Information Mart for Intensive Care Ⅳ (MIMIC-Ⅳ) database was used for model external validation [25]. MIMIC-Ⅳ is a relational database containing comprehensive information for intensive care unit (ICU) admissions at the Beth Israel Deaconess Medical Center (BIDMC) from 2008 to 2019.

Data Collection

Data on demographics, diagnoses, laboratory tests, interventions, and medications were collected. A total of 53 candidate predictor variables were identified that were considered clinically relevant to MAKE30, were available in the electronic health records (EHRs), and had a proportion of missing values less than 30% (Multimedia Appendix 1). Comorbidities were determined based on diagnoses encoded in the *ICD-9* (*International Classification of Diseases, Ninth Edition*) or *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*). The burden of comorbidities was assessed by the Charlson comorbidity index [26,27]. Laboratory tests were recorded as the measurements within 48 hours of and closest to the diagnosis of AKI. Medications were identified according to the Anatomic Therapeutic Chemical classification system and included if administered within 48 hours after the diagnosis of AKI [28]. The use of mechanical ventilation within 48 hours after the diagnosis of AKI was also collected.

Outcome Measures

The primary outcome was MAKE30, defined as a composite of death, new RRT, or PRD at hospital discharge or at 30 days after the diagnosis of AKI, whichever occurred first [9,10]. PRD was defined as a final inpatient SCr value ≥2 times baseline. Secondary outcomes were length of hospital stay as well as death within 30 days and 1 year after discharge. Survival data were extracted from the Chinese Center for Disease Control and Prevention cause-of-death reporting system.

Statistical Analysis

Statistical analyses were performed using R software (version 4.1.2; R Foundation for Statistical Computing). Categorical data are presented as numbers and percentages and were analyzed with the chi-square test or Fisher exact test, as appropriate. Continuous data are described with median (IQR) and were compared using the Mann-Whitney *U* test. Survival data were analyzed with the Kaplan-Meier method, and differences in survival were evaluated by the log-rank test. The 2-sided α level was set at .05. Missing data were analyzed by the missForest method, which is a nonparametric missing value imputation for mixed-type data [29]. The description of missing data can be found in Multimedia Appendix 2.

Patients from the Second Xiangya Hospital were randomly divided into a training set and an internal test set in a ratio of 7:3. The cohort of patients from the MIMIC-Ⅳ database was used as an external test set.

Feature Selection

To build a parsimonious model, feature selection was performed on the training set. First, we removed the predictor variables with near-zero variance. Near-zero variance was defined as a situation where a variable exhibits very little variation or almost no variability across its values, characterized by over 95% of its values being identical. Then the Boruta algorithm was used to select the most important predictor variables [30]. Boruta is an all-relevant feature selection wrapper algorithm, capable of working with classification methods that output variable importance measures; in this study, the Random Forest model was used. There were 30 variables ultimately selected for inclusion within the machine learning model (Multimedia Appendix 1).

Model Development and Validation

A machine learning XGBoost model was established on the training set to predict the development of MAKE30 [31]. XGBoost is an optimized implementation of the gradient boosting framework, which sequentially adds weak models (decision trees) to iteratively improve the overall prediction. XGBoost improves upon traditional gradient boosting algorithms by offering scalability, regularization techniques, optimized tree construction, and customization options. A set of model hyperparameters were optimized by running 5 random shuffles of 5-fold cross-validation, including eta, max_depth, min_child_weight, gamma, colsample_bytree, and subsample. The descriptions and search ranges for the hyperparameters are listed in Multimedia Appendix 3. The XGBoost model can be accessed at [LuoXiaoqin123/MAKE30-in-elderly-AKI] (GitHub).

The performance of the XGBoost model was evaluated on the internal and external test sets. Model discrimination was assessed by the area under the receiver operating characteristic curve (AUROC), which is an assessment metric that remains unaffected by incidence or threshold selection. Model calibration was evaluated using the calibration plot aggregated by deciles and the Brier score. The area under the precision-recall curve (AUPRC) was calculated, considering that it is a useful metric for class-imbalanced data in a problem setting where finding the adverse outcome is crucial. A decision curve analysis was performed to illustrate the clinical use of the model. In addition, we determined the optimal cutoff by the receiver operating characteristic (ROC) curve and the maximum Youden index (sensitivity + specificity – 1) in the training set. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), and negative likelihood ratio (NLR) at the optimal cutoff were calculated.

In the external validation set, there were 9 predictor variables with a proportion of missing values exceeding 50% (Multimedia Appendix 2). In this case, we simplified the model to include 21 variables, with model development performed as described above using the same training set.

Model Interpretations

We used the SHapley Additive exPlanations (SHAP) method to interpret model predictions [32]. SHAP is a game theoretic approach that helps in explaining the output of any machine learning model. It can provide consistent and locally accurate attribution values, the SHAP values, for each feature within a model. A positive SHAP value represents a higher risk of the outcome, whereas a negative SHAP value represents a lower risk of the outcome. The contribution of a feature to the predicted risk of the outcome can be explained by the cumulative effects of feature attribution in each observation.

Sensitivity Analyses

In order to mitigate the potential impact of age on our findings, we further conducted sensitivity analyses to assess the model's robustness across patients stratified by age. We evaluated the model's performance in distinct age groups: <70 years, 70 to 80 years, and ≥80 years.

Ethical Considerations

The medical ethics committee of the Second Xiangya Hospital of Central South University approved the study protocol (2022-K031) and waived the requirement for informed consent. This project has been registered in Chinese Clinical Trial Registry (ChiCTR2200061610). The institutional review boards of BIDMC and the Massachusetts Institute of Technology approved the project and waived the requirement for informed consent. The study followed the Declaration of Helsinki and the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis statement [33].

Results

Patient Characteristics

A total of 4266 patients from the Second Xiangya Hospital were enrolled, including 2973 patients in the training set and 1293 patients in the internal test set (Figure 1). The overall incidence of MAKE30 in this medical center was 28.3% (95% CI 26.9%-29.7%), including 12.3% (95% CI 11.4%-13.4%) with death, 10.2% (95% CI 9.3%-11.1%) with new RRT, and 17.7% (95% CI 16.6%-18.9%) with PRD. The external test set contained 11,864 patients from the MIMIC-Ⅳ database, 26.7% (95% CI 25.9%-27.5%) of whom developed MAKE30 (Multimedia Appendix 4). The incidence of death, new RRT, and PRD were 18.9% (95% CI 18.2%-19.6%), 4.9% (95% CI 4.5%-5.3%), and 12.8% (95% CI 12.2%-13.4%), respectively.

Table 1 shows the characteristics of patients from the Second Xiangya Hospital stratified by MAKE30. Patients with MAKE30 were older, had a higher burden of comorbidities, and had greater disease severity than patients without MAKE30. Compared with the nonMAKE30 group, the MAKE30 group had shorter length of hospital stay as well as higher 30-day and 1-year mortality after discharge. Multimedia Appendix 5 shows the Kaplan-Meier survival curves of death within 30 days and 1 year for patients alive at discharge stratified by PRD or RRT. Patients with versus without new RRT or PRD had higher 30-day and 1-year mortality after discharge (log-rank *P*<.001). Multimedia Appendix 6 provides the characteristics of patients from the MIMIC-Ⅳ database stratified by MAKE30. Multimedia Appendix 7 provides the characteristics of patients in the training, internal test, and external test sets.

Figure 1. Flow diagram of patient selection from the Second Xiangya Hospital. AKI: acute kidney injury; MAKE30, major adverse kidney events within 30 days; RRT: renal replacement therapy; SCr: serum creatinine.

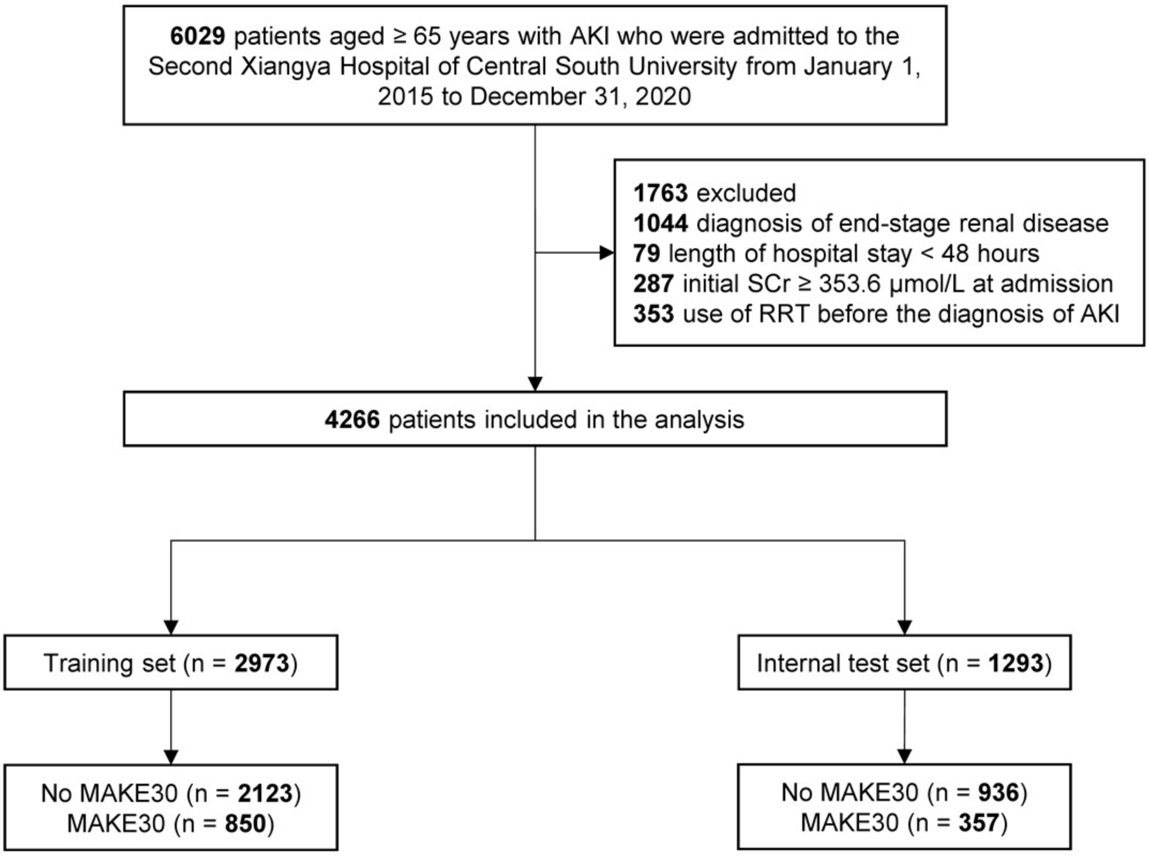


Table 1. Characteristics of patients from the Second Xiangya Hospital.

Variables	No MAKE30[a] (n=3059)	MAKE30[a] (n=1207)	*P* value
Age (years), median (IQR)	72 (68-77)	74 (68-80)	<.001
Sex, male, n (%)	1992 (65.1)	776 (64.3)	.64
Intensive care, n (%)	1284 (42.0)	765 (63.4)	<.001
Comorbidities, n (%)			
Sepsis	127 (4.2)	237 (19.6)	<.001
Hypertension	1685 (55.1)	676 (56.0)	.61
Myocardial infarction	346 (11.3)	170 (14.1)	.01
Congestive heart failure	932 (30.5)	502 (41.6)	<.001
Peripheral vascular disease	455 (14.9)	202 (16.7)	.14
Cerebrovascular disease	611 (20.0)	347 (28.7)	<.001
Dementia	45 (1.5)	41 (3.4)	<.001
Chronic pulmonary disease	515 (16.8)	259 (21.5)	<.001
Rheumatic disease	60 (2.0)	27 (2.2)	.65
Peptic ulcer disease	99 (3.2)	59 (4.9)	.01
Liver disease	630 (20.6)	308 (25.5)	.001
Diabetes	756 (24.7)	349 (28.9)	.005
Hemiplegia or paraplegia	10 (0.3)	3 (0.2)	.91
Renal disease	449 (14.7)	210 (17.4)	.03
Malignancy	913 (29.8)	296 (24.5)	.001
HIV/AIDS	2 (0.1)	0 (0.0)	.92
Charlson Comorbidity Index	2 (1-4)	3 (2-4)	<.001
Laboratory tests			
Red blood cells (×10^9/L), median (IQR)	3.4 (2.9-3.9)	3.2 (2.7-3.8)	<.001
Hemoglobin (g/L), median (IQR)	103 (86-118)	96 (80-114)	<.001
RDW-CV[b] (%), median (IQR)	14.0 (13.1-15.5)	15.0 (13.9-16.9)	<.001
White blood cells (×10^9/L), median (IQR)	9.8 (7.1-13.6)	11.4 (7.5-16.3)	<.001
Neutrophil percentage (%), median (IQR)	83.8 (75.6-89.3)	87.6 (80.8-91.9)	<.001
Lymphocyte percentage (%), median (IQR)	9.2 (5.6-15.1)	7.0 (4.1-11.9)	<.001
Platelets (×10^9/L), median (IQR)	175 (124-236)	147 (90-223)	<.001
Serum total protein (g/L), median (IQR)	58.4 (52.9-63.7)	56.5 (50.5-61.9)	<.001
Serum albumin (g/L), median (IQR)	32.3 (29.1-35.6)	30.0 (26.0-33.7)	<.001
Serum total bilirubin (μmol/L), median (IQR)	11.6 (7.9-18.4)	13.0 (8.0-23.4)	<.001
Serum direct bilirubin (μmol/L), median (IQR)	5.2 (3.3-8.6)	6.7 (3.9-14.1)	<.001
Alanine aminotransferase (U/L), median (IQR)	19.0 (11.5-34.6)	26.1 (13.4-68.2)	<.001
Aspartate aminotransferase (U/L), median (IQR)	26.5 (18.1-48.1)	42.3 (22.8-129.7)	<.001
Serum creatinine (μmol/L), median (IQR)	131.9 (97.7-186.7)	159.0 (109.4-240.2)	<.001
Blood urea nitrogen (mmol/L), median (IQR)	10.36 (7.16-15.36)	15.76 (10.79-22.96)	<.001
Blood uric acid (μmol/L), median (IQR)	331.3 (239.9-443.3)	385.2 (264.7-516.6)	<.001
Potassium (mmol/L), median (IQR)	4.2 (3.8-4.6)	4.3 (3.8-4.9)	<.001
Sodium (mmol/L), median (IQR)	139.5 (136.4-142.6)	141.0 (136.3-146.7)	<.001
Chloride (mmol/L), median (IQR)	103.2 (99.0-106.9)	103.1 (98.5-107.9)	.69

Variables	No MAKE30[a] (n=3059)	MAKE30[a] (n=1207)	*P* value
Calcium (mmol/L), median (IQR)	2.07 (1.96-2.20)	2.03 (1.90-2.17)	<.001
Mechanical ventilation, n (%)	442 (14.4)	564 (46.7)	<.001
Medications, n (%)			
Vasopressors	521 (17.0)	660 (54.7)	<.001
Diuretics	1699 (55.5)	849 (70.3)	<.001
ACEI[c]/ARB[d]	457 (14.9)	149 (12.3)	.03
NSAIDs[e]	1221 (39.9)	310 (25.7)	<.001
Proton pump inhibitors	2308 (75.4)	881 (73.0)	.10
Chemotherapeutic drugs	45 (1.5)	20 (1.7)	.76
Antiepileptic drugs	108 (3.5)	30 (2.5)	.10
Antituberculosis drugs	16 (0.5)	7 (0.6)	>.99
Nephrotoxic antibiotics	223 (7.3)	203 (16.8)	<.001
Antiviral drugs	41 (1.3)	35 (2.9)	.001
Antifungal drugs	186 (6.1)	232 (19.2)	<.001
Iodinated contrast media	159 (5.2)	50 (4.1)	.17
Outcomes			
Length of hospital stay (days), median (IQR)	8 (4-14)	7 (2-14)	<.001
Death within 30 days after discharge[f,g], n (%)	201 (6.8)	137 (21.5)	<.001
Death within 1 year after discharge[f,g], n (%)	637 (21.7)	247 (38.8)	<.001

[a] MAKE30: Major adverse kidney events within 30 days.

[b] RDW-CV: red blood cell distribution width-coefficient of variation.

[c] ACEI: angiotensin converting enzyme inhibitor.

[d] ARB: angiotensin Ⅱ receptor blocker.

[e] NSAID: non-steroidal anti-inflammatory drug.

[f] There were 165 patients with missing postdischarge survival data.

[g] A total of 526 patients who died at hospital discharge or 30 days, whichever occurred first, were excluded.

Model Performance

Figure 2 shows the ROC curves of the model in the training set and internal test set. The model exhibited good discrimination on both datasets, with the AUROC being 0.868 (95% CI 0.852-0.881) and 0.823 (95% CI 0.798-0.846), respectively. Figure 3 shows the calibration plots of the model in the training set and internal test set. The Brier scores on the 2 datasets were 0.127 and 0.145, respectively. Notably, despite the lower Brier score on the training set, the calibration plot exhibits pronounced deviations from the diagonal, especially at higher probability thresholds. In contrast, the calibration plot for the internal test set appears to demonstrate more consistent alignment with the ideal diagonal line across various probability ranges. Figure 4 shows the precision-recall curves of the model in the training set and internal test set. The model achieved an AUPRC of 0.778 and 0.639 on the 2 datasets, respectively. Figure 5 shows the decision curves of the model in the training set and internal test set. When the threshold probability was between 0.06 and 0.97, the model improved clinical decision-making in the training set. Similarly, within a narrower threshold probability range of 0.09 to 0.76, the model improved clinical decision-making in the internal test set. At the optimal cutoff of 0.350, the model achieved a sensitivity of 67.5%, a specificity of 82.5%, a PPV of 59.5%, and a NPV of 86.9% in the internal test set (Table 2).

Multimedia Appendix 8 presents the performance of the simplified model in the training set, the internal test set, and the external test set. It includes ROC curves, calibration plots, precision-recall curves, decision curves, and model performance at the optimal cutoff. The simplified model achieved an AUROC of 0.851 (95% CI 0.834-0.866) on the training set, 0.818 (95% CI 0.792-0.843) on the internal test set, and 0.744 (95% CI 0.735-0.754) on the external test set.

Figure 2. Receiver operating characteristic curves of the model in the training set (A) and internal test set (B). AUC: area under the curve.

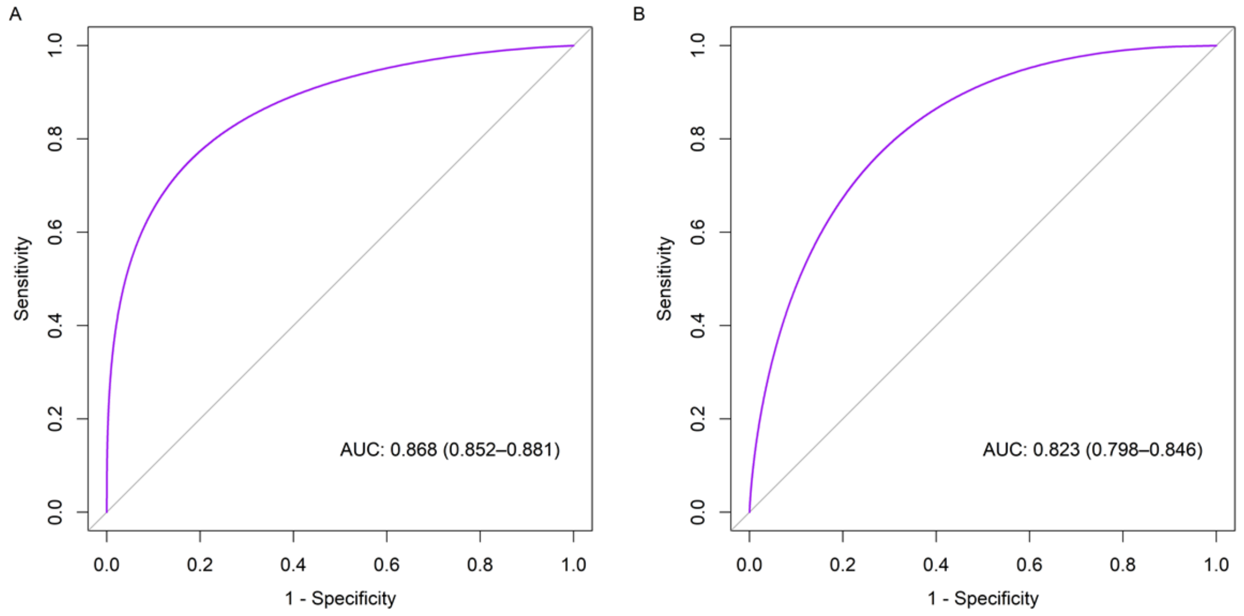


Figure 3. Calibration plots of the model in the training set (A) and internal test set (B).

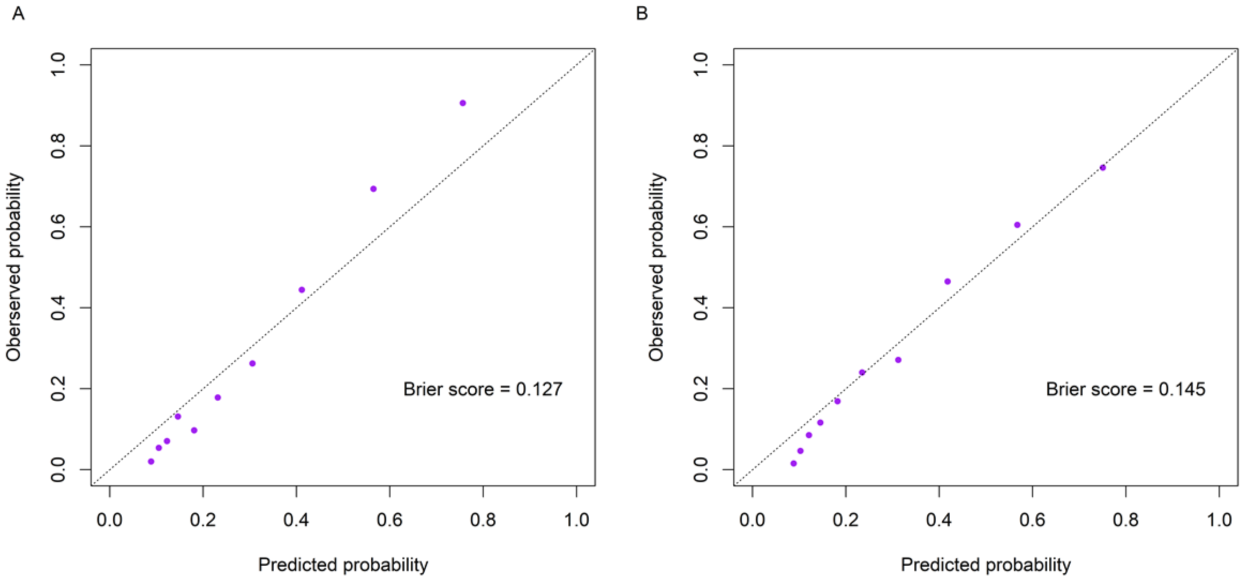


Figure 4. Precision-recall curves of the model in the training set (A) and internal test set (B). AUC: area under the curve.

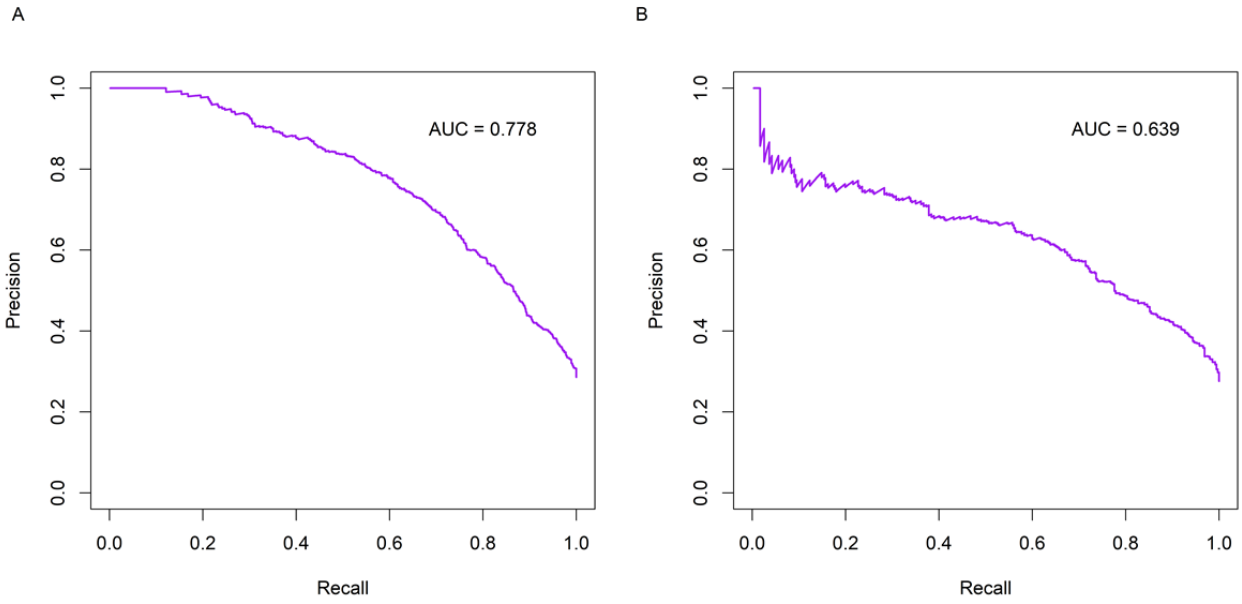


Figure 5. Decision curves of the model in the training set (A) and internal test set (B).

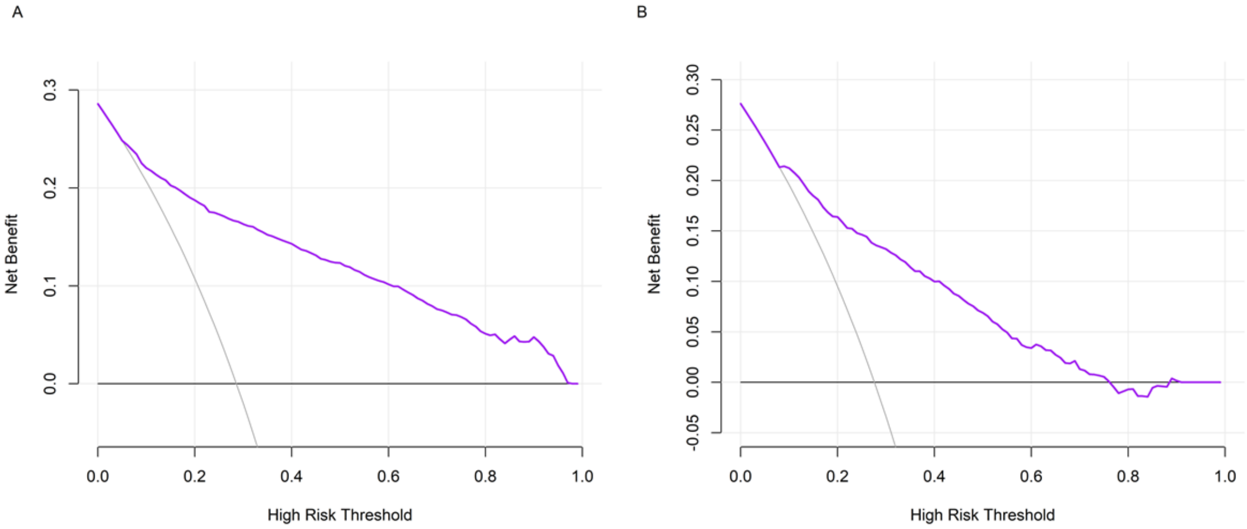


Table 2. Model performance at the optimal cutoff in the training and internal test sets.

Performance metrics	Training set	Internal test set
Cutoff value[a]	0.350	0.350
Sensitivity (%)	71.9	67.5
Specificity (%)	86.6	82.5
PPV[b] (%)	68.3	59.5
NPV[c] (%)	88.5	86.9
PLR[d]	5.37	3.86
NLR[e]	0.32	0.39

[a] The optimal cutoff was determined by the receiver operating characteristic curve and the maximum Youden index (sensitivity + specificity – 1) in the training set.

[b] PPV: positive predictive value.

[c] NPV: negative predictive value.

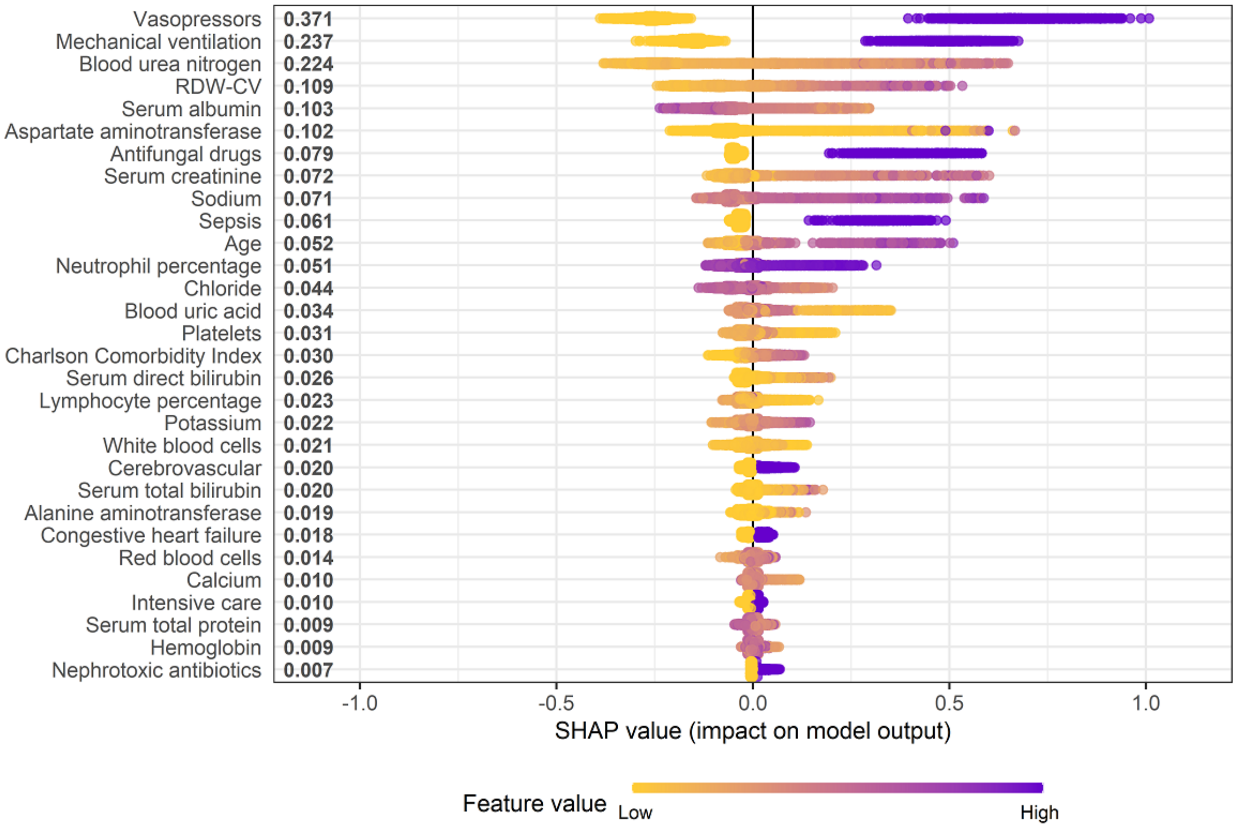
[d] PLR, positive likelihood ratio.

[e] NLR, negative likelihood ratio.

Model Interpretations

Figure 6 shows the SHAP summary plots of the model. The use of vasopressors, requirement for mechanical ventilation, increased blood urea nitrogen level, increased red blood cell distribution width coefficient of variation, decreased serum albumin level, elevated aspartate aminotransferase level, use of antifungal drugs, and increased SCr level were identified as important predictor variables associated with an increased risk of MAKE30. Multimedia Appendix 9 provides the SHAP dependence plots of the model that visualize how changes in each variable can affect model output.

Figure 6. SHapley Additive exPlanations summary plot of the model. RDW-CV: red blood cell distribution width-coefficient of variation.



Sensitivity Analyses

Multimedia Appendix 10 shows the ROC curves of the model and the simplified model in distinct age groups. In the internal test set, the model achieved AUROCs of 0.827 (95% CI 0.786-0.862), 0.835 (95% CI 0.794-0.869), and 0.770 (95% CI 0.708-0.837) in patients aged <70 years, 70 to 80 years, and ≥80 years, respectively. When assessing the simplified model across distinct age groups in the external test set, its performance exhibited overall stability.

Discussion

Principal Findings

This study showed that MAKE30 is common in hospitalized older patients with AKI. The main deliverable of this study is the development and validation of a machine learning–based model to predict MAKE30, representing a pioneering effort to address the critical need in this vulnerable population. The XGBoost model (or its simplified version) achieved adequate predictive performance in both internal and external validation. The model could be a useful tool for prognostic assessment, clinical decisionmaking, and the conduct of clinical trials involving AKI.

A National Institute of Diabetes and Digestive and Kidney Diseases Workshop recommended a composite of death, provision of dialysis, or sustained loss of kidney function at a discrete time point as a meaningful endpoint for trials [9]. This endpoint was later expanded to be MAKE30, which occurred at hospital discharge truncated at 30 days [34,35]. MAKE30 has been shown to be common, easily measurable, and a promising endpoint for patients with AKI. Our study further confirmed that MAKE30 is a prevalent endpoint among hospitalized older patients with AKI, with its nonfatal components being associated with long-term outcomes.

The prediction of MAKE30 is essential for risk stratification and clinical management of hospitalized older patients with AKI. While several prediction models for MAKE30 have been developed, none specifically focus on this population. A recent study developed a logistic regression model specifically for predicting MAKE30 among critically ill adults [36]. Our previous studies established machine learning–based models to predict MAKE30 in hospitalized children with AKI [37] and older patients in critical care [38]. To the best of our knowledge, this study is the first to use machine learning algorithm to predict MAKE30 in hospitalized older patients with AKI. This demographic often experiences impaired recovery of kidney function, resulting in significantly increased morbidity and mortality. By developing a prediction model tailored to this population, we aimed to enhance clinical outcomes and facilitate more effective interventions. The model can be seamlessly integrated into electronic medical record systems, assisting health care providers in risk stratification and informed clinical decision-making. In addition, in clinical trials involving AKI, the model can identify high-risk groups that are more likely to benefit from the intervention.

Compared with the existing studies, this study has several strengths. First, by using state-of-the-art XGBoost machine learning algorithm, our model represented a significant advancement over previous studies that primarily relied on clinical rationale and simpler statistical methods. XGBoost is particularly effective due to its ability to process large amounts of data and identify intricate patterns within complex datasets. Its gradient boosting framework enhances predictive accuracy by combining the strengths of multiple weak learners, leading to more robust and reliable outcomes. Furthermore, XGBoost's flexibility allows for easy tuning of hyperparameters, enabling us to optimize the model for our specific dataset and improve its performance in predicting MAKE30 in hospitalized older patients with AKI.

Second, this study used a larger sample size and conducted external validation to evaluate the model's performance. Notably, the results of the external validation showed that the simplified version of the model exhibited a gap of approximately 10% in AUROC between the training set and the external test set. However, the performance of the simplified model on the internal test set was comparable to that of the full model, suggesting that the observed gap may be attributed to the use of a different test set with distinct characteristics. Compared with the training set, the external test set comprised critically ill patients from various ethnic backgrounds and with distinct patient characteristics and data quality. These findings underscore the necessity for comprehensive testing across diverse datasets to ensure the model's reliability and applicability in real-world settings. Another noteworthy observation is that, although the Brier scores for both the training and test data suggested acceptable calibration, a deeper examination through visual inspection uncovered disparities. Factors contributing to the observed discrepancies may include overfitting during training, differences in underlying data distributions, model complexity, or nuances in evaluation methodology. Further investigation into the factors contributing to the observed discrepancies is warranted.

Finally, this study delved into the interpretability of the model. A major challenge in the clinical implementation of machine learning is how to uncover its "black box" nature. Usually, data come in and decisions go out, but the process between input and output is opaque. The advantage of our study is the use of SHAP method to explain the critical aspects of the data and fully understand the model. The SHAP method identified important predictor variables associated with the development of MAKE30, which could allow early intervention of modifiable factors to mitigate the risk of MAKE30. Another notable finding is the nonlinear relationship between the predictor variables and the risk of MEKE30, as shown in the SHAP dependence plots. This relationship is often ignored by traditional regression analysis, which requires a linearity between the independent variables and the outcome.

Limitations

Our study has several limitations. First, the MIMIC-Ⅳ database only contained data from critically ill patients admitted to the ICU. Model performance awaits further validation in larger samples of patients at different medical centers. Second, it remains unclear whether the model will perform well in individual prognostication and whether its clinical application will improve patient prognosis. Thus, clinical impact studies are needed to confirm the effectiveness of the model. Third, the unavailability of preadmission SCr data could lead to the omission of patients who could have been diagnosed with AKI, that is, community-acquired AKI. Fourth, the urine output criteria were not used to define AKI because hourly urine output data were unavailable for most patients. Finally, the missForest method, as a single imputation method, might affect the SHAP results by not considering sufficient variation for the imputation of missing data.

Conclusions

MAKE30 is common in hospitalized older patients with AKI. An interpretable machine learning XGBoost model was developed and validated to predict MAKE30. The model exhibited adequate predictive performance, which provides opportunities for risk stratification, clinical decision-making, and the conduct of clinical trials involving AKI. Future studies are needed to support the robustness and clinical effectiveness of the model.

Authors' Contributions

SBD designed and supervised the study and revised the manuscript. XQL performed the data extraction, analyzed and interpreted the data and drafted the manuscript. NYZ performed the data extraction and revised the manuscript critically for important intellectual content. YHD, HSW, and YXK interpreted the data and critically revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The list of 53 candidate predictor variables and 30 variables identified in feature selection.
[DOCX File , 50 KB - jmir_v27i1e52786_app1.docx]

Multimedia Appendix 2
Missing data in predictor variables.
[DOCX File , 55 KB - jmir_v27i1e52786_app2.docx]

Multimedia Appendix 3
The descriptions and search ranges for the hyperparameters of the XGBoost (eXtreme Gradient Boosting) model.
[DOCX File , 52 KB - jmir_v27i1e52786_app3.docx]

Multimedia Appendix 4
Flow diagram of patient selection from the MIMIC-Ⅳ (Medical Information Mart for Intensive Care Ⅳ) database.
[DOCX File , 603 KB - jmir_v27i1e52786_app4.docx]

Multimedia Appendix 5
Kaplan-Meier survival curves of death within 30 days and 1 year for patients alive at discharge from the Second Xiangya Hospital.
[DOCX File , 1248 KB - jmir_v27i1e52786_app5.docx]

Multimedia Appendix 6
Characteristics of patients from the MIMIC-Ⅳ (Medical Information Mart for Intensive Care Ⅳ) database.
[DOCX File , 55 KB - jmir_v27i1e52786_app6.docx]

Multimedia Appendix 7
Characteristics of patients in the training set, internal test set, and external test set.
[DOCX File , 56 KB - jmir_v27i1e52786_app7.docx]

Multimedia Appendix 8
Performance of the simplified model in the training set, internal test set, and external test set.
[DOCX File , 2620 KB - jmir_v27i1e52786_app8.docx]

Multimedia Appendix 9
SHAP (SHapley Additive exPlanations) dependence plots of the model.
[DOCX File , 1842 KB - jmir_v27i1e52786_app9.docx]

Multimedia Appendix 10
Receiver operating characteristic curves of the model and the simplified model in distinct age groups.

[DOCX File , 1300 KB - jmir_v27i1e52786_app10.docx]

References

1. Kellum JA, Romagnani P, Ashuntantang G, Ronco C, Zarbock A, Anders H. Acute kidney injury. Nat Rev Dis Primers 2021 Jul 15;7(1):52. [doi: 10.1038/s41572-021-00284-z] [Medline: 34267223]
2. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical Decision Support for In-Hospital AKI. J Am Soc Nephrol 2018 Feb;29(2):654-660 [FREE Full text] [doi: 10.1681/ASN.2017070765] [Medline: 29097621]
3. Hoste EAJ, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 2015 Aug;41(8):1411-1423. [doi: 10.1007/s00134-015-3934-7] [Medline: 26162677]
4. Ronco C, Bellomo R, Kellum JA. Acute kidney injury. Lancet 2019 Nov 23;394(10212):1949-1964. [doi: 10.1016/S0140-6736(19)32563-2] [Medline: 31777389]
5. Chronopoulos A, Cruz DN, Ronco C. Hospital-acquired acute kidney injury in the elderly. Nat Rev Nephrol 2010 Mar;6(3):141-149. [doi: 10.1038/nrneph.2009.234] [Medline: 20125094]
6. Stallone G, Infante B, Grandaliano G. Acute kidney injury in the elderly population. J Nephrol 2012;25 Suppl 19:S58-S66. [doi: 10.5301/jn.5000140] [Medline: 22760878]
7. Chang-Panesso M. Acute kidney injury and aging. Pediatr Nephrol 2021 Oct;36(10):2997-3006 [FREE Full text] [doi: 10.1007/s00467-020-04849-0] [Medline: 33411069]
8. Schmitt R, Coca S, Kanbay M, Tinetti ME, Cantley LG, Parikh CR. Recovery of kidney function after acute kidney injury in the elderly: a systematic review and meta-analysis. Am J Kidney Dis 2008 Aug;52(2):262-271. [doi: 10.1053/j.ajkd.2008.03.005] [Medline: 18511164]
9. Palevsky PM, Molitoris BA, Okusa MD, Levin A, Waikar SS, Wald R, et al. Design of clinical trials in acute kidney injury: report from an NIDDK workshop on trial methodology. Clin J Am Soc Nephrol 2012 May;7(5):844-850. [doi: 10.2215/CJN.12791211] [Medline: 22442182]
10. Semler MW, Rice TW, Shaw AD, Siew ED, Self WH, Kumar AB, et al. Identification of Major Adverse Kidney Events Within the Electronic Health Record. J Med Syst 2016 Jul;40(7):167 [FREE Full text] [doi: 10.1007/s10916-016-0528-z] [Medline: 27234478]
11. Heung M, Steffick DE, Zivin K, Gillespie BW, Banerjee T, Hsu CY, Centers for Disease ControlPrevention CKD Surveillance Team. Acute Kidney Injury Recovery Pattern and Subsequent Risk of CKD: An Analysis of Veterans Health Administration Data. Am J Kidney Dis 2016 May;67(5):742-752 [FREE Full text] [doi: 10.1053/j.ajkd.2015.10.019] [Medline: 26690912]
12. Fiorentino M, Tohme FA, Wang S, Murugan R, Angus DC, Kellum JA. Long-term survival in patients with septic acute kidney injury is strongly influenced by renal recovery. PLoS One 2018;13(6):e0198269 [FREE Full text] [doi: 10.1371/journal.pone.0198269] [Medline: 29870535]
13. De Corte W, Dhondt A, Vanholder R, De Waele J, Decruyenaere J, Sergoyne V, et al. Long-term outcome in ICU patients with acute kidney injury treated with renal replacement therapy: a prospective cohort study. Crit Care 2016 Aug 12;20(1):256 [FREE Full text] [doi: 10.1186/s13054-016-1409-z] [Medline: 27520553]
14. Kirwan CJ, Blunden MJ, Dobbie H, James A, Nedungadi A, Prowle JR. Critically ill patients requiring acute renal replacement therapy are at an increased risk of long-term renal dysfunction, but rarely receive specialist nephrology follow-up. Nephron 2015;129(3):164-170. [doi: 10.1159/000371448] [Medline: 25765730]
15. Self WH, Semler MW, Wanderer JP, Wang L, Byrne DW, Collins SP, SALT-ED Investigators. Balanced Crystalloids versus Saline in Noncritically Ill Adults. N Engl J Med 2018 Mar 01;378(9):819-828 [FREE Full text] [doi: 10.1056/NEJMoa1711586] [Medline: 29485926]
16. Semler MW, Self WH, Wanderer JP, Ehrenfeld JM, Wang L, Byrne DW, SMART Investigatorsthe Pragmatic Critical Care Research Group. Balanced Crystalloids versus Saline in Critically Ill Adults. N Engl J Med 2018 Mar 01;378(9):829-839 [FREE Full text] [doi: 10.1056/NEJMoa1711584] [Medline: 29485925]
17. Weiss SL, Balamuth F, Thurm CW, Downes KJ, Fitzgerald JC, Laskin BL. Major Adverse Kidney Events in Pediatric Sepsis. Clin J Am Soc Nephrol 2019 May 07;14(5):664-672 [FREE Full text] [doi: 10.2215/CJN.12201018] [Medline: 31000518]
18. Labarère J, Renaud B, Fine MJ. How to derive and validate clinical prediction models for use in intensive care medicine. Intensive Care Med 2014 Apr;40(4):513-527. [doi: 10.1007/s00134-014-3227-6] [Medline: 24570265]
19. Deo RC. Machine Learning in Medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
20. 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]
21. Chen T, Li X, Li Y, Xia E, Qin Y, Liang S, et al. Prediction and Risk Stratification of Kidney Outcomes in IgA Nephropathy. Am J Kidney Dis 2019 Sep;74(3):300-309. [doi: 10.1053/j.ajkd.2019.02.016] [Medline: 31031086]
22. Yue S, Li S, Huang X, Liu J, Hou X, Zhao Y, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. J Transl Med 2022 May 13;20(1):215 [FREE Full text] [doi: 10.1186/s12967-022-03364-0] [Medline: 35562803]

23. Luo XQ, Kang YX, Duan SB, Yan P, Song GB, Zhang NY, et al. Machine Learning-Based Prediction of Acute Kidney Injury Following Pediatric Cardiac Surgery: Model Development and Validation Study. J Med Internet Res 2023 Jan 05;25:e41142 [FREE Full text] [doi: 10.2196/41142] [Medline: 36603200]
24. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO clinical practice guideline for acute kidney injury. Kidney International Supplements 2012 Mar;2(1):1-138 [FREE Full text] [doi: 10.1038/kisup.2012.1]
25. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV (version 1.0). PhysioNet. 2021. URL: https://physionet.org/content/mimiciv/1.0/ [accessed 2024-11-23]
26. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
27. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005 Nov;43(11):1130-1139. [doi: 10.1097/01.mlr.0000182534.19832.83] [Medline: 16224307]
28. Xu X, Nie S, Zhang A, Mao J, Liu HP, Xia H, et al. Acute Kidney Injury among Hospitalized Children in China. Clin J Am Soc Nephrol 2018 Dec 07;13(12):1791-1800 [FREE Full text] [doi: 10.2215/CJN.00800118] [Medline: 30287424]
29. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 2012 Jan 01;28(1):112-118. [doi: 10.1093/bioinformatics/btr597] [Medline: 22039212]
30. Kursa MB, Rudnicki WR. Feature selection with boruta package. J. Stat. Soft 2010;36(11):1-13. [doi: 10.18637/jss.v036.i11]
31. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. : Association for Computing Machinery; 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 August 13-17; San Francisco, California, USA p. 785-794. [doi: 10.1145/2939672.2939785]
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Adv Neural Inf Process Syst. Red Hook, New York: Curran Associates, Inc; 2017.
33. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015 Jan 06;162(1):55-63 [FREE Full text] [doi: 10.7326/M14-0697] [Medline: 25560714]
34. Kashani K, Al-Khafaji A, Ardiles T, Artigas A, Bagshaw SM, Bell M, et al. Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. Crit Care 2013 Feb 06;17(1):R25 [FREE Full text] [doi: 10.1186/cc12503] [Medline: 23388612]
35. Hoste EAJ, McCullough PA, Kashani K, Chawla LS, Joannidis M, Shaw AD, Sapphire Investigators. Derivation and validation of cutoffs for clinical use of cell cycle arrest biomarkers. Nephrol Dial Transplant 2014 Nov;29(11):2054-2061 [FREE Full text] [doi: 10.1093/ndt/gfu292] [Medline: 25237065]
36. McKown AC, Wang L, Wanderer JP, Ehrenfeld J, Rice TW, Bernard GR, et al. Predicting Major Adverse Kidney Events among Critically Ill Adults Using the Electronic Health Record. J Med Syst 2017 Aug 31;41(10):156 [FREE Full text] [doi: 10.1007/s10916-017-0806-4] [Medline: 28861688]
37. Deng YH, Luo XQ, Yan P, Zhang NY, Liu Y, Duan SB. Outcome prediction for acute kidney injury among hospitalized children via eXtreme Gradient Boosting algorithm. Sci Rep 2022 May 27;12(1):8956 [FREE Full text] [doi: 10.1038/s41598-022-13152-x] [Medline: 35624143]
38. Wang L, Duan SB, Yan P, Luo XQ, Zhang NY. Utilization of interpretable machine learning model to forecast the risk of major adverse kidney events in elderly patients in critical care. Ren Fail 2023 Dec;45(1):2215329 [FREE Full text] [doi: 10.1080/0886022X.2023.2215329] [Medline: 37218683]

Abbreviations

AKI: Acute kidney injury
AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
BIDMC: Beth Israel Deaconess Medical Center
EHR: electronic health record
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10: International Statistical Classification of Diseases, Tenth Revision
ICU: intensive care unit
MAKE30: Major adverse kidney events within 30 days
MIMIC-Ⅳ: Medical Information Mart for Intensive Care Ⅳ
NLR: negative likelihood ratio
NPV: negative predictive value
PLR: positive likelihood ratio
PPV: positive predictive value

PRD: persistent renal dysfunction
ROC: receiver operating characteristic
RRT: renal replacement therapy
SCr: serum creatinine
SHAP: SHapley Additive exPlanations
XGBoost: eXtreme Gradient Boosting

Edited by A Mavragani; submitted 15.09.23; peer-reviewed by M Rezk, D Nelson, Q Yang, X Zhang; comments to author 26.02.24; revised version received 18.04.24; accepted 12.11.24; published 03.01.25.

Please cite as:
Luo XQ, Zhang NY, Deng YH, Wang HS, Kang YX, Duan SB
Major Adverse Kidney Events in Hospitalized Older Patients With Acute Kidney Injury: Machine Learning–Based Model Development and Validation Study
J Med Internet Res 2025;27:e52786
URL: https://www.jmir.org/2025/1/e52786
doi:10.2196/52786
PMID:

Original Paper

Recruitment of Young Gay, Bisexual, and Other Men Who Have Sex With Men for a Web-Based Human Papillomavirus Vaccination Intervention: Differences in Participant Characteristics and Study Engagement by Recruitment Source in a Randomized Controlled Trial

Daniel J Marshall[1], MSc; Amy L Gower[2], PhD; Mira L Katz[1,3], PhD; José A Bauermeister[4], PhD; Abigail B Shoben[1], PhD; Paul L Reiter[1,3], PhD

[1]College of Public Health, The Ohio State University, Columbus, OH, United States

[2]Division of General Pediatrics and Adolescent Health, Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN, United States

[3]Comprehensive Cancer Center, The Ohio State University, Columbus, OH, United States

[4]Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Paul L Reiter, PhD
College of Public Health
The Ohio State University
1841 Neil Ave.
Columbus, OH, 43210
United States
Phone: 1 6142924803
Email: reiter.36@osu.edu

Abstract

Background: Young gay, bisexual, and other men who have sex with men have been referred to as a "hard-to-reach" or "hidden" community in terms of recruiting for research studies. With widespread internet use among this group and young adults in general, web-based avenues represent an important approach for reaching and recruiting members of this community. However, little is known about how participants recruited from various web-based sources may differ from one another.

Objective: This study aimed to determine how young gay, bisexual, and other men who have sex with men recruited from various web-based sources differ from one another in terms of participant characteristics and study engagement.

Methods: Data were collected as part of a randomized controlled trial of Outsmart HPV, a web-based human papillomavirus (HPV) vaccination intervention for young gay, bisexual, and other men who have sex with men. From 2019 to 2021, we recruited young gay, bisexual, and other men who have sex with men in the United States who were aged 18-25 years and not vaccinated against HPV (n=1227) through various web-based avenues. We classified each participant as being recruited from either (1) social media (eg, Facebook, Instagram, Snapchat), (2) a dating app (eg, Grindr, Scruff), or (3) some other digital recruitment source (eg, existing research panel, university-based organization). Analyses compared participants from these 3 groups on demographic and health-related characteristics and metrics involving study engagement.

Results: Most demographic and health-related characteristics differed by web-based recruitment source, including race or ethnicity (*P*<.001), relationship status (*P*<.001), education level (*P*<.001), employment status (*P*<.001), sexual self-identity (*P*<.001), health insurance status (*P*<.001), disclosure of sexual orientation (*P*=.048), and connectedness to the LGBTQ (lesbian, gay, bisexual, transgender, queer) community (*P*<.001) The type of device used by participants during study enrollment also differed across groups, with smartphone use higher among participants recruited via dating apps (n=660, 96.6%) compared to those recruited via social media (n=318, 78.9%) or other digital sources (n=85, 60.3%; *P*<.001). Participants recruited via social media were more likely than those recruited via dating apps to complete follow-up surveys at 3 different timepoints (odds ratios 1.52-2.09, *P*=.001-.008). These participants also spent a longer amount of time viewing intervention content about HPV vaccination (3.14 minutes vs 2.67 minutes; *P*=.02).

Conclusions: We were able to recruit a large national sample of young gay, bisexual, and other men who have sex with men for a web-based HPV vaccination intervention via multiple methodologies. Participants differed on a range of demographic and health-related characteristics, as well as metrics related to study engagement, based on whether they were recruited from social media, a dating app, or some other digital recruitment source. Findings highlight key issues and considerations that can help researchers better plan and customize future web-based recruitment efforts of young gay, bisexual, and other men who have sex with men.

Trial Registration: ClinicalTrials.gov NCT04032106; https://clinicaltrials.gov/study/NCT04032106

International Registered Report Identifier (IRRID): RR2-10.2196/16294

(J Med Internet Res 2025;27:e64668) doi:10.2196/64668

KEYWORDS

study recruitment; gay and bisexual men; human papillomavirus; vaccination promotion; digital intervention; social media; dating apps; recruitment; young adults; adolescents; gay; bisexual; men who have sex with men

Introduction

Human papillomavirus (HPV) is the most common sexually transmitted infection in the United States [1], with infections having the potential to cause multiple types of cancer (ie, anal, oropharyngeal, and penile cancers) and genital warts in men [2]. Gay, bisexual, and other men who have sex with men tend to have higher rates of HPV infection and HPV-related disease compared to other men [3-5]. Routine HPV vaccination is currently recommended for those aged 11-12 years in the United States, while also recommended for everyone through the age of 26 years who has not already been vaccinated [6]. However, a recent review paper suggests that fewer than 40% of young gay, bisexual, and other men who have sex with men who are age-eligible for HPV vaccination have received any doses of the vaccine series [5]. In response, HPV vaccination interventions for young gay, bisexual, and other men who have sex with men have recently been developed to improve knowledge and increase vaccination rates [7-10].

One issue central to these interventions is reaching and recruiting young gay, bisexual, and other men who have sex with men as study participants [11]. Gay, bisexual, and other men who have sex with men have previously been referred to as a "hard-to-reach" or "hidden" population for study recruitment [12,13], and traditional recruitment approaches for this community often included snowball sampling, use of "gatekeeper" organizations associated with the community, and venue-based recruitment at community events, health clinics, or other settings [11,14,15]. These traditional approaches faced recruitment challenges including potential participants not being comfortable identifying as gay, bisexual, and other men who have sex with men, lack of inclusive language in recruitment materials, and concerns about stigmatization associated with participation [11-13,15,16].

Web-based recruitment methods, including the use of social media and dating apps, present an alternative approach that has gained popularity for recruiting young gay, bisexual, and other men who have sex with men [17-21]. Social media use is ubiquitous among young adults in the United States, with nearly all reporting prior use [22]. Further, about 80% of young gay, bisexual, and other men who have sex with men report using a dating app on at least a monthly basis [23]. Past studies have shown that digital venues can successfully reach and recruit large numbers of participants, including young gay, bisexual, and other men who have sex with men, and that the content of study advertisements on these platforms can affect recruitment metrics [17,24-29]. Studies have also provided a great deal of information about the costs associated with web-based recruitment [24,28,29] and shown that participants recruited via web-based methodologies are demographically different from those recruited through more traditional approaches [13,30-33]. For example, gay, bisexual, and other men who have sex with men participants recruited via web-based methodologies may differ in terms of age, race or ethnicity, and socioeconomic status (eg, education level) compared to those recruited via more traditional approaches [13,30-32]. Furthermore, a few studies have shown that gay, bisexual, and other men who have sex with men participants tend to be more diverse when multiple recruitment sources are used (eg, using Facebook and Craigslist) [18,21].

Even with our current amount of knowledge about the digital recruitment of young gay, bisexual, and other men who have sex with men, important research gaps remain. One key area to examine is how participants recruited from various web-based sources (eg, social media vs dating apps) may differ from one another. This is true not only for demographic characteristics but also for participants' study experience (eg, retention, engagement with study content). The latter is particularly relevant to studies that include the delivery of interventions or involve longitudinal data collection. In this report, we address these research gaps by analyzing data from Outsmart HPV, an HPV vaccination intervention study that recruited young gay, bisexual, and other men who have sex with men through various web-based venues. Results will inform and help guide future web-based recruitment efforts of young gay, bisexual, and other men who have sex with men for research studies.

Methods

Participants

All data were collected as part of a randomized controlled trial (RCT) of Outsmart HPV. The methods of the RCT have been described previously [7] and are briefly summarized below. We recruited a convenience sample of young gay, bisexual, and other men who have sex with men via social media sites, dating apps, and other web-based avenues. Advertisements on these digital platforms included a combination of images (eg, men in the targeted age range) and brief text about the study (eg, information about HPV, the study being web-based). Although we standardized advertisement content where possible, some content did differ across platforms due to varying advertisement requirements and options (eg, amount of text allowed, advertisement placement). For example, some of our advertisements through dating apps featured content that was sent directly to users' in-app inboxes over a 24-hour period, which was an approach that was not available across all platforms.

Interested individuals were linked via advertisements to a mobile-friendly project website to complete an eligibility screener. Eligibility criteria included self-identifying as (1) cisgender male; (2) aged 18-25 years; (3) either gay, bisexual, or queer; ever having oral or anal sex with another male; or being sexually attracted to other males; (4) living in the United States; (5) not having received any doses of HPV vaccine; and (5) not having previously participated in Outsmart HPV. Eligible individuals provided informed consent and created a project website account.

Participants next completed a baseline survey ("T1 survey") and were then immediately randomized using a 1:1:1 allocation scheme to 1 of 3 study groups. The 3 study groups included 2 groups that received intervention content and 1 control group [7]. Participants then viewed intervention content about HPV vaccination on the project website, with participants in both intervention groups viewing Outsmart HPV content about HPV vaccination and participants in the control group viewing standard information about HPV vaccination. After viewing the intervention content, participants completed a second survey ("T2 survey"). Additional follow-up surveys occurred 3 and 9 months later ("T3 survey" and "T4 survey," respectively). All study surveys were completed on the project website. A total of 1227 participants were randomized from October 2019 to June 2021.

Measures

The main independent variable for this report was the type of web-based recruitment source, with each participant categorized as being recruited from either: (1) social media (eg, Facebook, Instagram, Snapchat); (2) a dating app (eg, Grindr, Scruff); or (3) some other recruitment source (eg, existing research panel, university-based organization).

The T1 survey collected information on a range of demographic and health-related characteristics. This included participants' disclosure of their sexual orientation (3 items; α=.77; possible range 1-5) [34], concealment of their sexual orientation (3 items; α=.75; possible range 1-5) [34], connectedness to the LGBTQ (lesbian, gay, bisexual, transgender, queer) community (2 items; α=.72; possible range 1-4), and electronic health literacy (4 items; α=.83; possible range 1-5) [35]. Items were coded so that higher values indicate greater levels of a given construct. We examined the type of device that participants used during study enrollment, with each participant categorized as using either a smartphone, tablet device, or a personal computer (eg, desktop, laptop).

To examine study retention, we determined whether participants completed the T2, T3, and T4 surveys. For each survey separately, participants were categorized as having either completed the survey or not completed the survey. Among participants in the 2 intervention groups, we also examined additional metrics of study engagement by calculating the total amount of time (in minutes) spent viewing Outsmart HPV content about HPV vaccination and the total number of logins to the project website. Both of these metrics were winsorized so that the top 10% of values were set to equal the value corresponding to the 90th percentile.

Data Analysis

We first calculated descriptive statistics for all variables. We examined differences by recruitment source in participants' demographic and health-related characteristics and device type. To make these comparisons, we used the Pearson chi-square test for categorical dependent variables and 1-way ANOVA (analysis of variance) for continuous dependent variables. We made post hoc pairwise comparisons between the 3 web-based recruitment source groups and used the Bonferroni adjustment to account for multiple comparisons.

We then used regression models to examine differences by recruitment source in study retention and other study engagement metrics. Logistic regression models produced odds ratios (ORs) and 95% CIs for completion of the T2, T3, and T4 surveys. Linear regression models produced standardized β coefficients for the other study engagement metrics. The recruitment source of dating apps served as the referent group in regression models. These statistical tests were 2-tailed with a critical α value of .05. Data were analyzed using R (version 4.2.1; R Foundation for Statistical Computing), and Stata (version 15.0; StataCorp).

Ethical Considerations

The institutional review board at The Ohio State University approved this study (IRB 2019C0028), and the RCT is registered at ClinicalTrials.gov (NCT04032106). We based the reporting of results for this manuscript on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [36] since most of the included data were collected at baseline. Informed consent was obtained from all participants, and all data were deidentified. Participants could earn up to US $95 in gift cards for completing study surveys.

Results

Participant Characteristics

Overall, 52.8% (n=648) of participants indicated a minoritized racial or ethnic identity, 64.4% (n=790) were between 22 and 25 years of age, 69% (n=847) had at least some college education, and 66.4% (n=815) self-identified as gay. Most participants (n=974, 79.4%) had some form of health insurance, but more than half (n=666, 54.3%) had not had a preventive health visit in the prior year.

A total of 403 (32.8%) participants were recruited via social media, 683 (55.7%) participants via dating apps, and 141 (11.5%) participants via other recruitment sources. Most demographic and health-related characteristics differed by recruitment source (Table 1). Generally, a higher proportion of participants recruited via dating apps reported a minoritized racial or ethnic identity, were single and having sex or casually dating, had less education, reported having sex with a male partner in the past, and had public or no health insurance. Conversely, a higher proportion of participants recruited via social media indicated not being currently employed full-time or part-time, self-identified as gay, and reported being HIV-negative. Participants recruited via social media also reported a higher level of connectedness to the LGBTQ community and a higher level of disclosure of their sexual orientation. Table 1 shows the results of all pairwise comparisons.

The type of device used by participants during study enrollment also differed by recruitment source (*P*<.001; Table 1). Smartphone use was higher among participants recruited via dating apps (n=660, 96.6%) compared to those recruited via social media (n=318, 78.9%) or other web-based sources (n=85, 60.3%). Conversely, personal computer use was higher among participants recruited via social media (n=81, 20.1%) or other web-based sources (n=55, 39%) compared to those recruited via dating apps (n=17, 2.5%).

Table 1. Participant characteristics of young gay, bisexual, and other men who have sex with men in the United States by recruitment source for a web-based human papillomavirus vaccination intervention, 2019-2021 (N=1227)[a].

Characteristics	Social media (n=403)	Dating app (n=683)	Other online sources (n=141)	Test statistic	*P* value
Demographic characteristics					
Age (years), n (%)				1.9	.39
18-21	144 (35.7)	250 (36.6)	43 (30.5)		
22-25	259 (64.3)	433 (63.4)	98 (69.5)		
Race or ethnicity, n (%)				55.6	<.001[b,c]
Hispanic	84 (20.8)	244 (35.7)	24 (17)		
Non-Hispanic White	224 (55.6)	280 (41)	75 (53.2)		
Non-Hispanic Black	30 (7.4)	85 (12.5)	14 (9.9)		
Another race or ethnicity	65 (16.1)	74 (10.8)	28 (19.9)		
Relationship status, n (%)				50.5	<.001[b,c,d]
Single and not having sex	57 (14.1)	83 (12.2)	30 (21.3)		
Single and having sex or casually dating	239 (59.3)	492 (72)	62 (44)		
In a relationship	107 (26.6)	108 (15.8)	49 (34.8)		
Education level, n (%)				38.0	<.001[b,c]
High school or less	90 (22.3)	261 (38.2)	29 (20.6)		
Some college or more	313 (77.7)	422 (61.8)	112 (79.4)		
Employment status, n (%)				22.9	<.001[b,d]
Employed full time or part time	253 (62.8)	520 (76.1)	105 (74.5)		
Other	150 (37.2)	163 (23.9)	36 (25.5)		
Region of residence, n (%)				12.1	.06
Northeast	90 (22.3)	113 (16.5)	25 (17.7)		
Midwest	77 (19.1)	114 (16.7)	33 (23.4)		
South	136 (33.8)	245 (35.9)	48 (34)		
West	100 (24.8)	211 (30.9)	35 (24.8)		
Sexual identity, n (%)				40.2	<.001[b,c,d]
Gay	296 (73.5)	449 (65.7)	70 (49.7)		
Bisexual	76 (18.9)	194 (28.4)	47 (33.3)		
Another identity	31 (7.7)	40 (5.9)	24 (17.0)		
Ever had sex with a male, n (%)				43.3	<.001[b,c,d]
No	35 (8.7)	24 (3.5)	26 (18.4)		
Yes	368 (91.3)	659 (96.5)	115 (81.6)		
Sexually attracted to males, n (%)				5.9	.051[c]
No	12 (3.0)	17 (2.5)	9 (6.4)		
Yes	391 (97.0)	666 (97.5)	132 (93.6)		
Connectedness to LGBTQ[e] community, mean (SD)[f]	2.7 (0.8)	2.4 (0.9)	2.3 (0.8)	29.3	<.001[b,d]
Disclosure of sexual orientation, mean (SD)[g]	3.3 (1.6)	3.0 (1.6)	3.1 (1.5)	3.9	.048[b]
Concealment of sexual orientation, mean (SD)[h]	2.5 (1.2)	2.7 (1.2)	2.5 (1.1)	0.4	.53
Health-related characteristics					

Characteristics	Social media (n=403)	Dating app (n=683)	Other online sources (n=141)	Test statistic	*P* value
Health insurance, n (%)				34.1	<.001[b,c]
Private insurance	285 (70.7)	379 (55.5)	104 (73.8)		
Public insurance	53 (13.2)	134 (19.6)	19 (13.5)		
None or do not know	65 (16.1)	170 (24.9)	18 (12.8)		
Last preventive health visit, n (%)				0.7	.72
Within the last year	187 (46.4)	306 (44.8)	68 (48.2)		
More than a year ago	216 (53.6)	377 (55.2)	73 (51.8)		
HIV status, n (%)				7.7	.02[b]
Negative	392 (97.3)	638 (93.4)	133 (94.3)		
Positive	11 (2.7)	45 (6.6)	8 (5.7)		
Electronic health literacy, mean (SD)[i]	4.0 (0.7)	3.9 (0.8)	3.9 (0.8)	2.8	.10
Device information					
Device type, n (%)				175.9	<.001[b,c,d]
Smartphone	318 (78.9)	660 (96.6)	85 (60.3)		
Tablet device	4 (1)	6 (0.9)	1 (0.7)		
Personal computer	81 (20.1)	17 (2.5)	55 (39)		

[a] Due to rounding, percentages may not total 100%. Totals may not equal N=1227 due to missing data. Reported test statistics and *P* values are from Pearson chi-square test for categorical variables and 1-way ANOVA (analysis of variance) for continuous variables. Superscript letters b, c, and d indicate statistically significant differences of post hoc pairwise comparisons using the Bonferroni adjustment to account for multiple comparisons.

[b] Social media compared to dating app.

[c] Dating app compared to other web-based source.

[d] Social media compared to other web-based sources.

[e] LGBTQ: lesbian, gay, bisexual, transgender, queer.

[f] 2-item scale; items had a 4-point response scale ranging from 1="not at all" to 4="a lot."

[g] 3-item scale; items had a 5-point response scale ranging from 1="none" to 5="all."

[h] 3-item scale; items had a 5-point response scale ranging from 1="never" to 5="always."

[i] 4-item scale; items had a 5-point response scale ranging from 1="strongly disagree" to 5="strongly agree."

Study Engagement

Participant retention overall was 90% for the T2 survey (n=1001), 70% for the T3 survey (n=858), and 69% for the T4 survey (n=842). Compared to participants recruited via dating apps, retention was higher among participants recruited via social media for the T2 survey (94% vs 87%; OR 2.09, 95% CI 1.32-3.30), the T3 survey (75% vs 66%; OR 1.52, 95% CI 1.15-2.00), and the T4 survey (75% vs 65%; OR 1.66, 95% CI 1.26-2.19; Figure 1). There were no statistically significant differences in retention for participants recruited via other web-based sources.

Among participants in the 2 intervention groups (n=815), the average amount of time spent viewing intervention content about HPV vaccination was 2.83 minutes. Participants recruited via social media spent a longer amount of time viewing this content compared to those recruited via dating apps (3.14 minutes vs 2.67 minutes; 95% CI 0.03-0.33; *P*=.02; Table 2). The average number of project website logins was 4.94, and there were no differences in the number of logins by recruitment source (*P*=.83 for social media, *P*=.54 for other web-based sources; Table 2).

Figure 1. Survey retention of young gay, bisexual, and other men who have sex with men in the United States by recruitment source for a web-based human papillomavirus vaccination intervention (N=1227). Error bars represent SE.

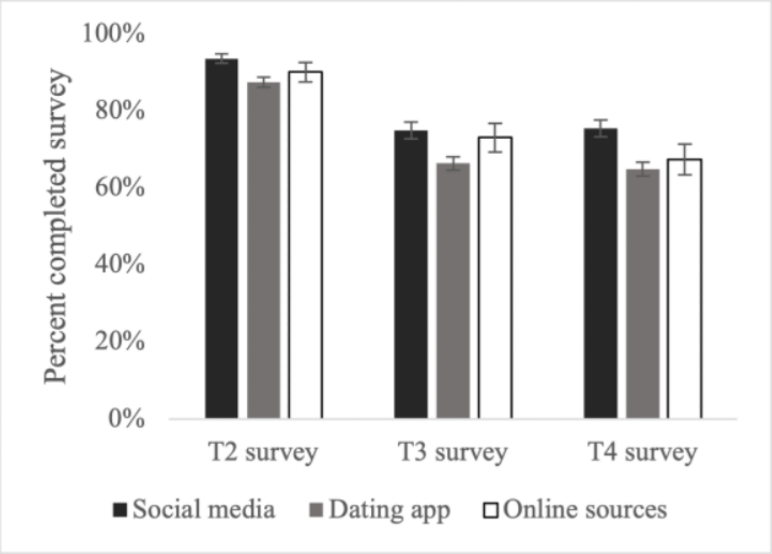


Table 2. Study engagement of young gay, bisexual, and other men who have sex with men in the United States by recruitment source for a web-based human papillomavirus vaccination intervention (n=815)[a,b].

	Time spent on intervention content (minutes)		Number of project website log-ins	
	Mean (SD)	β (95% CI)	Mean (SD)	β (95% CI)
Dating app	2.67 (2.44)	Reference	4.98 (2.40)	Reference
Social media	3.14 (2.80)	0.18[b] (0.03-0.33)	4.94 (2.42)	–0.02 (–0.17 to 0.13)
Other web-based sources	2.74 (2.56)	0.02 (–0.20 to 0.25)	4.81 (2.17)	–0.07 (–0.29 to 0.15)

[a] Table reports data for the 815 participants in the 2 intervention groups. β represents standardized regression coefficients from linear regression models.

[b] Statistically significant at the α=.05 level.

Discussion

Principal Results

With digital methodologies becoming increasingly popular for recruiting young gay, bisexual, and other men who have sex with men for research studies, [17-21] it is important to examine if different web-based recruitment sources yield varying types of participants. The Outsmart HPV study provided an opportunity to address this research topic by recruiting a national sample of young gay, bisexual, and other men who have sex with men through multiple web-based sources, and there are 2 main findings. First, most of the demographic and health-related characteristics examined differed by recruitment sources. Participants recruited via social media tended to be less diverse (eg, race or ethnicity, sexual self-identity), of higher socioeconomic status (eg, education level, health insurance status), and have higher levels of outness (eg, disclosure of sexual orientation, connectedness to the LGBTQ community) compared to participants recruited via dating apps. This pattern differs from a past study that found participants recruited from a social media site (Facebook) were highly similar to those recruited from a dating app (Grindr) in terms of race or ethnicity [18]. Interestingly, the characteristics of participants recruited via social media in our study better align with the characteristics of other national samples of young gay, bisexual, and other men who have sex with men and the larger gay, bisexual, and other men who have sex with men community in the United States than participants recruited via dating apps in our study [37-40]. A potential explanation for this pattern is that users of dating apps tend to be more demographically diverse compared to the general population [41]. Interestingly, some characteristics (eg, socioeconomic status) of participants that were recruited via social media for our study were more comparable to gay, bisexual, and other men who have sex with men recruited through in-person venues in past research than via dating apps in our study [42].

Second, study engagement also differed by recruitment source. Participants who were recruited via dating apps tended to interact less with study content and were less likely to complete follow-up surveys compared to participants recruited via social media. This may be attributable to users of dating apps having an emphasis on immediate gratification [43] and impulsivity [44] as a justification for their app usage (and study enrollment), which in turn may affect their awareness of and interest in engaging with study content and completing study-related activities (eg, follow-up surveys). Related, given that dating app users are more impulsive than nonusers, [44] young gay, bisexual, and other men who have sex with men who use dating apps might experience an initial interest in joining a research study and earning an initial study incentive. However, interest in completing study-related activities at later timepoints may not be as strong as their general preference for some level of anonymity [45] might dissuade them from wanting to continue discussing their health behaviors (eg, sexual history, vaccination status).

We think these 2 main findings can help researchers better plan and customize future digital recruitment efforts of young gay, bisexual, and other men who have sex with men by highlighting key issues and considerations. For example, researchers will now be able to better anticipate how the selection of web-based recruitment sources may affect the representativeness and diversity (and the potential balance between the two) of their study sample. Recruiting through social media may lead to a sample of participants that better reflect the larger young gay, bisexual, and other men who have sex with men population in the United States, but dating apps may be more effective in reaching more diverse young gay, bisexual, and other men who have sex with men, including those from communities that are often underrepresented in research studies [46]. Recruitment goals differ across studies in terms of representativeness and diversity, and our data on differences in demographic and health-related characteristics can help inform future efforts. Furthermore, studies involving interventions and/or longitudinal data collection should consider how various web-based recruitment sources can impact study retention and participants' engagement with study content. Some studies may want to maximize participant retention and engagement (ie, suggesting that recruitment primarily through social media may be appropriate), whereas other studies may be willing to have slightly lower study engagement metrics in exchange for being able to include additional recruitment sources, such as dating apps. Future studies that do include recruitment from dating apps may also want to consider prioritizing brief study-related activities and other strategies to promote engagement.

One final implication involves the types of devices that were used by participants in our study. Overall, a majority of participants used a smartphone to enroll in the study, including nearly all participants who were recruited via dating apps. This pattern is similar to national data showing that young adults use smartphones much more frequently than other device types [47,48]. Thus, it is important that future digital interventions and other research studies for young adults, including young gay, bisexual, and other men who have sex with men, ensure that study content and data collection instruments are optimized for use on smartphones. For example, to increase usability, it has been suggested that the interface of smartphone-based interventions have a minimal number of screens, limited manual data entry, limited pop-ups and notifications, and personal identification number-based entry (as compared to full passwords). The content of such interventions should be succinct with easy-to-read graphics [49]. Data collection instruments (eg, surveys) should have limited open-ended questions and different "pages" in the survey, rather than having participants scroll through one continuous page. Lastly, and perhaps most generally, mobile surveys should be pilot-tested prior to the start of data collection activities [50]. As smartphone usage continues to be a common approach for participants enrolling in and interacting with digital interventions, such recommendations can help optimize intervention delivery and data collection.

Limitations

Although we were able to compare different groupings of web-based recruitment sources, we were not able to examine potential differences between individual platforms (eg, Facebook vs Twitter vs Grindr) due to modest sample sizes for some of the platforms. These sample sizes also did not allow us to examine how differences between recruitment sources may have changed over time during the recruitment period for our study. Examining potential differences between individual platforms and potential temporal changes are important areas for future research. Our recruitment focused on the social media and dating apps that were among the most popular among young adults in the United States around the time of the start of our study [51,52]. It is likely that the use and popularity of various platforms will evolve over time, and future efforts should examine new and emerging platforms. We were not able to examine advanced metrics related to recruitment efficiency and cost due to a lack of available data. However, past studies have reported such information for various web-based platforms [17,27]. Study recruitment occurred both prior to and during the COVID-19 pandemic, and it is not known how the pandemic may have affected the interest of young gay, bisexual, and other men who have sex with men to enroll in a study focusing on vaccination (which would subsequently affect the characteristics of study participants). Fraudulent accounts with inaccurate self-reported data occur in digital studies, but we used several recommended strategies for detecting such accounts and minimizing this risk [53,54].

Conclusions

We were able to recruit a large national sample of young gay, bisexual, and other men who have sex with men for an HPV vaccination intervention via multiple web-based methodologies. Participants differed on a range of demographic and health-related characteristics based on whether they were recruited from social media, a dating app, or some other web-based recruitment source. Study retention and engagement with study content also differed by web-based recruitment source. These findings and patterns highlight key issues and potential tradeoffs that can help researchers better plan and customize future web-based recruitment efforts of young gay, bisexual, and other men who have sex with men.

Acknowledgments

The research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (R37CA226682). Additional support was provided by the Center for Health Communications Research at the University of Michigan (P30CA046592) and the Recruitment, Intervention, and Survey Shared Resource at The Ohio State University

Comprehensive Cancer Center (P30CA016058). Generative artificial intelligence was not used at any point in the development of this manuscript. The authors would also like to acknowledge the contributions of Annie-Laurie McRee to this research project.

Data Availability

The data for this study may be made available from the corresponding author on reasonable request.

Authors' Contributions

DJM, ALG, and PLR created the original concept and design of this manuscript, with DJM writing the initial draft and performing the initial analyses under the supervision of ALG and PLR. DJM, ALG, MLK, JAB, ABS, and PLR all contributed to the final analysis and interpretation of the data, with all authors also providing critical revision and feedback on the final drafts. All authors reviewed the final manuscript. All authors approve of the final version to be published and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. About genital HPV infection. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/sti/about/about-genital-hpv-infection.html [accessed 2024-04-15]
2. HPV and cancer. National Cancer Institute. 2023. URL: https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-and-cancer [accessed 2023-04-17]
3. Wei F, Gaisa MM, D'Souza G, Xia N, Giuliano AR, Hawes SE, et al. Epidemiology of anal human papillomavirus infection and high-grade squamous intraepithelial lesions in 29900 men according to HIV status, sexuality, and age: a collaborative pooled analysis of 64 studies. Lancet HIV 2021;8(9):e531-e543 [FREE Full text] [doi: 10.1016/S2352-3018(21)00108-9] [Medline: 34339628]
4. Machalek DA, Poynten M, Jin F, Fairley CK, Farnsworth A, Garland SM, et al. Anal human papillomavirus infection and associated neoplastic lesions in men who have sex with men: a systematic review and meta-analysis. Lancet Oncol 2012;13(5):487-500. [doi: 10.1016/S1470-2045(12)70080-3] [Medline: 22445259]
5. Meites E, Wilkin TJ, Markowitz LE. Review of human papillomavirus (HPV) burden and HPV vaccination for gay, bisexual, and other men who have sex with men and transgender women in the United States. Hum Vaccin Immunother 2022;18(1):2016007 [FREE Full text] [doi: 10.1080/21645515.2021.2016007] [Medline: 35294325]
6. Meites E, Szilagyi PG, Chesson HW, Unger ER, Romero JR, Markowitz LE. Human papillomavirus vaccination for adults: updated recommendations of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep 2019;68(32):698-702 [FREE Full text] [doi: 10.15585/mmwr.mm6832a3] [Medline: 31415491]
7. Reiter PL, Gower AL, Kiss DE, Malone MA, Katz ML, Bauermeister JA, et al. A web-based human papillomavirus vaccination intervention for young gay, bisexual, and other men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2020;9(2):e16294 [FREE Full text] [doi: 10.2196/16294] [Medline: 32130192]
8. Gerend MA, Madkins K, Crosby S, Korpak AK, Phillips GL, Bass M, et al. Evaluation of a text messaging-based human papillomavirus vaccination intervention for young sexual minority men: results from a pilot randomized controlled trial. Ann Behav Med 2021;55(4):321-332 [FREE Full text] [doi: 10.1093/abm/kaaa056] [Medline: 32914838]
9. Fontenot HB, White BP, Rosenberger JG, Lacasse H, Rutirasiri C, Mayer KH, et al. Mobile app strategy to facilitate human papillomavirus vaccination among young men who have sex with men: pilot intervention study. J Med Internet Res 2020;22(11):e22878 [FREE Full text] [doi: 10.2196/22878] [Medline: 33146621]
10. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Increasing human papillomavirus vaccination among young gay and bisexual men: a randomized pilot trial of the Outsmart HPV intervention. LGBT Health 2018;5(5):325-329 [FREE Full text] [doi: 10.1089/lgbt.2018.0059] [Medline: 29979642]
11. Meyer IH, Wilson PA. Sampling lesbian, gay, and bisexual populations. J Counsel Psychol 2009;56(1):23-31. [doi: 10.1037/a0014587]
12. Guillory J, Wiant KF, Farrelly M, Fiacco L, Alam I, Hoffman L, et al. Recruiting hard-to-reach populations for survey research: using Facebook and Instagram advertisements and in-person intercept in LGBT bars and nightclubs to recruit LGBT young adults. J Med Internet Res 2018;20(6):e197 [FREE Full text] [doi: 10.2196/jmir.9461] [Medline: 29914861]
13. 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;22(7):2296-2311. [doi: 10.1007/s10461-018-2071-z] [Medline: 29470808]
14. López DJ, García A, Engel R, Jackson JL. Challenges and opportunities to recruiting and engaging with gay male Latino sexual assault survivors. Field Methods 2022;35(1):48-63. [doi: 10.1177/1525822x221103736]

15. Ellard-Gray A, Jeffrey NK, Choubak M, Crann SE. Finding the hidden participant. Int J Qual Methods 2015;14(5):160940691562142. [doi: 10.1177/1609406915621420]
16. Sanchez T, Smith A, Denson D, Dinenno E, Lansky A. Internet-based methods may reach higher-risk men who have sex with men not reached through venue-based sampling. Open AIDS J 2012;6:83-89 [FREE Full text] [doi: 10.2174/1874613601206010083] [Medline: 23049657]
17. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Recruiting young gay and bisexual men for a human papillomavirus vaccination intervention through social media: the effects of advertisement content. JMIR Public Health Surveill 2017;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]
18. Buckingham L, Becher J, Voytek CD, Fiore D, Dunbar D, Davis-Vogel A, et al. Going social: success in online recruitment of men who have sex with men for prevention HIV vaccine research. Vaccine 2017;35(27):3498-3505 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.002] [Medline: 28526330]
19. Wilkerson JM, Shenk JE, Grey JA, Rosser BRS, 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]
20. Brown JR, Reid D, Howarth AR, Mohammed H, Saunders J, Pulford CV, et al. Difficulty accessing condoms because of the COVID-19 pandemic reported by gay, bisexual and other men who have sex with men in the UK: findings from a large, cross-sectional, online survey. Int J STD AIDS 2023;34(8):541-547 [FREE Full text] [doi: 10.1177/09564624231160804] [Medline: 36943893]
21. Sun CJ, Sutfin E, Bachmann LH, Stowers J, Rhodes SD. Comparing men who have sex with men and transgender women who use Grindr, other similar social and sexual networking apps, or no social and sexual networking apps: implications for recruitment and health promotion. J AIDS Clin Res 2018;9(2):757 [FREE Full text] [doi: 10.4172/2155-6113.1000757] [Medline: 29593933]
22. Social media fact sheet. Pew Research Center. 2024 Nov 13. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2024-11-13]
23. Badal HJ, Stryker JE, DeLuca N, Purcell DW. Swipe right: dating website and app use among men who have sex with men. AIDS Behav 2018;22(4):1265-1272. [doi: 10.1007/s10461-017-1882-7] [Medline: 28884248]
24. 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;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
25. Merchant RC, Romanoff J, Clark MA, Liu T, Rosenberger JG, Bauermeister J, et al. Variations in recruitment yield and characteristics of participants recruited across diverse internet platforms in an HIV testing study of young adult men-who-have-sex-with-men (YMSM). Am J Mens Health 2017;11(5):1342-1357 [FREE Full text] [doi: 10.1177/1557988317717383] [Medline: 28691552]
26. Moseson H, Kumar S, Juusola JL. Comparison of study samples recruited with virtual versus traditional recruitment methods. Contemp Clin Trials Commun 2020;19:100590 [FREE Full text] [doi: 10.1016/j.conctc.2020.100590] [Medline: 32637722]
27. Parker JN, Hunter AS, Bauermeister JA, Bonar EE, Carrico A, Stephenson R. Comparing social media and in-person recruitment: lessons learned from recruiting substance-using, sexual and gender minority adolescents and young adults for a randomized control trial. JMIR Public Health Surveill 2021;7(12):e31657 [FREE Full text] [doi: 10.2196/31657] [Medline: 34855613]
28. 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;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]
29. Sharma P, Lamb LE, Bartolone SN, Ward EP, Janicki JJ, Peters KM, et al. A cost, time, and demographic analysis of participant recruitment and urine sample collection through social media optimization. Urol Pract 2022;9(6):561-566. [doi: 10.1097/upj.0000000000000339]
30. Gibbs JJ, Schrager SM, Goldbach JT. Young men who have sex with men and substance use: a comparison of venue-based sampling and geosocial networking application sampling. J Homosex 2024;71(5):1163-1176. [doi: 10.1080/00918369.2022.2161086] [Medline: 36630486]
31. Lachowsky NJ, Lal A, Forrest JI, Card KG, Cui Z, Sereda P, et al. Including online-recruited seeds: a respondent-driven sample of men who have sex with men. J Med Internet Res 2016;18(3):e51 [FREE Full text] [doi: 10.2196/jmir.5258] [Medline: 26980147]
32. Bezerra DRB, Jalil CM, Jalil EM, Coelho LE, Netto EC, Freitas J, et al. Comparing web-based venues to recruit gay, bisexual, and other cisgender men who have sex with men to a large HIV prevention service in Brazil: evaluation study. JMIR Form Res 2022;6(8):e33309 [FREE Full text] [doi: 10.2196/33309] [Medline: 35925658]
33. Bender JL, Akinnibosun R, Puri N, D'Agostino N, Drake EK, Tsimicalis A, et al. A comparison of the sociodemographic, medical, and psychosocial characteristics of adolescents and young adults diagnosed with cancer recruited in-person and online: a Canadian cross-sectional survey. Digit Health 2023;9:20552076231205278 [FREE Full text] [doi: 10.1177/20552076231205278] [Medline: 37900258]
34. Meidlinger PC, Hope DA. Differentiating disclosure and concealment in measurement of outness for sexual minorities: the Nebraska Outness Scale. P Sex Orientation Gender Divers 2014;1(4):489-497. [doi: 10.1037/sgd0000080]

35. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
36. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147(8):573-577 [FREE Full text] [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
37. LGBT Demographic Data Interactive. The Williams Institute, UCLA School of Law. 2019 Jan. URL: https://williamsinstitute.law.ucla.edu/visualization/lgbt-stats/?topic=LGBT#about-the-data [accessed 2024-03-01]
38. Cole SW, Glick JL, Campoamor NB, Sanchez TH, Sarkar S, Vannappagari V, et al. Willingness and preferences for long-acting injectable PrEP among US men who have sex with men: a discrete choice experiment. BMJ Open 2024;14(4):e083837 [FREE Full text] [doi: 10.1136/bmjopen-2023-083837] [Medline: 38653510]
39. Strutz KL, Herring AH, Halpern CT. Health disparities among young adult sexual minorities in the U.S. Am J Prev Med 2015;48(1):76-88 [FREE Full text] [doi: 10.1016/j.amepre.2014.07.038] [Medline: 25241194]
40. Reiter PL, McRee A, Katz ML, Paskett ED. Human papillomavirus vaccination among young adult gay and bisexual men in the United States. Am J Public Health 2015;105(1):96-102. [doi: 10.2105/AJPH.2014.302095] [Medline: 25393178]
41. Vogels EA, McClain C. Key findings about online dating in the U.S. Pew Research Center. 2023 Feb 02. URL: https://www.pewresearch.org/short-reads/2023/02/02/key-findings-about-online-dating-in-the-u-s [accessed 2023-02-02]
42. Salway TJ, Morgan J, Ferlatte O, Hawkins B, Lachowsky NJ, Gilbert M. A systematic review of characteristics of nonprobability community venue samples of sexual minority individuals and associated methods for assessing selection bias. LGBT Health 2019;6(5):205-215. [doi: 10.1089/lgbt.2018.0241] [Medline: 31135260]
43. Gudelunas D. There’s an app for that: the uses and gratifications of online social networks for gay men. Sex Cult 2012;16(4):347-365. [doi: 10.1007/s12119-012-9127-4]
44. Sawyer AN, Smith ER, Benotsch EG. Dating application use and sexual risk behavior among young adults. Sex Res Soc Policy 2017;15(2):183-191. [doi: 10.1007/s13178-017-0297-6]
45. Waldman AE. Navigating privacy on gay-oriented mobile dating applications. In: Bailey J, Flynn A, Henry N, editors. The Emerald International Handbook of Technology-Facilitated Violence and Abuse. Bingley, United Kingdom: Emerald Publishing Limited; 2021.
46. Allison K, Patel D, Kaur R. Assessing multiple factors affecting minority participation in clinical trials: development of the clinical trials participation barriers survey. Cureus 2022;14(4):e24424 [FREE Full text] [doi: 10.7759/cureus.24424] [Medline: 35637812]
47. Cheung M, Lai JSK, Yip J. Influences of smartphone and computer use on health-related quality of life of early adolescents. Int J Environ Res Public Health 2022;19(4):2100 [FREE Full text] [doi: 10.3390/ijerph19042100] [Medline: 35206289]
48. Faverio M, Anderson M. For shopping, phones are common and influencers have become a factor—especially for young adults. Pew Research Center. 2022 Nov 21. URL: https://www.pewresearch.org/short-reads/2022/11/21/for-shopping-phones-are-common-and-influencers-have-become-a-factor-especially-for-young-adults/ [accessed 2022-11-21]
49. Gomolin A, Lebouché B, Engler K, Vedel I. Optimizing smartphone intervention features to improve chronic disease management: a rapid review. Health Informatics J 2020;26(3):1795-1809 [FREE Full text] [doi: 10.1177/1460458219891377] [Medline: 31826703]
50. Mobile survey optimization. Qualtrics. URL: https://www.qualtrics.com/support/survey-platform/survey-module/mobile-survey-optimization/ [accessed 2024-12-04]
51. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-04-07]
52. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. 2019 Apr 10. URL: https://www.pewresearch.org/short-reads/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-04-10]
53. Teitcher JEF, 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]
54. Xu Y, Pace S, Kim J. Threats to online surveys: recognizing, detecting, and preventing survey bots. Soc Work Res 2022;46(4):343-350. [doi: 10.1093/swr/svac023]

Abbreviations

ANOVA: analysis of variance
HPV: human papillomavirus
LGBTQ: lesbian, gay, bisexual, transgender, and queer
RCT: randomized controlled trial
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Mavragani; submitted 23.07.24; peer-reviewed by K Card, M Lim; comments to author 14.10.24; revised version received 04.11.24; accepted 24.11.24; published 03.01.25.

Please cite as:
Marshall DJ, Gower AL, Katz ML, Bauermeister JA, Shoben AB, Reiter PL
Recruitment of Young Gay, Bisexual, and Other Men Who Have Sex With Men for a Web-Based Human Papillomavirus Vaccination Intervention: Differences in Participant Characteristics and Study Engagement by Recruitment Source in a Randomized Controlled Trial
J Med Internet Res 2025;27:e64668
URL: https://www.jmir.org/2025/1/e64668
doi:10.2196/64668
PMID:

Original Paper

Urban-Suburban Differences in Public Perspectives on Digitalizing Pediatric Research: Cross-Sectional Survey Study

Heping Fang[1], MD, PhD; Ruoling Xian[2], MMeD; Juan Li[1], MD, PhD; Yingcun Li[3], MD, PhD; Enmei Liu[1], MD, PhD; Yan Zhao[2*], BD; Yan Hu[4,5*], MD, PhD

[1]Department of Respiratory Medicine, Pediatric Research Institute, Children's Hospital of Chongqing Medical University, Chongqing Key Laboratory of Child Rare Diseases in Infection and Immunity, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, Chongqing, China

[2]Department of Pediatrics, Chongqing Health Center for Women and Children, Women and Children's Hospital of Chongqing Medical University, Chongqing, China

[3]Department of General Surgery, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

[4]Department of Child Health Care, Children's Hospital of Chongqing Medical University, Chongqing Key Laboratory of Child Rare Diseases in Infection and Immunity, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, Chongqing, China

[5]Children's Health Care Center, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yan Hu, MD, PhD
Department of Child Health Care, Children's Hospital of Chongqing Medical University
Chongqing Key Laboratory of Child Rare Diseases in Infection and Immunity, Ministry of Education Key Laboratory of Child Development and Disorders
National Clinical Research Center for Child Health and Disorders
No. 20 Jinyu Avenue
Liangjiang New Area
Chongqing, 401122
China
Phone: 86 02368370551
Email: hy420@126.com

Abstract

Background: Recruiting and retaining participants in pediatric research has always been challenging, particularly in healthy populations and remote areas, leading to selection bias and increased health disparities. In the digital age, medical research has been transformed by digital tools, offering new opportunities to enhance engagement in clinical research. However, public perspectives on digitalizing pediatric research and potential differences between urban and suburban areas remain unclear.

Objective: This study aimed to investigate public perspectives on digitalizing pediatric research and compare differences between urban and suburban areas to help diversify participants and address health disparities.

Methods: A cross-sectional web-based survey targeting caregivers of kindergarten children (aged 2-7 years) in Chongqing was conducted between June and December 2023. A total of 4231 valid questionnaires were analyzed, with 25.1% (n=1064) of the children residing in urban areas and 74.9% (n=3167) in suburban areas. Descriptive statistics and intergroup comparisons were used for data analysis.

Results: Approximately 59.8% (n=2531) of the caregivers had first impressions of pediatric research, with 36.9% (n=1561) being positive and 22.9% (n=970) being negative. A total of 38.3% (n=1621) of caregivers recognized the growing popularity of digital tools, and 36.7% (n=1552) supported their use in pediatric research, but only 25.2% (n=1068) favored online-only research methods. The main concerns regarding the use of software in pediatric research were privacy issues (n=3273, 77.4%) and potential addiction (n=2457, 58.1%). Public accounts of research institutions (n=3400, 80.4%) were the most favored for online recruitment. Telephones (1916/3076, 62.3%) and social media apps (1801/3076, 58.6%) were the most popular for regular contact. Intergroup comparisons revealed that suburban caregivers had more positive first impressions of pediatric research (38.6% vs 32%; *P*<.001; adjusted odds ratio [aOR] 1.27, 95% CI 1.09-1.47) and faced fewer participation barriers: "worry about being an experimental subject" (70.9% vs 76.6%; *P*<.001; aOR 0.79, 95% CI 0.67-0.93), "pose a risk to children's health" (58.6% vs

67.8%; *P*<.001; aOR 0.71, 95% CI 0.61-0.83), “do not have enough background information” (55.2% vs 61.6%; *P*<.001; aOR 0.78, 95% CI 0.67-0.89), and “worry about recommending other products” (48.2% vs 55%; *P*<.001; aOR 0.78, 95% CI 0.67-0.89). They also showed greater support for online-only research methods (26% vs 22.9%; *P*=.045; aOR 1.19, 95% CI 1.01-1.41) and greater openness to unofficial online recruitment sources (social media friends: 24.7% vs 18.9%; *P*<.001; aOR 1.33, 95% CI 1.11-1.59; moments on social media: 15.5% vs 11.1%; *P*<.001; aOR 1.35, 95% CI 1.09-1.67).

Conclusions: In the digital age, enhancing recruitment and retention in pediatric research can be achieved by integrating both official and unofficial social media strategies, implementing a hybrid online-offline follow-up approach, and addressing privacy concerns.

(J Med Internet Res 2025;27:e60324) doi:10.2196/60324

KEYWORDS

pediatrics; pediatric research; digital health; public opinion; research; patient participation; urban; rural; caregiver attitudes; social media; mobile phone

Introduction

Recruitment and retention in pediatric research have consistently posed challenges [1-3]. In the United States, pediatric randomized clinical trials are often discontinued or unpublished, with patient recruitment difficulties (37%) being the primary reason [4]. Our experiences further corroborate this trend. For instance, in a cross-sectional hospital-based survey, we recruited only 65.8% of the caregivers [5]. Similarly, in a longitudinal study on the growth of critically ill children after liver transplantation, follow-up was completed for only 68.6% of the children one year after the procedure [6]. Recruitment and retention challenges frequently prevent studies from achieving their preset objectives, resulting in inefficiency and wasted personnel and financial resources.

While traditional research has focused on children with evident illnesses, emerging concerns such as overweight, internet addiction, and emotional or behavioral issues [7-10] highlight the importance of studying seemingly healthy children [11-13]. However, recruiting and retaining healthy children for research is more challenging than for those with specific diseases, as shown by pediatric reference interval studies [14]. Additionally, children in remote areas are frequently underrepresented in research [15-17], leading to selection bias and exacerbating health disparities. In China, geographical location has been a primary longstanding contributor to child health inequity [18]. Therefore, researchers need to devise strategies to address these challenges to meet the increasing demands of pediatric research. In this context, obtaining participants’ perspectives by seeking their opinions may provide more practical solutions.

Caregivers, typically parents, are key decision makers regarding younger children’s participation in research [19]. Numerous studies have examined parental decision-making in enrolling children in pediatric research, primarily focusing on traditional research methods [20-27]. In the digital age, medical research has been significantly transformed by digital tools and the process of digitalization [28,29]. Digital medicine has created new opportunities to enhance engagement in clinical research, especially in healthy populations and remote areas, which improves health outcomes for participants [30-32]. This study investigated public perspectives on pediatric research in the digital age and examined potential differences between urban and suburban areas. The goal was to provide insights to advance the digitalization of pediatric research, which may diversify the pool of participants and address health disparities among children.

Methods

Study Design and Participants

This study was a cross-sectional, web-based questionnaire survey. The study targeted caregivers of kindergarten children in Chongqing, a municipality in southwestern China, with a total of 12.58 million households and 995,239 kindergarten children. Chongqing is divided into urban and suburban areas based on geographical location and development level, with 3.07 million (23.9%) households in urban areas and 9.79 million (76.1%) in suburban areas [33]. Kindergarten children are typically between 3 and 6 years old, with exceptions depending on individual circumstances [34]. The inclusion criteria for kindergarten children were (1) residency in Chongqing, (2) kindergarten attendance, (3) aged 2 to 7 years, and (4) voluntary completion of the questionnaire by the caregiver. The exclusion criteria were (1) failure to pass validation questions, (2) logical errors in responses, and (3) self-reported lack of seriousness in answering.

Ethical Considerations

This study was approved by the Ethics Committee of the Children’s Hospital of Chongqing Medical University (Protocol #2023-236), with informed consent signatures waived. The collected data were anonymized, and no compensation was given.

Reporting Guidelines

The CHERRIES (Checklists for Reporting Results of Internet E-Surveys) and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines were followed for reporting the results [35,36].

Research Outcomes and Measurements

The primary outcome of this study was public perspectives on pediatric research, with differences between urban and suburban areas as the secondary outcome. These perspectives were divided into 4 key categories: “Facilitators and Barriers to Participation,” “Perspectives on Digital Medicine,” “Perspectives on Recruitment,” and “Perspectives on the Research Process.” Each category comprised single-choice and multiple-choice questions designed to collect data. In the intergroup comparison, participants were classified as urban or suburban (exposure) based on their responses to a residential address question (Multimedia Appendix 1).

Development of the Questionnaire

The questionnaire was initially developed and refined based on the literature and clinical experience [20-27,37-39], then revised and culturally adapted by 7 pediatric doctors with diverse professional backgrounds (see Acknowledgments). The final version was converted to a web-based format using “Lediaocha,” a web-based survey platform, with 50 mandatory questions spread across 4 pages. Of these, 11 questions focused on general characteristics, while the rest investigated perspectives on pediatric research, particularly digital medicine. To ensure data reliability, one question on caregivers’ roles was repeated at the beginning and end of the questionnaire, along with a self-assessment query on response seriousness before submission. Participation was voluntary, anonymous, and accessible through WeChat (Tencent) on smartphones. Device IDs were recorded to prevent duplicate entries. When caregivers started answering, the system marked them as “in progress.” Although it was technically feasible to save “in progress” answers automatically, ethical concerns led to a decision not to do so, as these were not final submissions. The process of developing questionnaires and the methodological considerations for web-based surveys followed our team’s established practices [40].

Determination of the Sample Size

Due to the lack of a clear sample size calculation method for multioutcome survey studies, we used the estimation method from quality-of-life studies, which suggests 5-10 questionnaires per question [41]. Therefore, with 50 questions, a minimum of 500 questionnaires were required. To mitigate potential selection bias [42], the minimum sample size was doubled to 1000. Further calculations took into account the number of households in Chongqing, with questionnaires planned to be collected at a 1:3 ratio between urban and suburban areas. As a result, a total of 4000 questionnaires were needed, with at least 1000 from urban areas and 3000 from suburban areas. Given the possibility of a high invalid response rate (approximately 50%) in public surveys [43], at least 8000 households were estimated to be needed. This sample size represented 0.4% of the entire population of kindergarten children in Chongqing, placing the sampling ratio between the national population dynamics survey (0.1%) and the sample survey (1%). Finally, the actual proportions of recruited participants from urban and suburban areas were 25.1% and 74.9%, respectively, which is consistent with the planned sample distribution.

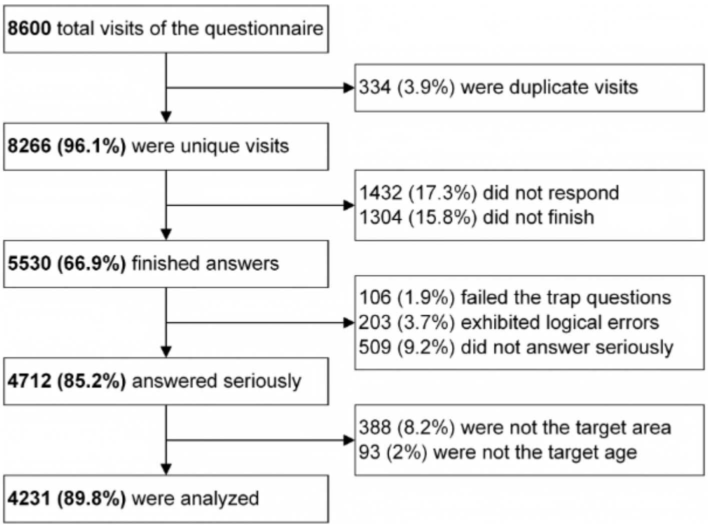
Survey Process

The survey was conducted between June and December 2023. Prior to the official launch, a pilot test with 50 participants was conducted to validate the survey process. A flexible survey was then conducted in urban and suburban areas over 20 days to accommodate different kindergarten schedules. The questionnaire was distributed to kindergartens and forwarded to caregivers, ensuring that recruitment focused on the target audience and minimized nontarget responses. Forwarding questionnaires by kindergarten teachers was not mandatory. Caregivers could view a short introduction and an anonymous, voluntary declaration of the study on the questionnaire homepage to avoid any sense of obligation. After the survey, researchers downloaded and saved the data from the cloud servers and then removed them from the cloud.

Data Processing

The collected questionnaires were assessed for adherence to the inclusion and exclusion criteria. Specifically, 8600 online visits were documented during the survey, with 8266 (96.1%) identified as unique after eliminating duplicates. Among the visitors, 6834 started the questionnaire (response rate of 82.7%), and 5530 completed it (completion rate of 66.9%). After unreliable and nontarget questionnaires were excluded, 4231 questionnaires were ultimately analyzed (inclusion rate of 76.5%), representing 0.43% of the entire population of kindergarten children (Figure 1).

Figure 1. The inclusion and exclusion process.



Statistical Analysis

Data analysis was performed using SPSS (version 25; IBM Corp). Qualitative data are presented as frequencies (percentages), while quantitative data are presented as medians (IQRs) after tests for normality. Intergroup comparisons were conducted using chi-square and Mann-Whitney tests. Multivariate logistic regression was used to adjust for potential confounders, including child age and sex, caregiver age, role, education level, and family size, when examining the relationship between region (independent variable) and public perspectives (outcome variables). Stepwise methods were not applied (Multimedia Appendix 2). The results of the logistic regression are presented as adjusted odds ratios (aORs) with 95% CIs. A *P* value less than .05 was considered statistically significant.

Results

General Characteristics

Table 1 presents the general characteristics of the children and caregivers. Among the children, 1064 (25.1%) resided in urban areas, while 3167 (74.9%) children were from suburban areas. Most children (n=3774, 89.2%) had never participated in pediatric research, while only 7.7% (n=326) and 5.4% (n=230) had taken part in offline and online research, respectively. Interestingly, suburban children had a greater proportion of online research participation than urban children (6% vs 3.9%; *P*=.008). Among the caregivers, 3547 (83.8%) were mothers, 2334 (55.2%) had an education level of senior high school or below, and 2297 (54.3%) lived in households with 4 or fewer members.

Table 1. General characteristics of the children and caregivers.

General characteristics	Overall	Urban	Suburban	*P* value
Age of the children (years), median (IQR)	4.8 (4.1-5.6)	4.6 (4.0-5.5)	4.9 (4.2-5.7)	<.001
Sex of the children, n (%)				.44
Male	2167 (51.2)	534 (50.2)	1633 (51.6)	
Female	2064 (48.8)	530 (49.8)	1534 (48.4)	
Ever participated in offline pediatric research, n (%)	326 (7.7)	72 (6.8)	254 (8)	.19
Ever participated in online pediatric research, n (%)	230 (5.4)	41 (3.9)	189 (6)	.008
Age of the caregivers (years), median (IQR)	33 (29-36)	33 (30-36)	32 (29-36)	.81
Role of the caregivers, n (%)				.005
Mother	3547 (83.8)	921 (86.6)	2626 (82.9)	
Father or others	684 (16.2)	143 (13.4)	541 (17.1)	
Education levels of the caregivers, n (%)				<.001
Senior high school or below	2334 (55.2)	455 (42.8)	1879 (59.3)	
College or above	1897 (44.8)	609 (57.2)	1288 (40.7)	
Number of people in the family, n (%)				<.001
2-4 people	2297 (54.3)	707 (66.4)	1590 (50.2)	
≥5 people	1934 (45.7)	357 (33.6)	1577 (49.8)	

Facilitators and Barriers to Participation

When questioned about their perspectives on research participation, 2531 (59.8%) caregivers expressed their first impressions, with 1561 (36.9%) caregivers inclined to participate and 970 (22.9%) caregivers inclined to decline. Suburban caregivers had more positive impressions than urban caregivers (38.6% vs 32%; *P*<.001; aOR 1.27, 95% CI 1.09-1.47). Additionally, caregivers with positive impressions reported higher rates of both offline (14.2% vs 2.9%; *P*<.001) and online (9.8% vs 2.7%; *P*<.001) research participation compared to those with negative impressions.

Table 2 presents the facilitators and barriers to participation, with most facilitators reporting rates lower than the barriers. The most significant facilitator was "be beneficial for children's health" (n=3579, 84.6%), which was more common among urban caregivers (87.5% vs 83.6%; *P*<.001; aOR 1.45, 95% CI 1.17-1.79). Interestingly, although "be suggested by close people" was deemed less important (n=635, 15%), it had a higher rate among suburban caregivers (16.2% vs 11.5%; *P*<.001; aOR 1.47, 95% CI 1.18-1.79). On the other hand, the most significant barriers were "worry about being an experimental subject" (Guinea pig concerns: n=3060, 72.3%) and "pose a risk to children's health" (n=2577, 60.9%). These barriers reflected a general apprehension about potential health risks. Furthermore, suburban caregivers had lower reporting rates for most barriers than urban caregivers, including "worry about being an experimental subject" (70.9% vs 76.6%; *P*<.001; aOR 0.79, 95% CI 0.67-0.93), "pose a risk to children's health" (58.6% vs 67.8%; *P*<.001; aOR 0.71, 95% CI 0.61-0.83), "do not have enough background information" (55.2% vs 61.6%; *P*<.001; aOR 0.78, 95% CI 0.67-0.89), and "worry about recommending other products" (48.2% vs 55%; *P*<.001; aOR 0.78, 95% CI 0.67-0.89).

Table 2. Facilitators and barriers to pediatric research participation.

Questions about facilitators and barriers	Overall, n (%)	Urban, n (%)	Suburban, n (%)	*P* value
Facilitators				
Be beneficial for children’s health	3579 (84.6)	931 (87.5)	2648 (83.6)	<.001[a]
Trust the hospital and researchers	1454 (34.4)	371 (34.9)	1083 (34.2)	.69
Acknowledge the importance of research	1409 (33.3)	358 (33.6)	1051 (33.2)	.78
Increase contact with doctors	1189 (28.1)	295 (27.7)	894 (28.1)	.75
Have small burden on the participants	1139 (26.9)	323 (30.4)	816 (25.8)	.003[a]
Be driven by altruism	1044 (24.7)	261 (24.5)	783 (24.7)	.90
Be suggested by close people	635 (15)	122 (11.5)	513 (16.2)	<.001[a]
Have economic subsidies	390 (9.2)	100 (9.4)	290 (9.2)	.81
Barriers				
Worry about being an experimental subject	3060 (72.3)	815 (76.6)	2245 (70.9)	<.001[a]
Pose a risk to children’s health	2577 (60.9)	721 (67.8)	1856 (58.6)	<.001[a]
Do not have enough background information	2404 (56.8)	655 (61.6)	1749 (55.2)	<.001[a]
Worry about personal privacy	2153 (50.9)	574 (53.9)	1579 (49.9)	.02
Worry about recommending other products	2112 (49.9)	585 (55)	1527 (48.2)	<.001[a]
Refuse procedures like venipuncture	1721 (40.7)	468 (44)	1253 (39.6)	.01
Cause disagreements among family members	1106 (26.1)	294 (27.6)	812 (25.6)	.20
Affect the purchase of medical insurance	803 (19)	201 (18.9)	602 (19)	.93

[a] Confirmed by multivariate logistic regression that adjusted general characteristics (Multimedia Appendix 2).

Perspectives on Digital Medicine

When questioned about digital medicine (Table 3), 36.4% (n=1542) of caregivers reported following online health-related accounts, while 38.4% (n=1626) had used online information for health-related decisions. Additionally, 38.3% (n=1621) and 36.7% (n=1552) of caregivers recognized the increasing popularity of digital tools in future health care practices and supported their integration into pediatric research. Nearly half of caregivers (n=2073, 49%) were willing to collect data daily using digital tools, and 61.7% (n=2610) supported the use of gamification in pediatric research. However, only 25.2% (n=1068) of caregivers supported online-only research methods, with slightly more support from suburban caregivers (26% vs 22.9%; *P*=.045; aOR 1.19, 95% CI 1.01-1.41). When asked about concerns regarding smartphone data collection software, privacy was the main issue (n=3273, 77.4%), followed by concerns about potential software addiction (n=2457, 58.1%). Multivariate logistic regression revealed privacy concerns as the barrier to supporting digital tools in pediatric research (aOR 0.61, 95% CI 0.52-0.71).

Table 3. Perspectives on digital medicine.

Questions about digital medicine	Overall, n (%)	Urban, n (%)	Suburban, n (%)	*P* value
Have followed online health-related accounts	1542 (36.4)	410 (38.5)	1132 (35.7)	.10
Can use online information to make health-related decisions	1626 (38.4)	327 (30.7)	1299 (41)	<.001[a]
Knowing that digital tools will become popular in future health care practices	1621 (38.3)	411 (38.6)	1210 (38.2)	.81
Support digital tools in pediatric research	1552 (36.7)	372 (35)	1180 (37.3)	.18
Support gamification of pediatric research	2610 (61.7)	680 (63.9)	1930 (60.9)	.09
Support online-only research methods	1068 (25.2)	244 (22.9)	824 (26)	.045[a]
Be willing to use digital tools daily to collect data	2073 (49)	484 (45.5)	1589 (50.2)	.008
Concerns about using smartphone software in research				
Personal privacy issues	3273 (77.4)	855 (80.4)	2418 (76.3)	.007[a]
Possible addiction of the tools	2457 (58.1)	606 (57)	1851 (58.4)	.39
Transparency of data	2159 (51)	564 (53)	1595 (50.4)	.14
Health impacts of the tools (such as radiation)	1830 (43.3)	439 (41.3)	1391 (43.9)	.13
Possible expenses	1465 (34.6)	367 (34.5)	1098 (34.7)	.92
Ease of use	1279 (30.1)	320 (30.1)	959 (30.3)	.90

[a] Confirmed by multivariate logistic regression that adjusted general characteristics (Multimedia Appendix 2).

Perspectives on Recruitment

Regarding recruitment (Table 4), 39.4% (n=1666) and 39.5% (n=1670) of caregivers expressed interest in reviewing recruitment information presented offline and online, respectively. Only 930 (22%) caregivers reported they would check both online and offline recruitment information. Notably, a greater proportion of suburban caregivers preferred checking offline recruitment information (41.2% vs 34%; *P*<.001; aOR 1.28, 95% CI 1.11-1.49). Additionally, 42.9% (n=1813) of caregivers expressed doubts about the reliability of online recruitment, while 58.1% (n=2457) doubted that research participation through online recruitment would lead to different treatments compared to offline methods.

When questioned about preferences, caregivers indicated that doctors with senior titles (n=2157, 51%) and specific researchers (n=2053, 48.5%) were the preferred recruiters, together accounting for 70.8% (n=2996) of caregivers. Research institution public accounts (n=3400, 80.4%) were the most favored source of online recruitment, with higher preferences among urban caregivers (85.9% vs 78.5%; *P*<.001; aOR 1.62, 95% CI 1.33-1.97). Although social media friends (n=984, 23.3%) and moments (n=608, 14.4%) were less popular sources for online recruitment, they were more frequently reported by suburban caregivers (friends: 24.7% vs 18.9%; *P*<.001; aOR 1.33, 95% CI 1.11-1.59; moments: 15.5% vs 11.1%; *P*<.001; aOR 1.35, 95% CI 1.09-1.67). Finally, the majority of caregivers (n=2963, 70%) preferred cartoon-style recruitment advertisements.

Table 4. Perspectives on recruitment.

Questions about recruitment	Overall, n (%)	Urban, n (%)	Suburban, n (%)	*P* value
Know that informed consent is required before beginning	3549 (83.9)	900 (84.6)	2649 (83.6)	.47
Be willing to review offline recruitment information	1666 (39.4)	362 (34)	1304 (41.2)	<.001[a]
Be willing to review online recruitment information	1670 (39.5)	408 (38.3)	1262 (39.8)	.39
Doubt the reliability of online recruitment information	1813 (42.9)	474 (44.5)	1339 (42.3)	.20
Believe that the treatment between online and offline recruitment is different	2457 (58.1)	603 (56.7)	1854 (58.5)	.29
Preferred role of recruiter				
Doctors (senior title)	2157 (51)	549 (51.6)	1608 (50.8)	.64
Specific researchers	2053 (48.5)	552 (51.9)	1501 (47.4)	.01
Medical students	786 (18.6)	188 (17.7)	598 (18.9)	.38
Doctors (ordinary title)	459 (10.8)	130 (12.2)	329 (10.4)	.10
Nurses	294 (6.9)	75 (7)	219 (6.9)	.88
Preferred source of online recruitment information				
Public accounts of research institutions	3400 (80.4)	914 (85.9)	2486 (78.5)	<.001[a]
Friends on social media	984 (23.3)	201 (18.9)	783 (24.7)	<.001[a]
Moments on social media	608 (14.4)	118 (11.1)	490 (15.5)	<.001[a]
Prefer cartoon version of the recruitment information	2963 (70)	752 (70.7)	2211 (69.8)	.60

[a] Confirmed by multivariate logistic regression that adjusted general characteristics (Multimedia Appendix 2).

Perspectives on the Research Process

When questioned about the research process (Table 5), 1445 (34.2%) caregivers reported that they knew they could withdraw from the study, and 79.8% (n=1153) of them would consult with the research team before doing so. Regarding follow-up, 3076 (72.7%) caregivers reported that regular contact was necessary. Telephones (n=1916, 62.3%) were the preferred method, followed by social media apps (n=1801, 58.6%). Social media apps were more common among urban participants (64.7% vs 56.4%; *P*<.001; aOR 1.39, 95% CI 1.17-1.65). For research feedback, 4017 (94.9%) caregivers expressed willingness to receive it, with social media apps (n=2624, 65.3%) being the preferred method, followed by email (n=2535, 63.1%). Additionally, we questioned at the beginning and end of the questionnaire whether caregivers were willing to provide phone numbers in anonymous surveys; a total of 3065 (72.4%) caregivers expressed willingness in offline surveys, a rate significantly higher than that in online surveys (72.4% vs 43.4%; *P*<.001).

Table 5. Perspectives on the research process.

Questions about research process	Overall, n (%)	Urban, n (%)	Suburban, n (%)	*P* value
Know that withdrawing midway is unconditionally allowed	1445 (34.2)	366 (34.4)	1079 (34.1)	.85
Would discuss with the research team before withdrawing	1153 (79.8)	300 (82.0)	853 (79.1)	.23
Believe that regular contact is necessary	3076 (72.7)	785 (73.8)	2291 (72.3)	.36
Preferred method (telephone)	1916 (62.3)	514 (65.5)	1402 (61.2)	.03
Preferred method (social media apps)	1801 (58.6)	508 (64.7)	1293 (56.4)	<.001[a]
Preferred method (video)	715 (23.2)	199 (25.4)	516 (22.5)	.11
Be willing to receive research feedback	4017 (94.9)	1024 (96.2)	2993 (94.5)	.03[a]
Preferred route (social media apps)	2624 (65.3)	687 (67.1)	1937 (64.7)	.17
Preferred route (mail)	2535 (63.1)	694 (67.8)	1841 (61.5)	<.001[a]
Preferred route (telephone)	2071 (51.6)	457 (44.6)	1614 (53.9)	<.001[a]
Be willing to provide phone number in anonymous surveys				
Offline surveys	3065 (72.4)	757 (71.1)	2308 (72.9)	.27
Online surveys	1835 (43.4)	452 (42.5)	1383 (43.7)	.50

[a] Confirmed by multivariate logistic regression that adjusted general characteristics (Multimedia Appendix 2).

Discussion

Principal Findings

This study investigated public perspectives on the digitalization of pediatric research, comparing views between urban and suburban areas. The findings indicate that the current support for digital approaches in pediatric research is suboptimal but still offers valuable strategies for using digital methods to reduce urban-suburban health disparities in children. This study is relatively rare in the literature and explores innovative ways to enhance participation and equity through digital approaches. Previous studies have emphasized the importance of understanding the barriers and facilitators that parents face when considering their children's participation in research [20-27]. As much of the research has already begun digitalization [28-32], it is beneficial to periodically pause and gather insights from participants by considering their perspectives.

Perspectives on Pediatric Research

A previous study on neonatal clinical trials revealed that one-quarter of parents made immediate enrollment decisions [27]. Our research found that 59.8% (n=2531) of caregivers formed initial impressions of pediatric research, with 36.9% (n=1561) expressing positive inclinations and 22.9% (n=970) expressing negative inclinations. These findings emphasize the importance of recruitment strategies during initial contact. Further analysis identified health benefits as the main facilitators and risks as the primary barriers, which aligns with previous studies from other regions [20]. This suggests the results could be applicable beyond Chongqing. The lower reporting rates of facilitators than barriers may explain challenges in pediatric research. Preferences in recruitment leaned toward doctors with senior titles and specific researchers as recruiters, emphasizing the influence of authority on caregivers' decision-making. Additionally, previous research has shown that cartoon images influence children's decision-making [44]. This study found that 70% (n=2963) of caregivers also preferred cartoon-style recruitment materials, possibly because they convey a sense of care for children, which resonates well with caregivers [45].

Perspectives on Digital Medicine

Public perspectives on digital medicine revealed suboptimal acceptance among caregivers. Approximately one-third of caregivers acknowledged the increasing popularity of digital tools and supported their application in pediatric research. This was reflected in a lower willingness to provide phone numbers in online surveys than in offline surveys. The transition from traditional to remote trials, particularly during the COVID-19 pandemic, has been successful [30,46], though some unsatisfactory outcomes have been reported [47,48]. Our results showed that public accounts from research institutions were the most preferred source of online recruitment, aligning with offline sources that emphasize authority. Given that 42.9% (n=1813) and 58.1% (n=2457) of caregivers doubted the reliability and consistency of online recruitment, establishing a credible public account by research institutions could enhance recruitment success on social media, a novel strategy compared to previous studies [49]. On the other hand, the most commonly mentioned methods for regular contact were telephone and social media apps. This may explain the effectiveness of the hybrid follow-up model, which encourages using both phone calls and social media [50]. However, only 25.2% (n=1068) of caregivers supported online-only research methods, highlighting the importance of face-to-face interactions. This raises important questions about the effectiveness of a fully remote model in postpandemic pediatric research. When questioned about gamification in research, 61.7% (n=2610) of caregivers were in favor. Studies on gamified medicine have shown positive outcomes [51,52], but concerns about potential risks remain [53,54]. In this study, 58.1% (n=2457) of caregivers expressed concerns about addiction. Additionally, 77.4% (n=3273) of caregivers expressed concerns about personal privacy when using smartphone software for research, which served as a barrier to supporting digital tools in pediatric research. Another advantage of digitalizing pediatric research is the availability of personalized real-time feedback, with 94.9% (n=4017) of participants willing to receive it. This meets caregivers' needs and serves as a strategy to increase motivation [55].

Differences Between Urban and Suburban Areas

When comparing urban and suburban areas, demographic differences were not adjusted using methods like propensity score matching, as these differences were inherent to the population. Instead, multivariate logistic regression was used to validate the results. Interestingly, suburban caregivers expressed greater interest in pediatric research and fewer barriers and concerns. These findings contradict common knowledge [15,16,56] but align with a previous study on congenital heart disease [57]. Another recent study also showed that rural caregivers had lower concerns about data privacy and security [58]. A possible explanation for suburban caregivers' higher enthusiasm is that the suburban areas in this study may not accurately represent resource-poor regions. These caregivers demonstrate sufficient health awareness and capabilities, making them value participation in pediatric research. This is evident in their willingness to access offline recruitment information, use digital tools for data collection, and support online-only research methods. Since suburban areas usually have more kindergarten children than urban areas (eg, Chongqing [33]), they represent a valuable potential pool of potential research participants. Further analysis revealed that suburban caregivers more frequently mentioned suggestions from close people. Similarly, suburban caregivers more often cited social media friends and moments as sources of online recruitment information. While these results raise concerns about caregivers' ability to assess the quality of online health information [59], they also highlight the potential for incorporating informal channels in recruitment strategies for suburban participants [60].

Strengths and Limitations

This study has several strengths. First, this study examined perspectives on digital medicine, which is distinct from conventional pediatric research topics [16,20-27] and provides essential data for the digital age. Second, this study included a more diverse sample from the general public, offering a broader representation compared to studies focused on specific illnesses. Third, the sample size was larger than that in previous studies, likely providing a more population-based overview. Finally, we performed a comparative analysis between urban and suburban areas, contributing to the understanding and mitigation of health inequalities among children in different geographical locations.

Despite these strengths, the limitations require careful consideration. First, this study was conducted in Chongqing, China. It is important to consider this when generalizing the findings to other regions. Therefore, further studies across different regions or countries are recommended to expand these findings. Second, although comparisons were made between urban and suburban areas, conditions in rural areas remain unknown and require further investigation. Third, perspectives were assessed using a structured questionnaire but lacked qualitative interview data. Additionally, despite cultural adaptations and a preliminary survey, individuals may still interpret certain questions differently based on their personal experiences. Finally, although smartphones are a prominent tool in digital health, the study's focus on them may introduce bias. However, less common digital tools may not be widely known to the general public, making it difficult to survey participants effectively about unfamiliar technologies.

Conclusions

In the digital age, approximately one-third of caregivers support pediatric research and endorse the use of digital tools in such studies. Caregivers in suburban areas show greater enthusiasm for research and digital medicine. Integrating official and unofficial social media recruitment strategies, implementing a hybrid online-offline follow-up approach, and addressing privacy concerns could help researchers increase participant recruitment and retention rates.

Acknowledgments

YH was affiliated with the Children's Hospital of Chongqing Medical University at the time of the study and is currently affiliated with Beijing Children's Hospital. The authors would like to thank Drs Shuyao Zhu, Keran Ling, Lin Chen, Zehao Li, Wei Zhang, Liyu Chen, and Yuxin Lv for their contributions to revising and culturally adapting the questionnaire. Special thanks are also extended to the kindergartens, teachers, and caregivers for their cooperation in distributing and responding to the questionnaire. This study was supported by the Major Program of the National Clinical Research Center for Child Health and Disorders in Children's Hospital of Chongqing Medical University (NCRCCHD-2023-MP-01) and the Program for Youth Innovation in Future Medicine, Chongqing Medical University, "Clinical Innovation Team for Standardized Management of Food Allergy in Children" (2022-26-W0120). The funder had no role in the preparation, review, or approval of the manuscript or the decision to submit it for publication. No generative artificial intelligence was used to generate the manuscript content.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

HF and RX contributed equally to this work and should be considered joint first authors. YZ and YH contributed equally to this work and should be considered joint corresponding authors. HF and RX conceptualized and designed the study, designed the data collection instruments, carried out the initial analyses, and drafted the initial manuscript. JL designed the data collection instruments, collected the data, carried out the initial analyses, and reviewed and revised the manuscript. YL and EL designed the data collection instruments, obtained funding, and reviewed and revised the manuscript. YZ and YH designed the data collection instruments, coordinated and supervised the data collection, obtained funding, and reviewed and revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire (in Chinese).
[PDF File (Adobe PDF File), 248 KB - jmir_v27i1e60324_app1.pdf]

Multimedia Appendix 2
Multivariate logistic regression.
[PDF File (Adobe PDF File), 172 KB - jmir_v27i1e60324_app2.pdf]

References

1. Wyatt TH, Li X, Fancher S, Mitoubsi AS, Pardue J. Recruitment barriers of an mHealth pediatric asthma pilot study. West J Nurs Res 2024;46(2):152-159. [doi: 10.1177/01939459231217919] [Medline: 38112079]
2. Niemeyer L, Mechler K, Buitelaar J, Durston S, Gooskens B, Oranje B, et al. 'Include me if you can'—reasons for low enrollment of pediatric patients in a psychopharmacological trial. Trials 2021;22(1):178 [FREE Full text] [doi: 10.1186/s13063-021-05119-6] [Medline: 33648579]
3. Fischer HS, Bührer C, Czernik C. Hazards to avoid in future neonatal studies of nasal high-frequency oscillatory ventilation: lessons from an early terminated trial. BMC Res Notes 2019;12(1):237 [FREE Full text] [doi: 10.1186/s13104-019-4268-2] [Medline: 31023363]
4. Pica N, Bourgeois F. Discontinuation and nonpublication of randomized clinical trials conducted in children. Pediatrics 2016;138(3):e20160223 [FREE Full text] [doi: 10.1542/peds.2016-0223] [Medline: 27492817]
5. Fang H, Chen L, Li J, Ren L, Yin Y, Chen D, et al. A web-based instrument for infantile atopic dermatitis identification (electronic version of the Modified Child Eczema Questionnaire): development and implementation. J Med Internet Res 2023;25:e44614 [FREE Full text] [doi: 10.2196/44614] [Medline: 37467020]
6. Fang H, Li Z, Xian R, Yin Y, Wang J, Guo H, et al. Early life growth and developmental trajectory in children with biliary atresia undergoing primary liver transplantation. Front Pediatr 2023;11:1198360 [FREE Full text] [doi: 10.3389/fped.2023.1198360] [Medline: 37377760]
7. Di Cesare M, Sorić M, Bovet P, Miranda JJ, Bhutta Z, Stevens GA, et al. The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. BMC Med 2019;17(1):212 [FREE Full text] [doi: 10.1186/s12916-019-1449-8] [Medline: 31760948]
8. Racine N, McArthur BA, Cooke JE, Eirich R, Zhu J, Madigan S. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. JAMA Pediatr 2021;175(11):1142-1150 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.2482] [Medline: 34369987]
9. Christakis DA. Internet addiction: a 21st century epidemic? BMC Med 2010;8(1):61 [FREE Full text] [doi: 10.1186/1741-7015-8-61] [Medline: 20955578]
10. Sayal K, Prasad V, Daley D, Ford T, Coghill D. ADHD in children and young people: prevalence, care pathways, and service provision. Lancet Psychiatry 2018;5(2):175-186. [doi: 10.1016/S2215-0366(17)30167-0] [Medline: 29033005]
11. Koren G. Healthy children as subjects in pharmaceutical research. Theor Med Bioeth 2003;24(2):149-159. [doi: 10.1023/a:1024694828858] [Medline: 12943269]
12. Postma L, Luchtenberg ML, Verhagen AAE, Maeckelberghe ELM. The attitudes of healthy children and researchers towards the challenges of involving children in research: an exploratory study. Res Involve Engage 2021;7(1):30. [doi: 10.1186/s40900-021-00263-4] [Medline: 33990230]
13. Di Carlo C, Mighton C, Clausen M, Joshi E, Casalino S, Kim THM, et al. Parents' attitudes towards research involving genome sequencing of their healthy children: a qualitative study. Eur J Hum Genet 2024;32(2):171-175. [doi: 10.1038/s41431-023-01476-9] [Medline: 37864046]
14. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian laboratory initiative on pediatric reference intervals: a CALIPER white paper. Crit Rev Clin Lab Sci 2017;54(6):358-413. [doi: 10.1080/10408363.2017.1379945] [Medline: 29017389]

15. Tanner A, Kim SH, Friedman DB, Foster C, Bergeron CD. Barriers to medical research participation as perceived by clinical trial investigators: communicating with rural and African American communities. J Health Commun 2015;20(1):88-96. [doi: 10.1080/10810730.2014.908985] [Medline: 25204763]
16. Watson SE, Smith P, Snowden J, Vaughn V, Cottrell L, Madden CA, et al. Facilitators and barriers to pediatric clinical trial recruitment and retention in rural and community settings: a scoping review of the literature. Clin Transl Sci 2022;15(4):838-853 [FREE Full text] [doi: 10.1111/cts.13220] [Medline: 35037409]
17. Chen Y, Sylvia S, Dill SE, Rozelle S. Structural determinants of child health in rural China: the challenge of creating health equity. Int J Environ Res Public Health 2022;19(21):13845 [FREE Full text] [doi: 10.3390/ijerph192113845] [Medline: 36360724]
18. Wu Q, Pan J. The trends of health equity in children aged 5-12 years and its determinants in China from 1991 to 2015 [in Chinese]. J Preventive Med Inf 2020;36(5):633-640.
19. COMMITTEE ON BIOETHICS. Informed consent in decision-making in pediatric practice. Pediatrics 2016;138(2):e20161484. [doi: 10.1542/peds.2016-1484] [Medline: 27456514]
20. Nathe JM, Oskoui TT, Weiss EM. Parental views of facilitators and barriers to research participation: systematic review. Pediatrics 2023;151(1):e2022058067 [FREE Full text] [doi: 10.1542/peds.2022-058067] [Medline: 36477217]
21. Engster SA, Fascetti C, Daw K, Reis EC. Parent perceptions of and preferences for participation in child health research: results from a pediatric practice-based research network. J Am Board Fam Med 2019;32(5):685-694 [FREE Full text] [doi: 10.3122/jabfm.2019.05.190030] [Medline: 31506364]
22. Tromp K, Zwaan CM, van de Vathorst S. Motivations of children and their parents to participate in drug research: a systematic review. Eur J Pediatr 2016;175(5):599-612 [FREE Full text] [doi: 10.1007/s00431-016-2715-9] [Medline: 27041121]
23. Sollo N, Ahlers-Schmidt CR, Davis AM, Smith TR, Dedeaux JA, McCulloh RJ. Perceived barriers to clinical trials participation: a survey of pediatric caregivers. Kans J Med 2022;15:139-143 [FREE Full text] [doi: 10.17161/kjm.vol15.16220] [Medline: 35646248]
24. Tu WJ, Li Y, Shi XD. The views of Chinese parents on research participation and informed consent. J Paediatr Child Health 2012;48(2):186-187. [doi: 10.1111/j.1440-1754.2011.02404.x] [Medline: 22320277]
25. McKenzie PL, Siegel DH, Cullen D, Castelo-Soccio L. Strategies to enhance pediatric clinical trial participation: a systematic review with narrative synthesis. Pediatr Dermatol 2021;38(6):1515-1522. [doi: 10.1111/pde.14817] [Medline: 34647357]
26. Weiss EM, Olszewski AE, Guttmann KF, Magnus BE, Li S, Shah AR, et al. Parental factors associated with the decision to participate in a neonatal clinical trial. JAMA Netw Open 2021;4(1):e2032106 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.32106] [Medline: 33433595]
27. Weiss EM, Guttmann KF, Olszewski AE, Magnus BE, Li S, Kim SY, et al. Parental enrollment decision-making for a neonatal clinical trial. J Pediatr 2021;239:143-149.e3 [FREE Full text] [doi: 10.1016/j.jpeds.2021.08.014] [Medline: 34400207]
28. Harmon DM, Noseworthy PA, Yao X. The digitization and decentralization of clinical trials. Mayo Clin Proc 2023;98(10):1568-1578. [doi: 10.1016/j.mayocp.2022.10.001] [Medline: 36669937]
29. Marra C, Chen JL, Coravos A, Stern AD. Quantifying the use of connected digital products in clinical research. NPJ Digital Med 2020;3:50 [FREE Full text] [doi: 10.1038/s41746-020-0259-x] [Medline: 32285011]
30. 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;117(3):241-245. [doi: 10.1016/j.anai.2016.07.012] [Medline: 27613456]
31. Alnasser Y, Proaño A, Loock C, Chuo J, Gilman RH. Telemedicine and pediatric care in rural and remote areas of middle-and-low-income countries: narrative review. J Epidemiol Glob Health 2024;14(3):779-786 [FREE Full text] [doi: 10.1007/s44197-024-00214-8] [Medline: 38478166]
32. Tan RKJ, Wu D, Day S, Zhao Y, Larson HJ, Sylvia S, et al. Digital approaches to enhancing community engagement in clinical trials. NPJ Digital Med 2022;5(1):37 [FREE Full text] [doi: 10.1038/s41746-022-00581-1] [Medline: 35338241]
33. Chongqing statistical yearbook 2022. Chongqing Municipal Statistics Bureau. URL: https://tjj.cq.gov.cn/zwgk_233/tjnj [accessed 2022-12-26]
34. Kindergarten work regulations. Ministry of Education of the People's Republic of China. URL: http://www.moe.gov.cn/jyb_xxgk/xxgk/zhengce/guizhang/202112/t20211206_585104.html [accessed 2016-01-05]
35. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
36. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
37. Jin H, Cui M, Liu J. Factors affecting people's attitude toward participation in medical research: a systematic review. Curr Med Res Opin 2020;36(7):1137-1143. [doi: 10.1080/03007995.2020.1760807] [Medline: 32329364]
38. El Achi D, Al Hakim L, Makki M, Mokaddem M, Khalil PA, Kaafarani BR, et al. Perception, attitude, practice and barriers towards medical research among undergraduate students. BMC Med Educ 2020;20(1):195 [FREE Full text] [doi: 10.1186/s12909-020-02104-6] [Medline: 32552801]

39. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Making 2016;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
40. Fang H, Lv Y, Chen L, Zhang X, Hu Y. The current knowledge, attitudes, and practices of the neglected methodology of web-based questionnaires among Chinese health workers: web-based questionnaire study. J Med Internet Res 2023;25:e41591 [FREE Full text] [doi: 10.2196/41591] [Medline: 36533302]
41. Lu Y, Fang J, Tian L, Jin H. Advanced Medical Statistics. Singapore: World Scientific Publishing CO Pte Ltd; 2015:1472.
42. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb) 2021;31(1):010502 [FREE Full text] [doi: 10.11613/BM.2021.010502] [Medline: 33380887]
43. Blumenberg C, Barros AJD. Response rate differences between web and alternative data collection methods for public health research: a systematic review of the literature. Int J Public Health 2018;63(6):765-773. [doi: 10.1007/s00038-018-1108-4] [Medline: 29691594]
44. Tzeng YF, Gau BS. Suitability of asthma education materials for school-age children: Implications for health literacy. J Clin Nurs 2018;27(5-6):e921-e930. [doi: 10.1111/jocn.13993] [Medline: 28793368]
45. Levetown M, American Academy of Pediatrics Committee on Bioethics. Communicating with children and families: from everyday interactions to skill in conveying distressing information. Pediatrics 2008;121(5):e1441-e1460. [doi: 10.1542/peds.2008-0565] [Medline: 18450887]
46. Simmons LA, Phipps JE, Whipps M, Smith P, Carbajal KA, Overstreet C, et al. From hybrid to fully remote clinical trial amidst the COVID-19 pandemic: strategies to promote recruitment, retention, and engagement in a randomized mHealth trial. Digit Health 2022;8:20552076221129065 [FREE Full text] [doi: 10.1177/20552076221129065] [Medline: 36185388]
47. Zapata KA, Wang-Price SS, Fletcher TS, Johnston CE. Factors influencing adherence to an app-based exercise program in adolescents with painful hyperkyphosis. Scoliosis Spinal Disord 2018;13:11 [FREE Full text] [doi: 10.1186/s13013-018-0159-x] [Medline: 30027121]
48. Haynes SC, Kamerman-Kretzmer R, Khan SS, Crossen S, Lieng MK, Marcin JP, et al. Telemedicine use for pediatric asthma care: a mixed methods study. J Asthma 2022;59(12):2431-2440. [doi: 10.1080/02770903.2021.2019265] [Medline: 34913803]
49. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022;24(8):e38015 [FREE Full text] [doi: 10.2196/38015] [Medline: 35925655]
50. Qiao S, Tang L, Zhang W, Tian S, Liu M, Yang L, et al. Nurse-led follow-up to outpatients with cancer pain treated with opioids at home-telephone calls plus WeChat versus telephone calls only: a quasi-experimental study. Patient Prefer Adherence 2019;13:923-931 [FREE Full text] [doi: 10.2147/PPA.S203900] [Medline: 31239650]
51. Pimentel-Ponce M, Romero-Galisteo RP, Palomo-Carrión R, Pinero-Pinto E, Merchán-Baeza JA, Ruiz-Muñoz M, et al. Gamification and neurological motor rehabilitation in children and adolescents: a systematic review. Neurologia (Engl Ed) 2024;39(1):63-83 [FREE Full text] [doi: 10.1016/j.nrleng.2023.12.006] [Medline: 38065433]
52. Suleiman-Martos N, García-Lara RA, Membrive-Jiménez MJ, Pradas-Hernández L, Romero-Béjar JL, Dominguez-Vías G, et al. Effect of a game-based intervention on preoperative pain and anxiety in children: a systematic review and meta-analysis. J Clin Nurs 2022;31(23-24):3350-3367 [FREE Full text] [doi: 10.1111/jocn.16227] [Medline: 35075716]
53. Boendermaker WJ, Maceiras SS, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]
54. Holtz BE, Murray K, Park T. Serious games for children with chronic diseases: a systematic review. Games Health J 2018;7(5):291-301. [doi: 10.1089/g4h.2018.0024] [Medline: 29957082]
55. Wu Q, Wang X, Zhang J, Zhang Y, van Velthoven MH. The effectiveness of a WeChat-based self-assessment with a tailored feedback report on improving complementary feeding and movement behaviour of children aged 6-20 months in rural China: a cluster randomized controlled trial. Lancet Reg Health West Pac 2023;37:100796 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100796] [Medline: 37273963]
56. Stewart E, Milton A, Yee HF, Song MJ, Roberts A, Davenport T, et al. eHealth tools that assess and track health and well-being in children and young people: systematic review. J Med Internet Res 2022;24(5):e26015 [FREE Full text] [doi: 10.2196/26015] [Medline: 35550285]
57. Lin WH, Chen YK, Lin SH, Cao H, Chen Q. Parents' understanding and attitudes toward the use of the WeChat platform for postoperative follow-up management of children with congenital heart disease. J Cardiothorac Surg 2023;18(1):66 [FREE Full text] [doi: 10.1186/s13019-023-02153-0] [Medline: 36755263]
58. Chow E, Virani A, Pinkney S, Abdulhussein FS, van Rooij T, Görges M, et al. Caregiver and youth characteristics that influence trust in digital health platforms in pediatric care: mixed methods study. J Med Internet Res 2024;26:e53657 [FREE Full text] [doi: 10.2196/53657] [Medline: 39467279]
59. Li P, Chen B, Deveaux G, Luo Y, Tao W, Li W, et al. Cross-verification of COVID-19 information obtained from unofficial social media accounts and associated changes in health behaviors: web-based questionnaire study among Chinese netizens. JMIR Public Health Surveill 2022;8(5):e33577 [FREE Full text] [doi: 10.2196/33577] [Medline: 35486529]

60. Vos SC, Adatorwovor R, Roberts MK, Sherman DL, Bonds D, Dunfee MN, et al. Community engagement through social media: a promising low-cost strategy for rural recruitment? J Rural Health 2024;40(3):467-475. [doi: 10.1111/jrh.12809] [Medline: 37985592]

Abbreviations

aOR: adjusted odds ratio
CHERRIES: Checklists for Reporting Results of Internet E-Surveys
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

---

Edited by T de Azevedo Cardoso; submitted 07.05.24; peer-reviewed by L Giacomelli, SL Coven; comments to author 09.09.24; revised version received 15.09.24; accepted 21.11.24; published 07.01.25.

Please cite as:
Fang H, Xian R, Li J, Li Y, Liu E, Zhao Y, Hu Y
Urban-Suburban Differences in Public Perspectives on Digitalizing Pediatric Research: Cross-Sectional Survey Study
J Med Internet Res 2025;27:e60324
URL: https://www.jmir.org/2025/1/e60324
doi:10.2196/60324
PMID:

Original Paper

Association Between Intergenerational Support, Technology Perception and Trust, and Intention to Seek Medical Care on the Internet Among Chinese Older Adults: Cross-Sectional Questionnaire Study

Hengjiang Jin[1], PhD; Ying Qu[1], MA

[1] School of Journalism, Chongqing University, Chongqing, China

Corresponding Author:
Hengjiang Jin, PhD
School of Journalism
Chongqing University
Number 174, Shazheng Street, Shapingba District
Chongqing
China
Phone: 86 15523596050
Email: jinhj@cqu.edu.cn

Abstract

Background: Avoiding technological innovation does not simplify life. In fact, using internet-based medical services can enhance the quality of life for older adults. In the context of an aging population and the growing integration of information technology, the demand for internet-based medical services among older adults is gaining increased attention. While scholars have highlighted the important role of intergenerational support in promoting digital inclusion for older adults, research on the relationship between intergenerational support and older adults' intentions to seek online care remains limited.

Objective: This study aims (1) to explore the association between intergenerational support, online medical information, and older adults' intention to seek medical care online, and (2) to examine the mediating role of technology perception and trust, as well as the moderating role of eHealth literacy.

Methods: A cross-sectional survey was conducted in China, collecting 958 valid responses from older adults aged 60 years and above. A vast majority of participants were between the ages of 60 and 75 years (771/958, 80.5%). Of the 958 participants, 559 (58.4%) resided in rural areas, while 399 (41.6%) lived in urban areas. The survey included questions on intergenerational support, perceived usefulness, perceived ease of use, trust, online medical information, eHealth literacy, and the intention to seek medical care online. Structural equation modeling and linear regression analysis were applied to explore the relationship between intergenerational support and the intention to seek medical care on the internet.

Results: Intergenerational support was positively associated with perceived ease of use (β=.292, *P*<.001), perceived usefulness (β=.437, *P*<.001), trust (β=.322, *P*<.001), and the intention to seek medical care online (β=.354, *P*<.001). Online medical information also positively affected the intention to seek medical care online among older adults (β=.109, *P*<.001). Perceived ease of use (β=.029, 95% CI 0.009-0.054), perceived usefulness (β=.089, 95% CI 0.050-0.130), and trust (β=.063, 95% CI 0.036-0.099) partially mediated the association between intergenerational support and the intention to seek medical care online. Further analysis found that perceived ease of use, perceived usefulness, and trust played a chain mediating role between intergenerational support and the intention to seek medical care online (β=.015, 95% CI 0.008-0.027; β=.022, 95% CI 0.012-0.036). Additionally, eHealth literacy played a moderating role in the relationship between intergenerational support and perceived ease of use (β=.177, *P*<.001), perceived usefulness (β=.073, *P*<.05), trust (β=.090, *P*<.01), and the intention to seek medical care online (β=.124, *P*<.001).

Conclusions: An integrated model of health communication effects was constructed and validated, providing empirical support for the intention to seek medical care online and for the impact of health communication. This model also helps promote the role of technology in empowering the lives of seniors.

(J Med Internet Res 2025;27:e65065) doi:10.2196/65065

KEYWORDS

intergenerational support; older adults; internet medical intentions; perceived technology; trust

Introduction

Background

The internet has become an important medium for providing various eHealth services and a source of medical information [1]. Smartphones bring the convenience of telemedicine to users' fingertips [2]. According to the 52nd China Internet Network Information Center report [3], China has 364 million internet medical care users, accounting for 33.8% of all internet users. The number of online hospitals continues to grow, and the development of large-scale online medical care platforms shows strong momentum.

The World Health Organization has encouraged member states to actively use digital technologies to improve medical, health, and social services, thereby promoting active aging [4]. Given that older adults are the primary users of health care services [5], there is significant potential for internet-based medical services to enhance geriatric health. Currently, China's population is aging rapidly [6]. As of the end of 2023, the population of older adults in China (aged ≥60 years) has reached 297 million, accounting for 21.1% of the total population. For older adults, proficient use of technology is key to accessing internet-based medical services [7]. However, digital disparities—such as difficulties in acquiring, operating, and using basic equipment, as well as insufficient knowledge—limit their ability to equally benefit from technological advancements. Given the trend of population aging, the challenge of successfully integrating older adults into the digital health ecosystem has become a pressing issue.

Some prior studies have explored this question, noting that older adults are particularly concerned about the quality and risks of internet-based medical services [8]. Additionally, older adults' habits play an important role in explaining their use of such services [9]. However, further discussion is needed to clarify the complexities and mechanisms that influence older adults' intention to seek medical care online.

Seeking Medical Care Online

Intention to adopt internet-based medical services refers to individuals' subjective beliefs and willingness to seek care through such services when facing health problems [10]. It reflects the importance individuals place on their health, as well as their trust in and reliance on internet-based medical services. This intention is influenced by multiple factors, including disease severity, individual perceptions and attitudes toward the illness, and the quality and availability of health care services.

Examinations of the intention to adopt internet-based medical services currently follow 3 approaches. The first approach is based on the Technology Acceptance Model or the Unified Theory of Acceptance and Use of Technology [11], which assesses user acceptance of online health platforms and the willingness of different population groups to use internet-based medical services [12]. The second approach constructs an analytical model centered on trust within the doctor-patient relationship, focusing on factors such as personal trust preferences and website trustworthiness [13]. The third approach is based on Social Exchange Theory, which examines the physicians' perspectives on the potential impact of interactive tools on their careers [14], as well as how regulatory systems, reputation systems, and communication exchanges shape their motivation to use these services [15].

Examining older adults' access to online health care in greater detail, Mansson et al [16] argued that the use of mobile health (mHealth) apps may reduce treatment costs. However, Askari et al [17] identified several factors influencing older adults' willingness to use mHealth apps, including perceived usefulness (PU), ease of use, social connections, social norms, and anxiety.

Intergenerational Support

Intergenerational Support Theory asserts that the flow of resources among family members is bidirectional [18], encompassing economic, technological, and emotional exchanges [19]. In the context of internet-based medical services, intergenerational support focuses on the technical assistance provided by the children to their parents in using these services.

Perceived ease of use (PEOU) measures older adults' ratings of technology's friendliness versus their beliefs about ease of mastery [20]. However, many older adults are unfamiliar with smartphone operation, are often excluded from various internet services, and are unable to benefit equally from digital technologies [21]. Additionally, they perceive apps as complex [22], resulting in low overall levels of PEOU. The lack of digital literacy and the challenges of aging are significant barriers to older adults' intention to use mHealth technology [23]. In the context of online-based medical services, children can assist older adults by guiding them through the process of using online platforms—such as downloading, registering, finding, and consulting internet-based medical services. They can also help resolve technical issues, thereby reducing frustration [24]. This support will make it easier for older adults to use these services, leading to an increase in their assessment of PEOU. Based on this, the following hypotheses are proposed.

- Hypothesis 1a: Intergenerational support is positively associated with perceived ease of use of internet-based medical services among older adults.

PU reflects older people's views that a particular technology can improve their health care experience. It encompasses their belief that internet-based medical services can meet their health management needs, such as online appointment booking, teleconsultation, and e-prescription flow. Studies have shown that older adults are more likely to adopt technology when they gain insight into its usefulness and potential benefits. Through direct operation or demonstration by children, older adults can independently or semi-independently use the internet-based medical service platform to experience its advantages in simplifying clinical services, saving time and energy for medical treatment, and enhancing their knowledge and acceptance of the platform. Once older adults experience the tangible benefits of internet-based medical services, it may stimulate positive feedback mechanisms that gradually ease the psychological barriers to adopting new technologies. This, in turn, can enhance psychological adaptation and cognitive health [25], thereby deepening the PU of the internet-based medical service platform.

- Hypothesis 1b: Intergenerational support is positively associated with the perceived usefulness of internet-based medical services for older adults.

Trust typically builds up gradually through continuous and intimate interactions, which can enhance patients' positive expectations [26]. However, many older adults have conservative and negative attitudes toward new technologies and are hesitant to use online services. As a result, it can be challenging for services such as internet-based medical services to gain the trust of older adults [27]. According to Ma et al [28], trust serves as a critical criterion for older adults in identifying individuals who can help them enhance their digital skills. Consequently, the younger generation plays a significant role in building trust and security for older adults. This support also involves a sense of being recognized and respected [29]. We hypothesize that greater intergenerational support is more likely to reduce older adults' unfamiliarity and thus increase their trust in these services.

- Hypothesis 1c: Intergenerational support is positively associated with older adults' trust in internet-based medical service.

According to Li and Kostka [30], social support plays a significant role in affecting older adults' digital engagement. Many older adults who maintain close contact with their families tend to receive more effective support for digital learning and engagement [30]. Additionally, given the importance of filial piety in Chinese culture, adult children are expected to assist disadvantaged family members [31]. This includes providing help and counseling regarding digital access, use, and literacy, as well as facilitating bottom-up technology transfer [32] to help their parents accomplish reverse socialization [33]. As a result, support and feedback can effectively enhance the digital competence of older adults [34] and promote the adoption of new technologies.

- Hypothesis 1d: Intergenerational support is positively associated with older adults' intention to seek medical care on the internet.

Perceived Usefulness and Perceived Ease of Use

As the key elements determining users' behavioral intentions, PU indicates the extent to which using a particular system or operation can improve task performance or achieve a specific goal. PEOU reflects the user's subjective evaluation of the convenience and ease provided by using something [35].

Some studies have shown that the PU of social media affects people's trust in the platform as well as the channel [36]. Acharya et al [37] further found that the PU of recommender systems can directly impact consumers' trust. It has also been confirmed that PEOU is positively correlated with consumer trust [38]. This study suggests, therefore, that older adults may trust internet-based medical services if they perceive them as useful or easy to use.

- Hypothesis 2a: Perceived ease of use is positively associated with older adults' trust in internet-based medical service.
- Hypothesis 2b: Perceived usefulness is positively associated with older adults' trust in internet-based medical service.

Currently, technology perception has been studied in various contexts, such as the adoption of smart home technologies [39] and the willingness to use internet applications [40]. Naidoo and Leonard [41] found that the continued willingness to use e-services is entirely dependent on high levels of PU. Further research indicates that the PU of the internet is an important predictor of an individual's use of eHealth solutions and plays a significant role in patients' sustained use of online health care communities [42]. For older adults, the functions of internet-based medical services are far more complex than simple services such as browsing short videos. Older patients tend to consider using these services only if the transition to new medical treatments is easy [43]. This study therefore argues that older adults are more likely to use internet-based medical services if they perceive them to be useful or easy to use.

- Hypothesis 3a: Perceived ease of use is positively associated with older adults' intention to seek medical care on the internet.
- Hypothesis 3b: Perceived usefulness is positively associated with older adults' intention to seek medical care on the internet.

Trust

Schoorman et al [44] suggested that trust in internet-based medical services is the tendency for people to believe that these services can fulfill their health needs and to be willing to use them, reflecting the level of acceptance for such services [45]. Trust is a crucial factor for assessing users' willingness to use internet-based medical services [46], and establishing trust improves the likelihood of older adults using these services [47]. This study therefore suggests that older adults who trust internet-based medical services are more likely to be willing to access such services. We therefore propose the following hypothesis:

- Hypothesis 4: Trust is positively associated with older adults' intention to seek medical care on the internet.

Online Medical Information

Online medical information—encompassing all types of medical-related content delivered and shared through internet platforms such as search engines, medical websites, and online forums—is highly sought after. The causes and treatments of diseases are among the most popular topics for medical information searches [48]. Notably, 60% of individuals view online medical information as equally good or better than information provided by physicians [49]. Online medical information enhances the efficiency of medical knowledge dissemination and is rapidly replacing traditional approaches for seeking advice and information [50]. It helps to improve individuals' knowledge, promote disease prevention, and facilitate access to appropriate medical services. To pursue more informed health decisions [51], older adults often seek health-related medical information online, which influences their health behaviors [52]. Consequently, older adults may be motivated to seek guidance from physicians practicing online [53]. The following research hypothesis is therefore proposed:

- Hypothesis 5: Online medical information is positively associated with older adults' intention to seek internet-based medical services.

eHealth Literacy

eHealth literacy refers to the ability needed to access, understand, and evaluate health information from digital resources and make informed health decisions [54]. It encompasses active information-seeking, 2-way interactive communication, and information utilization/sharing [55]. Although older adults are increasingly relying on the internet to access health-related services, they often struggle to meet the necessary eHealth literacy requirements [56]. Families play an important role in promoting eHealth literacy among older adults [57]. By strengthening intergenerational support, they can help older adults better learn and apply new technologies. When confronted with a complex internet health care delivery system, families can assist in quickly getting up to speed and becoming proficient in its operation. At this level, internet health care services are easy to use. At the same time, families can help older adults obtain rich health information from the platform [58] and manage their daily health needs, which can enhance their perception of the usefulness of these services. Therefore, we propose the following:

- Hypothesis 6a: eHealth literacy positively moderates the relationship between intergenerational support and perceived ease of use.
- Hypothesis 6b: eHealth literacy positively moderates the relationship between intergenerational support and perceived usefulness.

A study found that social interaction, access to technology, and digital literacy are positively correlated. With the assistance of their children, older adults can significantly improve their ability to access, understand, and apply digital health information [59]. High levels of eHealth literacy can increase older adults' confidence and ability to assess the reliability and validity of internet-based health care services, thereby enhancing their trust in such services.

- Hypothesis 6c: eHealth literacy positively moderates the relationship between intergenerational support and trust.

According to Health Empowerment Theory, eHealth literacy and social support can promote self-care behaviors among older adults [60]. The findings of Hsu et al [61] further indicate that individuals with higher levels of eHealth literacy have a greater potential to make informed health choices, improved health care competence, and ultimately higher quality of life. Therefore, we propose the following:

- Hypothesis 6d: eHealth literacy positively moderates the relationship between intergenerational support and intention to seek medical care on the internet.

Methods

Data Collection and Participants

In this study, a questionnaire (Multimedia Appendix 1) was used to validate the research model. To ensure reliability and validity, all items were borrowed from well-established domestic and international scales and adjusted according to the theme of internet medical care. The questionnaire comprises a total of 7 latent variables, involving 24 items, all measured on a 7-point Likert scale. Before the official survey, a presurvey was conducted with 50 older adults to ensure a high degree of internal consistency. Based on their feedback, questions with ambiguities or semantic problems were adjusted.

Between March 20, 2024, and March 31, 2024, we officially launched the distribution of the questionnaire using a combination of online and offline methods. This approach ensures that the data (Multimedia Appendix 2) collected are broad and representative, providing a more comprehensive understanding of the attitudes and needs of older adults regarding the use of the internet for health care services. The study was conducted among individuals aged 60 years and older who had autonomous behavioral skills and some experience with internet use, encompassing both rural and urban areas. Those who lacked basic knowledge of the internet or were unable to participate due to daily activities or cognitive limitations were excluded.

Online questionnaires offer the advantages of low cost, autonomy, and comprehensive documentation, as well as the ability to effectively expand the range of respondents. In this study, an online self-report questionnaire was designed using Questionnaire Star (Changsha Ranxing Information Technology Co., Ltd.). The questionnaire was distributed through a convenience sample with a snowball approach. An online survey link was shared across various social media platforms and WeChat (Tencent Holdings Limited) app groups, where participation was voluntary. This approach covered older adult groups from regions such as Chongqing, Hunan, Guizhou, and Zhejiang, among others. To address the unfamiliarity of some older adults with the operation of the Questionnaire Star and the limitation of their text reading comprehension ability, this study required them to complete the questionnaire in the presence and under the guidance of their middle-aged or young relatives or friends. This approach ensured the accuracy and reliability of the data by providing assistance with understanding and navigating the questions.

Although many older people already have smart devices, they may be wary of clicking on a link to participate in an online questionnaire. Therefore, face-to-face interviews are particularly necessary. In the offline questionnaire survey, we adopted a dual strategy: on the one hand, the research team personally visited communities and villages in Chongqing and Hunan, directly communicating with older adult groups through on-site visits to ask them in detail about their willingness to use internet health care services. On the other hand, with the support of governmental agencies where the researchers are domiciled, we commissioned local staff to assist in distributing some of the questionnaires. For the older adults interviewed who had the ability to write, we asked them to fill in the questionnaire themselves; for patients who were unable to fill in the questionnaire by themselves, their primary caregiver or interviewer filled in the questionnaire on their behalf. The questionnaires were collected on-site after completion.

During the data collection phase, a total of 1200 questionnaires were collected. After eliminating invalid questionnaires (including online questionnaires that took less <90 seconds to complete, failed the honesty test, had missing values, or provided the same answer to all questions, as well as offline questionnaires that were obviously invalid, such as choosing “strongly agree” for all of them), a total of 958 valid questionnaires were ultimately obtained for an effective response rate of 79.8%.

Measurements

Intergenerational Support

The measurement items for this dimension were derived from Lang and Schütze’s study [62] and consisted of 3 items. It was measured with good reliability (α=.887) [63]. This factor collectively explained 81.6% of the variance (Kaiser-Meier-Olkin [KMO]=0.748). Specific items included “My children encourage me to seek medical care on the internet, and I am willing to try it,” “My children guide I would be willing to try,” and “When I have problems with internet health care, my children help me solve the problem, so I would be willing to try it.”

Online Medical Information

Based on a study by de Boer et al [64], 4 items were used to measure the medical information that older adults sought online. The measurement items included “I would search for medical information on the internet” and “Online medical information has taught me something about health.” The results of factor analysis for the 4 items showed a KMO value of 0.843, with the factor explaining a total of 76.2% of the variance; these items also had good reliability (α=.895).

Trust

This dimension is based on the studies by Fogel and Nehmad [65] and Deng et al [66] and consists of 3 items. Specific items included “Most of the doctors on the online medical service platforms are health experts in their fields, and I have no doubt about their professionalism” and “Generally, I trust the health advice or tips from doctors on the internet.” The overall performance of these items indicated good reliability (α=.868); this factor collectively explained 79.2% of the variance (KMO=0.735).

Perceived Ease of Use

The 5 items were based on the study by Deng et al [66]. Typical items include “I don’t think it is difficult to use the internet for health counseling” or “In general, I think internet health care is easy to use.” The results of factor analysis for the 3 items showed a KMO value of 0.885, explaining a total of 72.7% of the variance. The 5 items showed good reliability (α=.906).

Perceived Usefulness

The items for this factor were based on those in the study by Deng et al [66]. The items included “I believe that using the internet for medical care can improve the efficiency of health care” and “Overall, I think online medical care is useful for health management.” The results of the analysis showed a KMO value of 0.746, explaining a total of 81.7% of the variance, and good reliability (α=.888).

Intention to Seek Medical Care on the Internet

These 3 items referred to the study by Deng et al [66]., including “When I face health problems, I think I will solve them through internet health care” and “I am willing to use internet health care services for health counseling, such as disease control.” These analyses showed a KMO value of 0.751, explaining a total of 82.8% of the variance; all 3 items also had good reliability (α=.895).

eHealth Literacy

This dimension combines health and media literacy and references the study by Norman and Skinner [67]. The 3 items included “I know how to find useful health resources and messages online” or “I know how to use the health care-type information I find on the internet to help myself,” with 82.6% variance explained, a KMO value of 0.751, and good reliability (α=.894).

All 24 items were rated on a 7-point Likert scale (1=strongly disagree, and 7=strongly agree). All measurement items can be found in Table S1 in Multimedia Appendix 3.

Data Analysis

Cronbach α was used to determine the reliability of the scales. Construct validity and reliability were examined using exploratory factor analysis with varimax rotation, Bartlett test, and KMO statistics, as well as confirmatory factor analysis (CFA).

For the analysis of structural equation modeling, Amos 26.0 (IBM Corp.) was used to analyze the data and the effect of the model was measured by the fit indices, which included chi-square/degree of freedom (χ^2/df), the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), the incremental fit index (IFI), the comparative fit index (CFI), the nonnormed fit index (NNFI), and root mean square error of approximation (RMSEA). Additionally, the bootstrap method was used to test the mediation model with 5000 iterations of repeated sampling. Results were also obtained using SPSS 26.0 (IBM Corp.) to test for interaction effects.

Ethical Considerations

Ethical approval for this study was obtained from the academic committee (acting as the ethics committee) of the School of Journalism and Communication at Chongqing University (approval number 20240320). Informed consent was obtained from all participants before their participation in the survey. Participants were provided with a detailed informed consent form outlining the purpose of the study, the duration of the investigation, the procedures, and the potential risks and benefits. They were informed that their participation was voluntary and that they had the right to withdraw at any time without facing any negative consequences. Participants were assured of the confidentiality and anonymity of their responses.

Results

Participant Characteristics

The sample demographics are as follows: of the 958 participants, 771 (80.5%) were between the ages of 60 and 75 years, 511 (53.3%) were females, and 447 (46.7%) were males. In terms of residential area, 559 (58.4%) participants resided in a rural area and 399 (41.6%) in an urban area. Most participants (812/958, 84.8%) had between 1 and 3 children, about one-quarter (214/958, 22.3%) of the participants reported an education level of high school and above, nearly half (440/958, 45.9%) rated their self-assessed health status as fair, and 186 (19.4%) rated it as poor or very poor. Detailed demographic information is presented in Table 1.

Table 1. Participant characteristics (N=958).

Characteristics	Values, n (%)
Age (years)	
60-65	444 (46.3)
66-70	183 (19.1)
71-75	144 (15.0)
76-80	123 (12.8)
>80	64 (6.7)
Gender	
Women	511 (53.3)
Male	447 (46.7)
Status of residence	
Residence with spouse	424 (44.3)
Large family living together	188 (19.6)
Living with children	184 (19.2)
Living alone	137 (14.3)
Others	25 (2.6)
Educational level	
Secondary schools	353 (36.8)
Junior high school	243 (25.4)
Not attending school	148 (15.4)
High school/secondary school	138 (14.4)
Undergraduate	53 (5.5)
Graduate students and above	23 (2.4)
Occupation now or before retirement	
Peasants	475 (49.6)
Workers	112 (11.7)
Others	109 (11.4)
Public officials/units	107 (11.2)
Sole trader/freelancer	100 (10.4)
Business managers/office staff	55 (5.7)
Monthly income (CNY[a])	
≤2000	472 (49.3)
2001-4000	286 (29.9)
4001-6000	133 (13.9)
6001-8000	44 (4.6)
≥8001	23 (2.4)
Urban/rural	
Countryside	559 (58.4)
Cities and towns	399 (41.6)
Number of children	
2	452 (47.2)
3	192 (20.0)
1	168 (17.5)

Characteristics	Values, n (%)
4	87 (9.1)
5	30 (3.1)
0	18 (1.9)
≥6	11 (1.1)
Self-assessed health status	
General	440 (45.9)
Better	288 (30.1)
Rather poor	170 (17.7)
Rare	44 (4.6)
Very poor	16 (1.7)

[a] 1 CNY=US $0.14.

Structural Model

The CFA model was fit to the data (Table S2 in Multimedia Appendix 3): χ^2_{231}=525.021, GFI=0.958, AGFI=0.945, IFI=0.980, CFI=0.980, NNFI=0.976, and RMSEA=0.036. The model fit appeared to be good, and all items met the criteria proposed by Hu and Bentler [68] and Kline [69].

According to Bagozzi and Youjae [70] and Hair [71], composite reliability (ρ)≥0.60 and average variance extracted ≥0.50 indicate good internal consistency and convergent validity. We therefore used composite reliability to measure the reliability of the constructs, with the following results: intergenerational support=0.887, online medical information=0.896, trust=0.869, PU=0.888, PEOU=0.906, eHealth literacy=0.894, and intention to seek medical care on the internet=0.896. As shown in Table S3 in Multimedia Appendix 3, the factor loadings for all question items were greater than 0.60. Convergent validity was assessed using the average variance extracted, which yielded the following values: 0.723 for intergenerational support, 0.684 for online medical information, 0.690 for trust, 0.726 for PU, 0.659 for PEOU, 0.739 for eHealth literacy, and 0.742 for intention to seek medical care on the internet. These results indicate good convergent validity. As shown in Table S4 in Multimedia Appendix 3, the square root of the average variance extracted for each construct is greater than the correlation coefficients between the variables, indicating good discriminant validity [72].

Based on the final CFA model, a full structural equation modeling was conducted to test the hypotheses. The model fit indices (Table S5 in Multimedia Appendix 3) were as follows: χ^2_{179}=498.30, GFI=0.954, AGFI=0.940, IFI=0.975, CFI=0.975, Tucker-Lewis Index=0.971, and RMSEA=0.043. These indices indicate a good model fit, and the structural equation modeling approach allowed the constructed and hypothesized model to be tested in a more satisfactory manner.

Association of Intergenerational Support and Other Variables and Intention to Seek Medical Care on the Internet

Table 2 shows the results of the path analysis. Intergenerational support was positively associated with PEOU, PU, trust, and older adults’ intention to seek medical care on the internet (β=.292, *P*<.001; β=.437, *P*<.001; β=.322, *P*<.001; and β=.354, *P*<.001, respectively); therefore, hypotheses H1a, H1b, H1c, and H1d were supported. PEOU and PU were positively associated with older adults’ trust in internet-based medical services (β=.263, *P*<.001 and β=.261, *P*<.001, respectively), supporting H2a and H2b. PEOU and PU were also positively correlated with older adults’ intention to seek medical care on the internet (β=.099, *P*=.002 and β=.204, *P*<.001, respectively), supporting H3a and H3b. Trust was positively correlated with older adults’ intention to seek medical care on the internet (β=.197, *P*<.001), supporting H4. Online medical information was also positively correlated with older adults’ intention to seek medical care on the internet (β=.109, *P*<.001), supporting H5.

Table 2. Results for direct relationships.

Relationship	Path coefficient	SE	Critical ratio	*P* value
PEOU ← IS[a]	0.292	0.037	8.204	<.001
PU ← IS	0.437	0.035	12.388	<.001
T ← IS	0.322	0.034	8.543	<.001
T ← PEOU[b]	0.263	0.029	8.003	<.001
T ← PU[c]	0.261	0.032	7.345	<.001
SMCI[d] ← IS	0.354	0.04	9.265	<.001
SMCI ← T[e]	0.197	0.046	4.961	<.001
SMCI ← PEOU	0.099	0.032	3.088	.002
SMCI ← PU	0.204	0.037	5.809	<.001
SMCI ← OMI[f]	0.109	0.034	3.781	<.001

[a] IS: intergenerational support.

[b] PEOU: perceived ease of use.

[c] PU: perceived usefulness.

[d] SMCI: intention to seek medical care on the internet

[e] T: Trust

[f] OMI: online medical information.

Table 3 shows the mediating relationship between PU, PEOU, and trust. The total effect value of the relationship between intergenerational support and willingness to use internet-based medical services was 0.573, while the direct effect value was 0.354. Both had a positive 95% CI and a *P* value of less than 0.05, indicating that the total and direct effects were significant. The indirect effect of trust was 0.063 (95% CI 0.036-0.099). More specifically, trust partially mediated the association between intergenerational support and older adults' willingness to use internet-based medical services. PEOU, PU, and trust also played a chain mediating role between intergenerational support and willingness to use internet-based medical services, with specific mediation effect sizes of 0.015 and 0.022, respectively. The 95% CI for the PEOU path was 0.008-0.027, and for the PU path, it was 0.012-0.036. The *P* values for both chain mediation paths were less than 0.05, suggesting that intergenerational support can enhance PEOU and PU among older adults. This, in turn, enhances trust, which subsequently increases the willingness to use internet-based medical services.

Table 3. Total, direct, and indirect effects.[a]

Mediation effect and path	Estimate	95% CI	*P* value
Total effect			
IS[b] → SMCI[c]	0.573	0.509-0.632	<.001
Direct effect			
IS → SMCI	0.354	0.278-0.428	<.001
Indirect effect			
IS → T[d] → SMCI	0.063	0.036-0.099	<.001
IS → PEOU[e] → SMCI	0.029	0.009-0.054	0.005
IS → PU[f] → SMCI	0.089	0.050-0.130	<.001
IS → PEOU → T → SMCI	0.015	0.008-0.027	<.001
IS → PU → T → SMCI	0.022	0.012-0.036	<.001

[a] Standardized estimation of 5000 bootstrap samples.

[b] IS: intergenerational support.

[c] SMCI: intention to seek medical care on the internet.

[d] T: trust.

[e] PEOU: perceived ease of use.

[f] PU: perceived usefulness.

Table 4 shows the moderating relationships of eHealth literacy. The interaction term between intergenerational support and eHealth literacy was a significant positive predictor of PEOU (β=.177, *P*<.001) and PU (β=.073, *P*=.018). This suggests that eHealth literacy moderates the effect of intergenerational support on PEOU and PU, thereby supporting H6a and H6b. The interaction term between intergenerational support and eHealth literacy was a significant positive predictor of trust (β=.09, *P*<.01), suggesting that eHealth literacy moderates the effect of intergenerational support on trust, thus supporting H6c. The interaction term between intergenerational support and eHealth literacy had a significant positive predictive effect on the intention to seek medical care on the internet (β=.124, *P*<.001). This indicates that eHealth literacy moderates the influence of intergenerational support on the intention to seek medical care on the internet, thereby supporting H6d. A further simple slope analysis was then conducted (Figure 1). The slope was shallower when eHealth literacy was low (solid line) and steeper when eHealth literacy was high (dashed line). This suggests that intergenerational support was a more significant positive predictor of the intention to seek medical care on the internet, PEOU, PU, and trust at high levels of eHealth literacy.

Table 4. Moderated effect.

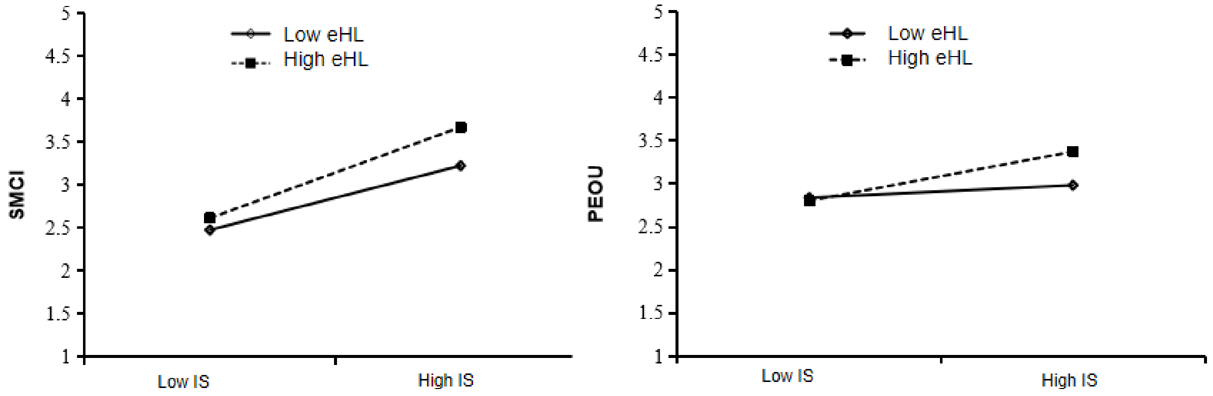
Path	Intention to seek medical care on the internet, β (*P* value)	Perceived ease of use, β (*P* value)	Perceived usefulness, β (*P* value)	Trust, β (*P* value)
Intergenerational support	.447 (<.001)	.183 (<.001)	.337 (<.001)	.389 (<.001)
Electronic health literacy[a]	.15 (<.001)	.094 (.002)	.167 (<.001)	.173 (<.001)
IS[b] × eHL[c]	.124 (<.001)	.177 (<.001)	.073 (.018)	.09 (.003)

[a] Electronic health literacy was used as a moderating variable in the model after controlling for the variables age, gender, education, residential status, occupation now or before retirement, monthly income, urban /rural, number of children, self-assessed health status.

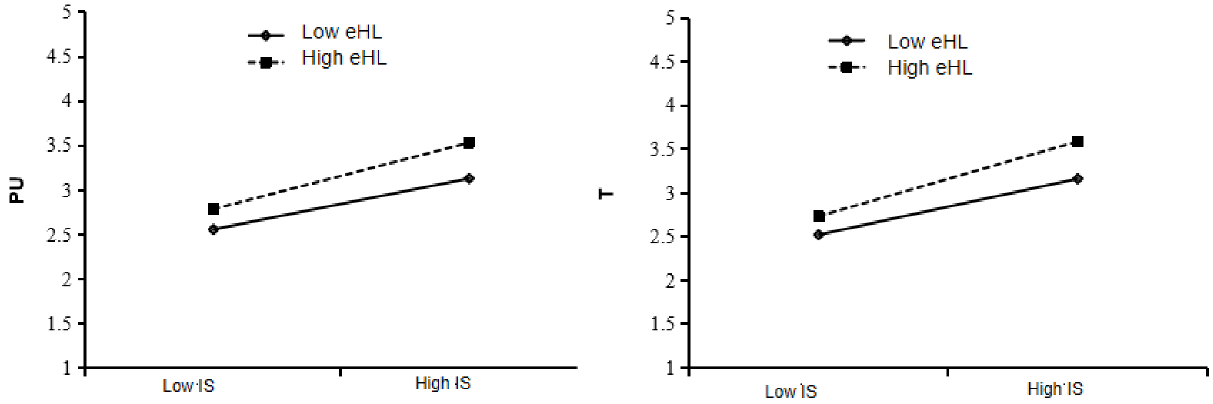
[b] IS: intergenerational support.

[c] eHL: eHealth literacy.

Figure 1. Simple slope plot. eHL: eHealth literacy; IS: intergenerational support; PEOU: perceived ease of use; PU: perceived usefulness; SMCI: intention to seek medical care on the internet; T: trust.



The moderating role of eHL between IS and SMCI The moderating role of eHL between IS and PEOU



The moderating role of eHL between IS and PU The moderating role of eHL between IS and T

Discussion

Principal Findings

Older people are often referred to as digital refugees [73]. A lack of digital skills may contribute to the risk of widening health disparities and unequal access to services [74]. Studies have shown that many older individuals face numerous barriers to using smartphones, often due to a lack of knowledge about apps and their features [75] and technology anxiety [76]. Most older adult patients are also accustomed to face-to-face outpatient services, making the adoption of internet-based health care not easily achievable for this population. Although the digital divide hinders active aging by reducing the social participation, as well as the physical and mental health, of older adults [77], the active bridging provided by intergenerational support helps older adults overcome their fears, improve their digital skills, and enhance their sense of well-being and accessibility. This, in turn, facilitates their integration into the digital age by fostering a positive and optimistic mindset that promotes effective active aging [78].

Intergenerational support to address the technical aspect occurs when children explain to older adults how to operate the online medical platform, help them understand error messages, and provide tips on various operations such as registration, logging in, booking appointments, and online counseling [79]. This support involves not only technical guidance but also spiritual care and comfort. This kind of benign emotional exchange can promote older people's perception of satisfaction, enhancing their acceptance of new technologies. Therefore, online medical care with intergenerational support is not only a reflection of filial piety but also an endorsement of positive intergenerational relationships. Children's counseling and assistance can improve the cognitive abilities [80] and overall health awareness of older adults, alleviate their fear of technology, reduce digital loneliness, and strengthen their confidence in using internet-based medical services.

The results of this study show that the higher the level of intergenerational support, the greater the PEOU and usefulness of internet health care platforms among older adults, which in turn enhances their willingness to access these services. Through the help and guidance of family members, older adults are able to master internet technologies more quickly and effectively utilize these tools for health management. This support not only increased older adults' PEOU of internet-based health care services but also boosted their confidence and security, leading to a greater willingness to try and use these services. In addition, this support implies encouragement and companionship on an emotional and psychological level [81]. When older adults feel the concern and attention from their family members or later generations regarding their health problems, and the possibility of obtaining a convenient and professional health care experience through internet health care services, they are more likely to develop positive psychological expectations that these services are valuable and can practically solve their health issues [82]. This positive cognitive shift directly enhances older adults' acceptance of and reliance on internet health care services, thereby strengthening their willingness to seek medical care.

There is also a transformational path between social support factors that generate the intention to adopt internet-based medical services; trust mediates between intergenerational support and the intention to adopt internet-based medical services. Trust, in this context, refers to the extent to which an individual relies on an internet-based medical service to meet health management needs [83], encompassing aspects such as the professionalism of the health care service, platform security, and a positive user experience. Cao et al [84] suggested that mistrust is a common reason for digital disengagement and that trust is crucial for an individual's ongoing behavior when using information technology. This is also true in the internet-based medical service environment [85]. Studies have shown that the realization of internet-based medical services is divided into 2 phases: first, patients develop perceptions about the technical aspects of the health care platform, including its functionality, usefulness, and convenience [86]; second, patients evaluate the trustworthiness of the care provided and the service process [87]. Strong behavioral intentions only occur if older adults have a positive perception of the trustworthiness of the care.

Based on previous studies, this study further clarifies the influence of intergenerational support on older adults' intention to adopt internet-based medical services. It constructs and validates an integrated model of the effect of health communication: intergenerational support → PEOU → PU → trust → intention to adopt internet-based medical services. As a new approach to medical care, internet-based medical services improve convenience and efficiency. With intergenerational support, older adults perceive the technical aspects (PEOU and PU) of these services more positively. This, in turn, enhances their trust and intention to use internet-based medical services, facilitating the integration of online care into health care.

Intergenerational support had a greater effect on the willingness of older adults to use internet-based medical services among those with high eHealth literacy. According to Jung et al [55], eHealth literacy enhances older adults' ability to manage health-related issues effectively. Health literacy has also been positively associated with help-seeking intentions [88], older adults' choice of physician, and their understanding of the physician's recommendations [89]. Older adults with high eHealth literacy are thus more likely to have a higher intention to seek medical care on the internet when they receive support from their children and are more likely to participate in internet-based medical services actively for a better health care experience [90].

Previous studies have shown that PU and trust in online medical information [91] can improve eHealth literacy. However, this study demonstrated that the degree of intergenerational support for perceived technology, as well as trust in internet-based medical services, is deeper when older adults have higher eHealth literacy. According to Nie et al [92], people with higher eHealth literacy better assess the usefulness of searching for health information online. However, it may be that eHealth literacy increases older adults' interest in health, their knowledge, and their expectations of health care [93] while inducing confidence and a sense of the efficacy of internet-based medical services.

Much of the health and medical information provided by internet-based medical service platforms originates from authoritative health care institutions and professionals, ensuring a high degree of accuracy. These platforms can support long-term disease management and health monitoring [94], allowing older people to access the health information they need. Bundorf et al [95] argued that this shift to searching for medical information online has obvious impacts on health behaviors as well as on treatment choices [96], improving older adults' health beliefs and motivating them to engage in healthy activities [97]. There is also evidence that online medical information not only encourages self-initiated participation in health care [98] but also increases the likelihood that older adults will take an active role in disease management. It is thus reasonable to believe, based on the present results, that online medical information can play an active role in disease management and increase the willingness of older adults to use internet-based medical services.

Limitations

Although this study initially verified the proposed theoretical model, certain limitations remain. First, the study focused exclusively on the PEOU and PU of internet-based medical services, leaving other related perceptions (eg, risks and benefits) unexplored. Moreover, it considered only the factors influencing older adults' willingness to use internet-based medical services, without addressing motivation, sustained willingness to use, or actual usage. Finally, this study examined only the influence of the core variable of intergenerational support. Other factors, such as peer influence, social influence (eg, doctors' recommendations), and subjective norms, could be explored in greater depth and incorporated into models in future research.

Implications

Examining technology use among older adults is also a valuable starting point for addressing the range of social issues associated with aging [99]. This study's focus on the needs of older adults regarding internet-based medical services, as well as their intentions to use such services, offers constructive insights for improving health equity and enhancing the inclusiveness and equity of internet-based medical services in particular [100]. Perceptions of and trust in technology are crucial factors in facilitating older adults' adoption of internet-based medical services. Efforts to encourage older adults to use these services may first need to ensure that they recognize such services as not only useful and easy to use but also trustworthy.

Medical institutions and internet-based medical platforms can take the following measures: in platform design, they should fully consider the characteristics and needs of older adults, implement age-adapted designs, lower the barriers to use, and create an age-friendly online environment. Additionally, medical institutions and platforms must establish a highly credible environment by ensuring transparency in the treatment process and maintaining honesty and openness with patients. Protecting the personal information of older adults and preventing data breaches are essential to enhancing their confidence in using internet-based health care services.

Furthermore, the role of intergenerational support in enhancing older adults' intention to seek medical care online must be emphasized. This highlights the importance of understanding the barriers older adults may face when using internet-based medical services, as well as the critical role children/young adults can play in facilitating the adoption of new technologies by older generations.

Within families, the younger generation should take the initiative to help older adults adapt to digital technology by offering emotional support and technical guidance while leveraging the advantages of the internet for health management.

The use of internet-based medical services by older adults holds significant value for management. First, it improves management efficiency. Internet-based medical services simplify traditional medical processes, enhancing hospital management efficiency. These platforms can allocate medical resources more effectively based on the health status and needs of older adults, optimizing resource utilization. Second, it fosters an innovative service model. Internet-based medical platforms can leverage technologies such as big data and artificial intelligence to deliver more precise medical services. Additionally, they can collaborate with other industries, such as care services for older adults, to innovate and expand service models. Third, it facilitates policy implementation. As a platform for policy dissemination, internet-based medical services can help raise awareness among older adults about national health and pension policies, thereby supporting the effective implementation of these policies. Fourth, it reflects social value. Internet-based medical services address the health needs of older adults, demonstrating social care and respect for older adults. This contributes to the transformation and upgrading of the national health care industry, ultimately improving the overall quality of medical services.

Conclusions

Population aging is a long-term national issue in China, and the availability of medical services is directly linked to the quality of life for older adults. This study provides evidence that intergenerational support is a key factor in promoting the adoption of internet health care services by older adults. Encouraging children to offer guidance and assistance to their parents is essential. To increase older adults' willingness to use internet health care, future efforts should focus not only on intergenerational support but also on improving the PEOU and usefulness of the technology, building trust, and fostering eHealth literacy among older adults.

Acknowledgments

This study was supported by the Project of Humanities and Social Sciences Research of the Ministry of Education (grant 23XJC860002) and the Fundamental Research Fund for the Central Universities (grant 2021CDJSKJC33).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[ZIP File (Zip Archive), 182 KB - jmir_v27i1e65065_app1.zip]

Multimedia Appendix 2
Data (centralized processing).
[ZIP File (Zip Archive), 190 KB - jmir_v27i1e65065_app2.zip]

Multimedia Appendix 3
Additional analysis.
[ZIP File (Zip Archive), 19 KB - jmir_v27i1e65065_app3.zip]

References

1. Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. J Med Internet Res 2011 Dec 14;13(4):e110 [FREE Full text] [doi: 10.2196/jmir.1796] [Medline: 22169526]
2. Zhang C. Smartphones and telemedicine for older people in China: opportunities and challenges. Digit Health 2022;8:20552076221133695 [FREE Full text] [doi: 10.1177/20552076221133695] [Medline: 36353695]
3. China Internet Network Information Center (CNNIC). The 52nd China statistical report on the development of internet in China. CNNIC. 2023 Aug 28. URL: https://www.cnnic.cn/n4/2023/0828/c199-10830.html [accessed 2024-07-14]
4. World Health Organization (WHO). Draft global strategy on digital health 2020-2024. WHO. Geneva, Switzerland: WHO; 2020. URL: https://www.who.int/docs/default-source/documents/gs4dh0c510c483a9a42b1834a8f4d276c6352.pdf [accessed 2024-07-14]

5. Resnick B, Young HM, Fick DM, Kagan SH. Making care for older people the choice of nurses today, tomorrow, and forever. Geriatr Nurs 2022;46:A1-A2 [FREE Full text] [doi: 10.1016/j.gerinurse.2022.04.014] [Medline: 35606212]
6. Sun X, Yan W, Zhou H, Wang Z, Zhang X, Huang S, et al. Internet use and need for digital health technology among the elderly: a cross-sectional survey in China. BMC Public Health 2020 Sep 11;20(1):1386 [FREE Full text] [doi: 10.1186/s12889-020-09448-0] [Medline: 32917171]
7. 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]
8. Klaver NS, van de Klundert J, van den Broek RJGM, Askari M. Relationship between perceived risks of using mHealth applications and the intention to use them among older adults in the Netherlands: cross-sectional study. JMIR Mhealth Uhealth 2021 Aug 30;9(8):e26845 [FREE Full text] [doi: 10.2196/26845] [Medline: 34459745]
9. Palas JU, Sorwar G, Hoque MR, Sivabalan A. Factors influencing the elderly's adoption of mHealth: an empirical study using extended UTAUT2 model. BMC Med Inform Decis Mak 2022 Jul 24;22(1):191 [FREE Full text] [doi: 10.1186/s12911-022-01917-3] [Medline: 35871682]
10. Maxham JG. Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions. Journal of Business Research 2001 Oct;54(1):11-24. [doi: 10.1016/s0148-2963(00)00114-4]
11. Ma Y, Luo M. Older people's intention to use medical apps during the COVID-19 pandemic in China: an application of the Unified Theory of Acceptance and Use of Technology (UTAUT) model and the Technology of Acceptance Model (TAM). Ageing and Society 2022 Oct 17;44(7):1515-1532. [doi: 10.1017/s0144686x22000423]
12. Tu J, Luo SC, Lee Y, Shih M, Chiu S. Exploring usability and patient attitude towards a smart hospital service with the Technology Acceptance Model. Int J Environ Res Public Health 2022 May 16;19(10):6059 [FREE Full text] [doi: 10.3390/ijerph19106059] [Medline: 35627595]
13. Fu Y, Tang T, Long J, Lin B, Li J, Quan G, et al. Factors associated with using the internet for medical information based on the doctor-patient trust model: a cross-sectional study. BMC Health Serv Res 2021 Nov 24;21(1):1268 [FREE Full text] [doi: 10.1186/s12913-021-07283-6] [Medline: 34819044]
14. Ren D, Ma B. Effectiveness of interactive tools in online health care communities: social exchange theory perspective. J Med Internet Res 2021 Mar 12;23(3):e21892 [FREE Full text] [doi: 10.2196/21892] [Medline: 33709940]
15. Ren D, Ma B. Influences of governance mechanisms on patients' usage intention: a study on web-based consultation platforms. Health Informatics J 2023;29(1):14604582231153509 [FREE Full text] [doi: 10.1177/14604582231153509] [Medline: 36657942]
16. Mansson L, Wiklund M, Öhberg F, Danielsson K, Sandlund M. Co-creation with older adults to improve user-experience of a smartphone self-test application to assess balance function. Int J Environ Res Public Health 2020 May 26;17(11):3768 [FREE Full text] [doi: 10.3390/ijerph17113768] [Medline: 32466484]
17. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use medical apps among older adults in the Netherlands: cross-sectional study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]
18. Morgan SP, Hirosima K. The persistence of extended family residence in Japan: anachronism or alternative strategy? American Sociological Review 1983 Apr;48(2):269-281. [doi: 10.2307/2095111]
19. Chen J, Jordan LP. Intergenerational support and life satisfaction of young-, old- and oldest-old adults in China. Aging Ment Health 2018 Mar;22(3):412-420. [doi: 10.1080/13607863.2016.1261798] [Medline: 27918203]
20. Kim SD. Application and challenges of the technology acceptance model in elderly healthcare: insights from ChatGPT. Technologies 2024 May 13;12(5):68. [doi: 10.3390/technologies12050068]
21. McGaughey RE, Zeltmann SM, McMurtrey ME. Motivations and obstacles to smartphone use by the elderly: developing a research framework. IJEF 2013;7(3/4):177. [doi: 10.1504/ijef.2013.058601]
22. Yuan Z, Jia G. Profiling the digital divide of the elderly based on internet big data: evidence from China. Data Science and Management 2021 Sep;3:33-43. [doi: 10.1016/j.dsm.2021.10.001]
23. Almulhem JA. Factors, barriers, and recommendations related to mobile health acceptance among the elderly in Saudi Arabia: a qualitative study. Healthcare (Basel) 2023 Nov 23;11(23):3024. [doi: 10.3390/healthcare11233024] [Medline: 38063592]
24. Wang J, Katz I, Li J, Wu Q, Dai C. Mobile digital divide and older people’s access to ‘Internet plus social work’: implications from the COVID-19 help-seeking cases. Asia Pacific Journal of Social Work and Development 2020 Nov 26;31(1-2):52-58. [doi: 10.1080/02185385.2020.1850332]
25. Sharifi S, Babaei Khorzoughi K, Khaledi-Paveh B, Rahmati M. Association of intergenerational relationship and supports with cognitive performance in older adults: a systematic review. Geriatr Nurs 2023;52:146-151. [doi: 10.1016/j.gerinurse.2023.05.014] [Medline: 37307634]
26. Wei D, Xu A, Wu X. The mediating effect of trust on the relationship between doctor-patient communication and patients' risk perception during treatment. Psych J 2020 Jun;9(3):383-391. [doi: 10.1002/pchj.327] [Medline: 31795011]
27. Lee OE, Kim D. Bridging the digital divide for older adults via intergenerational mentor-up. Research on Social Work Practice 2018 Nov 05;29(7):786-795. [doi: 10.1177/1049731518810798]

28. Ma T, Zhang S, Zhu S, Ni J, Wu Q, Liu M. The new role of nursing in digital inclusion: reflections on smartphone use and willingness to increase digital skills among Chinese older adults. Geriatr Nurs 2022;48:118-126 [FREE Full text] [doi: 10.1016/j.gerinurse.2022.09.004] [Medline: 36155310]
29. Pan Z, Chen J. Association of received intergenerational support with subjective well-being among elderly: the mediating role of optimism and sex differences. Int J Environ Res Public Health 2022 Jun 22;19(13):7614 [FREE Full text] [doi: 10.3390/ijerph19137614] [Medline: 35805273]
30. Li H, Kostka G. Navigating the digital age: the gray digital divide and digital inclusion in China. Media, Culture & Society 2024 Feb 21;46(6):1181-1199. [doi: 10.1177/01634437241229382]
31. Yeh K, Yi C, Tsao W, Wan P. Filial piety in contemporary Chinese societies: a comparative study of Taiwan, Hong Kong, and China. International Sociology 2013 May 14;28(3):277-296. [doi: 10.1177/0268580913484345]
32. Correa T. Bottom-up technology transmission within families: exploring how youths influence their parents' digital media use with dyadic data. J Commun 2013 Dec 02;64(1):103-124. [doi: 10.1111/jcom.12067]
33. Ekström KM. Parental consumer learning or ‘keeping up with the children’. J of Consumer Behaviour 2007 Sep 13;6(4):203-217. [doi: 10.1002/cb.215]
34. Ma J, Fang G, Guo K. Bridging the digital divide: the influence of digital feedback on the digital capabilities of the rural elderly. Information Development 2023 Dec 28;39:02666669231222208. [doi: 10.1177/02666669231222208]
35. 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]
36. Harrigan M, Feddema K, Wang S, Harrigan P, Diot E. How trust leads to online purchase intention founded in perceived usefulness and peer communication. J of Consumer Behaviour 2021 Mar 03;20(5):1297-1312. [doi: 10.1002/cb.1936]
37. Acharya N, Sassenberg A, Soar J. Consumers’ behavioural intentions to reuse recommender systems: assessing the effects of trust propensity, trusting beliefs and perceived usefulness. JTAER 2022 Dec 22;18(1):55-78. [doi: 10.3390/jtaer18010004]
38. Martínez-Navalón J, Fernández-Fernández M, Alberto FP. Does privacy and ease of use influence user trust in digital banking applications in Spain and Portugal? Int Entrep Manag J 2023 Feb 22;19(2):781-803. [doi: 10.1007/s11365-023-00839-4]
39. Kim Y, Park Y, Choi J. A study on the adoption of IoT smart home service: using value-based adoption model. Total Quality Management & Business Excellence 2017 Apr 12;28(9-10):1149-1165. [doi: 10.1080/14783363.2017.1310708]
40. Stone R, Baker Eveleth L. User's perceptions of perceived usefulness, satisfaction, and intentions of mobile application. IJMC 2020;18(1):1-18. [doi: 10.1504/ijmc.2020.10025192]
41. Naidoo R, Leonard A. Perceived usefulness, service quality and loyalty incentives: effects on electronic service continuance. SAJBM 2007 Sep 30;38(3):39-48. [doi: 10.4102/sajbm.v38i3.587]
42. Wu B. Patient continued use of online health care communities: web mining of patient-doctor communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
43. Chen Y, Xu Q. The willingness to use mobile health and its influencing factors among elderly patients with chronic heart failure in Shanghai, China. Int J Med Inform 2021 Dec 16;158:104656. [doi: 10.1016/j.ijmedinf.2021.104656] [Medline: 34933173]
44. Schoorman FD, Mayer RC, Davis JH. An integrative model of organizational trust: past, present, and future. AMR 2007 Apr;32(2):344-354. [doi: 10.5465/amr.2007.24348410]
45. Doney PM, Cannon JP. An examination of the nature of trust in buyer-seller relationships. Journal of Marketing 1997 Apr;61(2):35-51. [doi: 10.2307/1251829]
46. Bu W, Xing L, Xiao M, Chang P, Du Y, Xie H. A survey on the willingness of Ganzhou residents to participate in "internet + nursing services" and associated factors. J Multidiscip Healthc 2022;15:897-906 [FREE Full text] [doi: 10.2147/JMDH.S351071] [Medline: 35509727]
47. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011 Feb 16;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
48. Roshandel D, Rezailashkajani M, Ansari S, Zali MR. Internet use by a referral gastroenterology clinic population and their medical information preferences. Int J Med Inform 2005 Jul;74(6):447-459. [doi: 10.1016/j.ijmedinf.2005.05.003] [Medline: 15963757]
49. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]
50. Smith AD, Manna DR. Exploring the trust factor in e‐medicine. Online Information Review 2004 Oct;28(5):346-355. [doi: 10.1108/14684520410564271]
51. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
52. 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]

53. Peng Y, Yin P, Deng Z, Wang R. Patient-physician interaction and trust in online health community: the role of perceived usefulness of health information and services. Int J Environ Res Public Health 2019 Dec 24;17(1):139 [FREE Full text] [doi: 10.3390/ijerph17010139] [Medline: 31878145]
54. 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]
55. Jung SO, Son YH, Choi E. E-health literacy in older adults: an evolutionary concept analysis. BMC Med Inform Decis Mak 2022 Jan 31;22(1):28 [FREE Full text] [doi: 10.1186/s12911-022-01761-5] [Medline: 35101005]
56. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
57. Lanxin W, Yan Z, Yutong T, Lixue M, Li L, Ting Z. Potential profiling of self-management skills in older co-morbid patients. BMC Geriatr 2024 Jun 25;24(1):555 [FREE Full text] [doi: 10.1186/s12877-024-05137-4] [Medline: 38918703]
58. Liu S, Zhao H, Fu J, Kong D, Zhong Z, Hong Y, et al. Current status and influencing factors of digital health literacy among community-dwelling older adults in Southwest China: a cross-sectional study. BMC Public Health 2022 May 17;22(1):996. [doi: 10.1186/s12889-022-13378-4] [Medline: 35581565]
59. Arora S, Huda RK, Verma S, Khetan M, Sangwan RK. Challenges, barriers, and facilitators in telemedicine implementation in India: a scoping review. Cureus 2024 Aug;16(8):e67388. [doi: 10.7759/cureus.67388] [Medline: 39310647]
60. Wong AKC, Bayuo J, Wong FKY. Investigating predictors of self-care behavior among homebound older adults: the role of self-efficacy, eHealth literacy, and perceived social support. J Nurs Scholarsh 2022 May;54(3):278-285. [doi: 10.1111/jnu.12730] [Medline: 34766694]
61. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
62. Lang FR, Schütze Y. Adult children’s supportive behaviors and older parents’ subjective well–being—a developmental perspective on intergenerational relationships. Journal of Social Issues 2003 Feb 12;58(4):661-680. [doi: 10.1111/1540-4560.00283]
63. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997 Feb 22;314(7080):572 [FREE Full text] [doi: 10.1136/bmj.314.7080.572] [Medline: 9055718]
64. de Boer MJ, Versteegen GJ, van Wijhe M. Patients' use of the internet for pain-related medical information. Patient Educ Couns 2007 Sep;68(1):86-97. [doi: 10.1016/j.pec.2007.05.012] [Medline: 17590563]
65. Fogel J, Nehmad E. Internet social network communities: risk taking, trust, and privacy concerns. Computers in Human Behavior 2009 Jan;25(1):153-160. [doi: 10.1016/j.chb.2008.08.006]
66. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: empirical study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
67. 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]
68. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
69. Kline R. Principles and Practice of Structural Equation Modeling. New York, NY: Guilford Publications; 2023.
70. Bagozzi RP, Yi Y. On the evaluation of structural equation models. JAMS 1988 Mar;16(1):74-94. [doi: 10.1007/bf02723327]
71. Hair J. Multivariate Data Analysis. 7th edition. London, UK: Pearson Prentice Hall; Feb 23, 2009.
72. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 1981 Feb;18(1):39. [doi: 10.2307/3151312]
73. Peine A, Neven L. From intervention to co-constitution: new directions in theorizing about aging and technology. Gerontologist 2019 Jan 09;59(1):15-21. [doi: 10.1093/geront/gny050] [Medline: 29850812]
74. Bakhtiar M, Elbuluk N, Lipoff JB. The digital divide: how COVID-19's telemedicine expansion could exacerbate disparities. J Am Acad Dermatol 2020 Nov;83(5):e345-e346 [FREE Full text] [doi: 10.1016/j.jaad.2020.07.043] [Medline: 32682890]
75. Hsieh M, Ho C, Lee I. Effects of smartphone numeric keypad designs on performance and satisfaction of elderly users. International Journal of Industrial Ergonomics 2022 Jan;87:103236. [doi: 10.1016/j.ergon.2021.103236]
76. Caspi A, Daniel M, Kavé G. Technology makes older adults feel older. Aging Ment Health 2019 Aug;23(8):1025-1030. [doi: 10.1080/13607863.2018.1479834] [Medline: 30475067]
77. Liu L, Wu F, Tong H, Hao C, Xie T. The digital divide and active aging in China. Int J Environ Res Public Health 2021 Dec 01;18(23):12675 [FREE Full text] [doi: 10.3390/ijerph182312675] [Medline: 34886400]
78. Amaral I, Daniel F. The use of social media among senior citizens in Portugal: active ageing through an intergeneration approach. Las Vegas, NV: Champringer International Publishing; 2018 Presented at: 4th International Conference, ITAP 2018, Held as Part of HCI International 2018; July 15-20, 2018; Las Vegas, NV p. 422-434. [doi: 10.1007/978-3-319-92034-4_32]

79. Martínez C, Olsson T. The warm expert—a warm teacher? Learning about digital media in intergenerational interaction. Convergence: The International Journal of Research into New Media Technologies 2022 Feb 03;28(6):1861-1877. [doi: 10.1177/13548565211070409]
80. Li Y, Guo M. Filial piety matters: a study of intergenerational supports and parental health. SSM Popul Health 2022 Jun;18:101096 [FREE Full text] [doi: 10.1016/j.ssmph.2022.101096] [Medline: 35493409]
81. Zhou C, Qian Y, Kaner J. A study on smart home use intention of elderly consumers based on technology acceptance models. PLoS One 2024;19(3):e0300574 [FREE Full text] [doi: 10.1371/journal.pone.0300574] [Medline: 38536849]
82. Wang J, Fu Y, Lou V, Tan SY, Chui E. A systematic review of factors influencing attitudes towards and intention to use the long-distance caregiving technologies for older adults. Int J Med Inform 2021 Sep;153:104536. [doi: 10.1016/j.ijmedinf.2021.104536] [Medline: 34325206]
83. Thom DH, Stanford Trust Study Physicians. Physician behaviors that predict patient trust. J Fam Pract 2001 Apr;50(4):323-328. [Medline: 11300984]
84. Cao X, Zhang H, Zhou B, Wang D, Cui C, Bai X. Factors influencing older adults' acceptance of voice assistants. Front Psychol 2024 Mar 07;15(2):1376207. [doi: 10.3389/fpsyg.2024.1376207] [Medline: 38515974]
85. Zhang X, Yan X, Cao X, Sun Y, Chen H, She J. The role of perceived e-health literacy in users' continuance intention to use mobile healthcare applications: an exploratory empirical study in China. Information Technology for Development 2017 Mar 09;24(2):198-223. [doi: 10.1080/02681102.2017.1283286]
86. Joshi K. A model of users' perspective on change: the case of information systems technology implementation. MIS Quarterly 1991 Jun;15(2):229-242. [doi: 10.2307/249384]
87. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
88. Suka M, Yamauchi T, Sugimori H. Help-seeking intentions for early signs of mental illness and their associated factors: comparison across four kinds of health problems. BMC Public Health 2016 Apr 07;16:301 [FREE Full text] [doi: 10.1186/s12889-016-2998-9] [Medline: 27056546]
89. Xie B. Older adults' health information wants in the internet age: implications for patient-provider relationships. J Health Commun 2009 Sep;14(6):510-524. [doi: 10.1080/10810730903089614] [Medline: 19731124]
90. Waterworth S, Honey M. On-line health seeking activity of older adults: an integrative review of the literature. Geriatr Nurs 2018;39(3):310-317. [doi: 10.1016/j.gerinurse.2017.10.016] [Medline: 29198622]
91. Campbell RJ, Nolfi DA. Teaching elderly adults to use the internet to access health care information: before-after study. J Med Internet Res 2005 Jun 30;7(2):e19 [FREE Full text] [doi: 10.2196/jmir.7.2.e19] [Medline: 15998610]
92. Nie L, Oldenburg B, Cao Y, Ren W. Continuous usage intention of mobile health services: model construction and validation. BMC Health Serv Res 2023 May 05;23(1):442 [FREE Full text] [doi: 10.1186/s12913-023-09393-9] [Medline: 37143005]
93. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and validity of the telephone-based eHealth literacy scale among older adults: cross-sectional survey. J Med Internet Res 2017 Oct 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
94. Yuan Y, Wang S, Tao C, Gu Z, Kitayama A, Yanagihara K, et al. Mapping trends and hotspots regarding the use of telenursing for elderly individuals with chronic diseases: a bibliometric analysis. Medicine (Baltimore) 2024 Mar 01;103(9):e37313 [FREE Full text] [doi: 10.1097/MD.0000000000037313] [Medline: 38428870]
95. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
96. Chen Y, Li C, Liang J, Tsai C. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. J Med Internet Res 2018 Feb 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]
97. Dutta-Bergman MJ. Health attitudes, health cognitions, and health behaviors among internet health information seekers: population-based survey. J Med Internet Res 2004 May 28;6(2):e15 [FREE Full text] [doi: 10.2196/jmir.6.2.e15] [Medline: 15249264]
98. Sassenberg K, Greving H. Internet searching about disease elicits a positive perception of own health when severity of illness is high: a longitudinal questionnaire study. J Med Internet Res 2016 Mar 04;18(3):e56 [FREE Full text] [doi: 10.2196/jmir.5140] [Medline: 26944335]
99. Amini R, Chee KH, Mendieta M, Parker S. Online engagement and cognitive function among older adults. Geriatr Gerontol Int 2019 Sep;19(9):918-923 [FREE Full text] [doi: 10.1111/ggi.13749] [Medline: 31368165]
100. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476 [FREE Full text] [doi: 10.2196/18476] [Medline: 32788144]

Abbreviations

AGFI: adjusted goodness of fit index

CFA: confirmatory factor analysis
CFI: comparative fit index
eHL: eHealth literacy
GFI: goodness of fit index
IFI: incremental fit index
KMO: Kaiser-Meier-Olkin
mHealth: mobile health
NNFI: nonnormed fit index
OMI: online medical information
PEOU: perceived ease of use
PU: perceived usefulness
RMSEA: root mean square error of approximation
SMCI: intention to seek medical care on the internet

Edited by A Schwartz; submitted 04.08.24; peer-reviewed by X Zou, Y Liu; comments to author 07.10.24; revised version received 28.10.24; accepted 29.11.24; published 06.01.25.

Please cite as:
Jin H, Qu Y
Association Between Intergenerational Support, Technology Perception and Trust, and Intention to Seek Medical Care on the Internet Among Chinese Older Adults: Cross-Sectional Questionnaire Study
J Med Internet Res 2025;27:e65065
URL: https://www.jmir.org/2025/1/e65065
doi:10.2196/65065
PMID:

Short Paper

Two-Layer Retrieval-Augmented Generation Framework for Low-Resource Medical Question Answering Using Reddit Data: Proof-of-Concept Study

Sudeshna Das[1*], PhD; Yao Ge[1*], PhD; Yuting Guo[1], MS; Swati Rajwal[2], MTech; JaMor Hairston[1], MS; Jeanne Powell[1], PhD; Drew Walker[1], PhD; Snigdha Peddireddy[3], MPH; Sahithi Lakamana[1], MS; Selen Bozkurt[1], PhD; Matthew Reyna[1], PhD; Reza Sameni[1,4], PhD; Yunyu Xiao[5], PhD; Sangmi Kim[6], PhD; Rasheeta Chandler[6], PhD; Natalie Hernandez[7], PhD; Danielle Mowery[8], PhD; Rachel Wightman[9], MD; Jennifer Love[10], MD; Anthony Spadaro[11], MD; Jeanmarie Perrone[12], MD; Abeed Sarker[1,4], PhD

[1]Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, United States
[2]Department of Computer Science and Informatics, Emory University, Atlanta, GA, United States
[3]Department of Behavioral, Social & Health Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA, United States
[4]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, United States
[5]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, United States
[6]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, United States
[7]Center for Maternal Health Equity, Morehouse School of Medicine, Atlanta, GA, United States
[8]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[9]Department of Emergency Medicine, Warren Alpert Medical School of Brown University, Providence, RI, United States
[10]Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States
[11]Department of Emergency Medicine, Rutgers New Jersey Medical School, Newark, NJ, United States
[12]Department of Emergency Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States
[*]these authors contributed equally

Corresponding Author:
Sudeshna Das, PhD
Department of Biomedical Informatics
School of Medicine
Emory University
101 Woodruff Circle
Atlanta, GA, 30322
United States
Phone: 1 4047270229
Email: sudeshna.das@emory.edu

Abstract

Background: The increasing use of social media to share lived and living experiences of substance use presents a unique opportunity to obtain information on side effects, use patterns, and opinions on novel psychoactive substances. However, due to the large volume of data, obtaining useful insights through natural language processing technologies such as large language models is challenging.

Objective: This paper aims to develop a retrieval-augmented generation (RAG) architecture for medical question answering pertaining to clinicians' queries on emerging issues associated with health-related topics, using user-generated medical information on social media.

Methods: We proposed a two-layer RAG framework for query-focused answer generation and evaluated a proof of concept for the framework in the context of query-focused summary generation from social media forums, focusing on emerging drug-related information. Our modular framework generates individual summaries followed by an aggregated summary to answer medical queries from large amounts of user-generated social media data in an efficient manner. We compared the performance of a quantized large language model (Nous-Hermes-2-7B-DPO), deployable in low-resource settings, with GPT-4. For this proof-of-concept study, we used user-generated data from Reddit to answer clinicians' questions on the use of xylazine and ketamine.

Results: Our framework achieves comparable median scores in terms of relevance, length, hallucination, coverage, and coherence when evaluated using GPT-4 and Nous-Hermes-2-7B-DPO, evaluated for 20 queries with 76 samples. There was no statistically significant difference between GPT-4 and Nous-Hermes-2-7B-DPO for coverage (Mann-Whitney U=733.0; n_1=37; n_2=39; P=.89 two-tailed), coherence (U=670.0; n_1=37; n_2=39; P=.49 two-tailed), relevance (U=662.0; n_1=37; n_2=39; P=.15 two-tailed), length (U=672.0; n_1=37; n_2=39; P=.55 two-tailed), and hallucination (U=859.0; n_1=37; n_2=39; P=.01 two-tailed). A statistically significant difference was noted for the Coleman-Liau Index (U=307.5; n_1=20; n_2=16; P<.001 two-tailed).

Conclusions: Our RAG framework can effectively answer medical questions about targeted topics and can be deployed in resource-constrained settings.

(J Med Internet Res 2025;27:e66220) doi:10.2196/66220

KEYWORDS

retrieval-augmented generation; substance use; social media; large language models; natural language processing; artificial intelligence; GPT; psychoactive substance

Introduction

Large language models (LLMs) present opportunities for solving complex biomedical natural language processing problems, such as medical question answering (MQA). However, operational challenges (eg, high computational resource requirements) hinder their real-life deployment and use. Another issue with LLM-generated text for MQA is "hallucination": generated text that is plausible-sounding but nonsensical or incorrect [1]. Chain-of-thought prompting [2], self-reflection [1], and retrieval-augmented generation (RAG) are forerunners in mitigating hallucination. RAG also aids in constraining generated texts and improves in-context learning [3]. LLMs in RAG frameworks have been used in the biomedical domain owing to the need for timely, accurate, and transparent responses [4]. As LLMs become increasingly integrated into clinical practice [5], it is important to ensure their operability in low-resource settings [6] while generating accurate and coherent texts.

We present a proof-of-concept study for a two-layer RAG framework for MQA that ingests user-generated medical information from Reddit. We used smaller, quantized, open-source LLMs that can run on personal computers without specialized hardware, allowing our framework to be used in low-resource settings, thus ensuring equitable access to timely medical information.

Methods

Study Design

We evaluated our proof-of-concept study in a setting where copious amounts of data are available for a topic but gathering insights and answering questions require substantial manual work—the topic of emerging drugs from Reddit. Reddit has ~52 million daily active users, is commonly used to study emerging medical themes [7], and features numerous discussions on the nonmedical uses of substances. Recently, Reddit data have been leveraged to study novel psychoactive substances since such information is not typically available elsewhere. We chose two substances that have gained attention recently—xylazine (because of its increasing impact and association with the US opioid crisis) and ketamine (because of its recent popularity as a treatment for depression). We collected all available data (~2.5 billion posts) from Reddit until December 31, 2023, and extracted all posts mentioning xylazine (n=177,684) and ketamine (n=7699) for our retrieval engine. Based on clinician-driven interests, we formulated 20 queries (Table 1).

Table 1. Queries used for evaluating the framework.

Query ID	Query
1	What are the side effects of xylazine?
2	What does xylazine do to the skin?
3	How does xylazine impact rehab?
4	What is xylazine withdrawal like?
5	What drugs contain xylazine?
6	What treatments work for xylazine?
7	What drugs are mixed or cut with xylazine?
8	What areas of the United States are impacted by xylazine?
9	How is xylazine different from pure heroin?
10	What is the general sentiment associated with xylazine?
11	Does narcan or naloxone work for xylazine overdose?
12	What are the side effects of ketamine?
13	What is ketamine withdrawal like?
14	What are k cramps like?
15	How do the users describe k hole?
16	Does ketamine work for depression?
17	What drugs are ketamine coused with recreationally?
18	Is ketamine effective for the treatment of suicidal behavior?
19	How can you treat ketamine addiction?
20	Does ketamine use cause cramps?

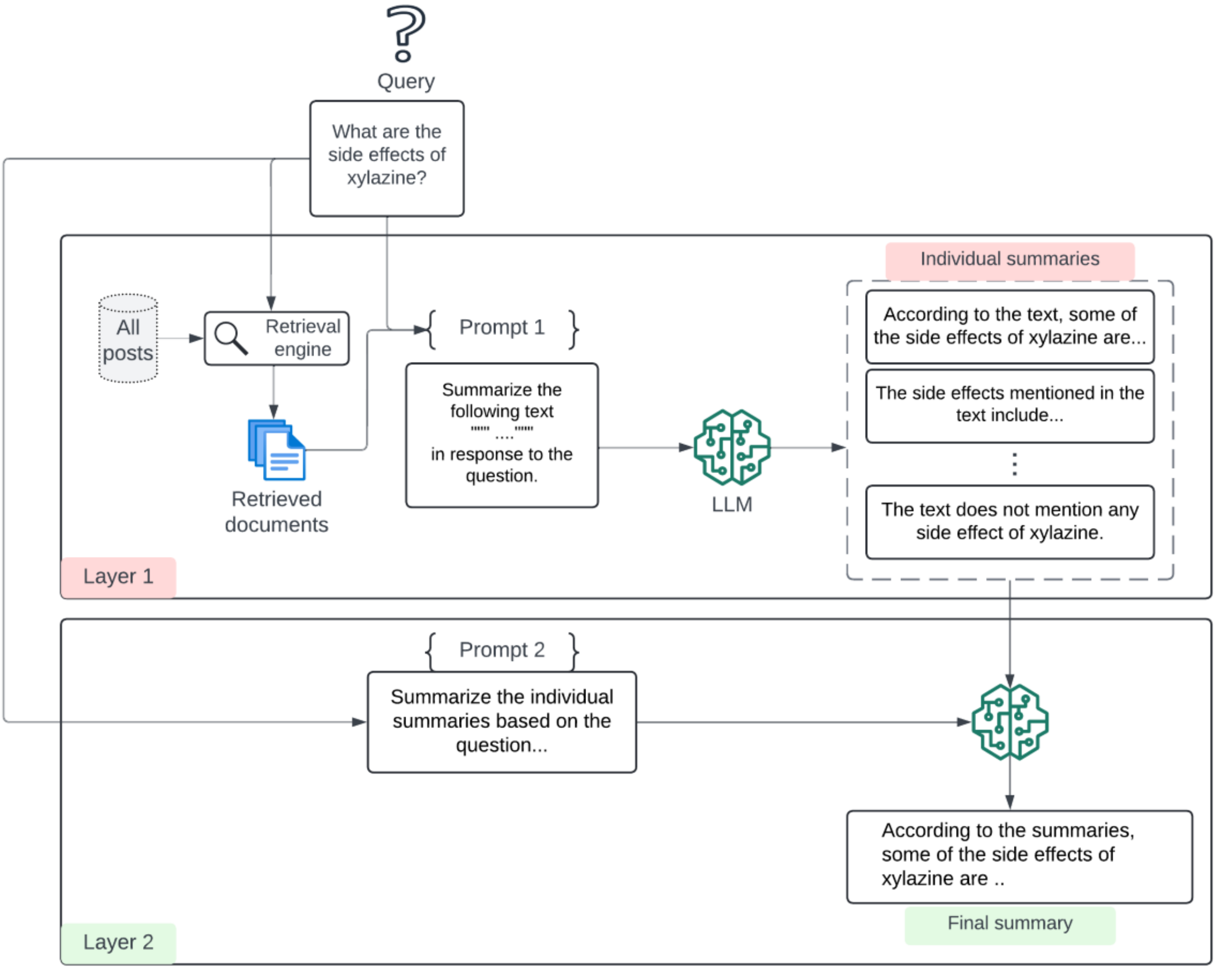
System Architecture

As depicted in Figure 1, the user submits a query to be parsed by the information retrieval engine, which returns a ranked list of documents. The top n documents are chosen to be sources for answer generation. In the first layer, the LLM is provided with (1) a query, (2) text from the retrieved documents, and (3) a prompt that embeds the text and instructs the LLM to summarize it (Multimedia Appendix 1). Since the prompt context window is finite, feeding the LLM all the retrieved text for answer generation is typically impossible. Even single documents can be too long. Thus, the framework allows for the specification of segment lengths for the retrieved text in each iteration, ensuring the framework is applied to relatively small LLMs with shorter context lengths. The first layer generates short, query-focused summaries (Figure 1). The LLM states if the retrieved text segment does not contain an answer to the question. Examples of this first-layer summarization are provided in Multimedia Appendix 2.

The second layer takes as input the original query and individual short summaries embedded within a second prompt (Multimedia Appendix 1) while ignoring summaries that the LLM states did not contain the answer. Figure 1 depicts an example of the final, synthesized summary.

We used the 8-bit quantized model Nous-Hermes-2-7B-DPO as our LLM, which is tuned on 1,000,000 high-quality instructions [8]. To test the performance of the proposed framework with larger models, we also performed an evaluation using GPT-4 [9].

Figure 1. Overview of the two-layer RAG framework. The first layer generates individual summaries based on retrieved posts relevant to the original query. The second layer generates the final summary based on the individual summaries generated in the first layer. LLM: large language model; RAG: retrieval-augmented generation.



RAG Architecture

Since the retrieval aspect is not our primary focus, we used a simple keyword-based approach using the default search settings provided by our information retrieval package "Whoosh," which uses Okapi BM25F ranking [10]. The top 50 retrieved documents were chosen for generating the first-layer summaries. This number may be adjusted without changes to the architecture. The number of text segments is typically higher when posts do not fit within the context window of the LLM after being embedded within the prompt.

Evaluation

Our evaluation focused on the architecture's summary generation quality, rather than retrieval performance. Commonly used automatic summary evaluation methods, such as Recall-Oriented Understudy for Gisting Evaluation (ROUGE) [11] and bilingual evaluation understudy (BLEU) [12], primarily focus on text overlap between generated and gold-standard summaries. In the absence of gold-standard summaries, subject matter experts manually and qualitatively evaluated the important nuances of generative summaries, which is impossible with ROUGE or BLEU. We used Likert-scale evaluations (Table 2). Each query–individual summary–final summary triplet was evaluated by ≥2 evaluators (at least a master's degree in medicine, public health, informatics, or allied fields). Overall, 21 experts generated 76 evaluations for 20 unique queries.

We also assessed the readability of the final summaries using the Coleman-Liau Readability Index (CLI) [13], which approximates the US grade level required to comprehend text.

We performed nonparametric tests for proportions (Mann-Whitney *U* test) with the null hypothesis (H_0: "The two populations are equal") to determine if the scores assigned to answers generated by GPT-4 and Nous-Hermes-2-7B-DPO vary significantly. All tests were performed using the *SciPy* package [14]. The null hypothesis (H_0) was rejected if $P<.05$ (two-tailed).

Table 2. Evaluation criteria and scales presented to annotators.

Criteria	Question	Evaluation scale
Coverage	Does the final summary accurately represent the information present in the original text?	• 5: Yes; the final summary covers all the important information present in the original text. • 4: Mostly; the final summary covers most, but not all of the important information. • 3: Somewhat; the final summary covers some of the important information, but also misses some of them. • 2: Not really; the final summary misses most of the important information. • 1: No; the final summary does not cover any of the important information present in the original text.
Coherence	Is the final summary coherent?	• 5: Yes; the final summary is easy to read and understand. • 4: Mostly; the final summary is readable, but not straightforward to understand. • 3: Somewhat; the final summary is readable but confusing. • 2: Not really; the final summary has some grammatical errors or nonsequiturs. • 1: No; the final summary is unintelligible or incomprehensible.
Relevance	Does the final summary answer the original question?	• 3: Yes; the summary answers the original question. • 2: Partially; the summary answers the original question, but not fully. • 1: No; the summary does not answer the original question.
Length	Is the length of the final summary appropriate?	• 3: Yes; the summary is appropriate in length. • 2: Somewhat; the summary could be shorter or longer. • 1: No; the summary is long-winded or too short.
Hallucination	Does the summary contain information not present in the original text?	• 0: No; the summary does not contain information not present in the original text. • 1: Yes; the summary contains information not present in the original text.

Ethical Considerations

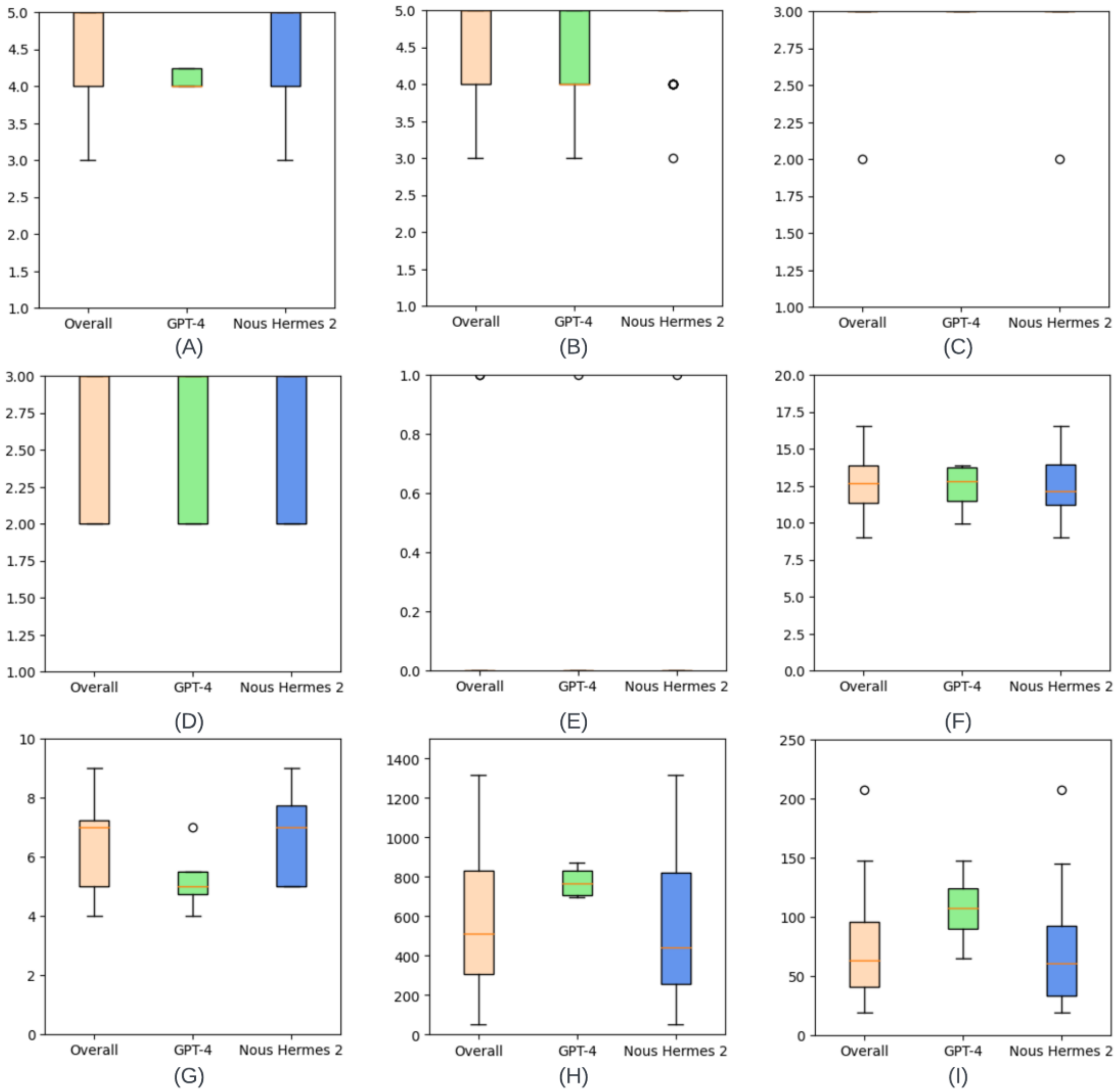
This study was deemed to be exempt from review per the Emory University Institutional Review Board's guidelines. The data used in this study are anonymous by default. We ensured that self-disclosed, personally identifiable information is not used by only reporting aggregated data. We removed posts that were deleted by the user.

Results

We conducted extensive expert evaluations of the generated answers for coverage, coherence, relevance, length, and hallucination (Figure 2). Annotators were not made aware of which LLM was used to generate the summaries for fair evaluation. On a 5-point Likert scale, median coverage scores were 5 (IQR 4-5) for both models; the distributions did not differ significantly (U=733.0; n_1=37; n_2=39; P=.89). Median coherence scores were 5 (IQR 5-5 for GPT-4; IQR 4-5 for Nous-Hermes-2-7B-DPO) for both; they did not differ significantly (U=670.0; n_1=37; n_2=39; P=.49). On a 3-point Likert scale, the median relevance scores were 3 (IQR 3-3) for both; they did not differ significantly (U=662.0; n_1=37; n_2=39; P=.15). Median length scores were 3 (IQR 2-3) for both; they did not differ significantly (U=672.0; n_1=37; n_2=39; P=.55). On a binary Likert scale, median hallucination scores were 0 (IQR 0-0) for both; they did not differ significantly (U=859.0; n_1=37; n_2=39; P=.10). The median CLIs were 16.635 (IQR 13.860-17.675) for GPT-4 and 12.125 (IQR 11.02-13.98) for Nous-Hermes-2-7B-DPO; there was a statistically significant difference (U=307.5; n_1=20; n_2=16; P<.001).

Median token counts for queries posed to GPT-4 and Nous-Hermes-2-7B-DPO were 5 (IQR 5-7) and 7 (IQR 5-8), respectively; there was no significant difference (U=165.0; n_1=20; n_2=16; P=.66). Median lengths of responses generated by GPT-4 and Nous-Hermes-2-7B-DPO were 1118 (IQR 709-2986) and 441 (IQR 231-695) for the combined individual summaries and 141.5 (IQR 115-159) and 61 (IQR 28-87) for the final summaries, respectively. Both were significantly different (U=300.0; n_1=20; n_2=16; P<.001, and U=145.5; n_1=20; n_2=16; P<.001).

Figure 2. Box plots illustrating the distribution of scores for the evaluation criteria used. (A) Coverage on a 5-point Likert scale. (B) Coherence on a 5-point Likert scale. (C) Relevance on a 3-point Likert scale. (D) Length on a 3-point Likert scale. (E) Hallucination on a binary scale. (F) Values for the Coleman-Liau Index. (G) Token counts for Questions. (H) Token counts for combined individual summaries. (I) Token counts for the final summary.



Discussion

Principal Findings

This study presents a novel two-layer RAG framework for MQA that uses user-generated content from Reddit. Our findings demonstrate that the framework effectively synthesizes accurate and contextually relevant answers even in low-resource settings, aligning with our goal to create an accessible, computationally lightweight tool. Focusing on small, quantized, open-source LLMs ensures equitable access to valuable insights about emerging trends, potential side effects, and general perception of substances, as reflected in Reddit posts.

The modular structure of the framework enables good performance without requiring specialized hardware, which is critical in low-resource environments. This modularity also supports using different retrieval engines or LLMs, providing flexibility to adapt to various use cases. The system's ability to answer nuanced queries (eg "What are k cramps like?"—which would require extensive manual curation) illustrates its potential for real-world applications. The framework's ability to specify temporal ranges in queries allows it to track trends over time, offering opportunities for longitudinal studies and misinformation detection.

Unlike previous work [15], where segments of text are generated chronologically, we performed segmentation at the post level without accounting for chronology. Compared with existing literature, which predominantly relies on large LLMs for MQA [4,16], this study underscores the potential of smaller models for tasks requiring domain-specific, contextually accurate outputs. Prior work often focuses on high-resource settings [17,18] with robust computational infrastructure, leaving gaps in applicability for low-resource environments. We fill this gap by showing that reliable performance can be achieved with computationally efficient architectures, expanding the reach of artificial intelligence tools to underresourced regions. Although smaller LLMs have been used [19], summarization from large volumes of text with aggregated information has not been addressed before.

Limitations

Despite its promising features, the framework has limitations. It relies on the accuracy and representativeness of the Reddit data it ingests. Reddit posts may include biases, inaccuracies, or misinformation that could influence the system's output. While faithfully summarizing misinformation is valuable for transparency, users need to exercise caution in interpreting the results. Additionally, we evaluated the framework using a small set of queries pertaining to substance use; further validation is necessary to assess performance across diverse medical domains.

Conclusions

This study demonstrates that a modular, lightweight RAG framework can effectively address complex MQA using social media data in low-resource settings. By enabling clinicians to rapidly extract insights about substance use trends and potential side effects from Reddit posts, the framework holds significant potential for improving public health.

Acknowledgments

Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health (NIH) under award R01DA057599. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Data Availability

All data used in this study were publicly available from Reddit at the time of data collection. The second-level summaries are available in Multimedia Appendix 3. The original posts and social media posts analyzed during this study are available from the corresponding author upon reasonable request and the completion of a data use agreement.

Authors' Contributions

SD and Y Ge led analysis, evaluation, visualization, and original draft preparation. Y Guo, SR, JH, J Powell, DW, SP, and SL contributed to the evaluation and manuscript preparation. SB, MR, RS, YX, SK, RC, NH, DM, RW, JL, A Spadaro, and J Perrone worked on the evaluation. A Sarker conceptualized the study, led model design and implementation, and supervised the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prompts used.
[DOCX File , 12 KB - jmir_v27i1e66220_app1.docx]

Multimedia Appendix 2
Sample first-layer individual summaries.
[DOCX File , 14 KB - jmir_v27i1e66220_app2.docx]

Multimedia Appendix 3
Final summaries generated by the framework for each of the 20 queries used for evaluation.
[DOCX File , 30 KB - jmir_v27i1e66220_app3.docx]

References

1. Ji Z, Yu T, Xu Y, Lee N, Ishii E, Fung P. Towards mitigating LLM hallucination via self reflection. 2023 Presented at: Findings of the Association for Computational Linguistics: EMNLP 2023; December 6-10, 2023; Singapore, Singapore p. 1827-1843 URL: https://aclanthology.org/2023.findings-emnlp.123/ [doi: 10.18653/v1/2023.findings-emnlp.123]
2. Dhuliawala S, Komeili M, Xu J, Raileanu R, Li X, Celikyilmaz A, et al. Chain-of-verification reduces hallucination in large language models. arXiv Preprint posted online on September 20, 2023 [FREE Full text] [doi: 10.18653/v1/2024.findings-acl.212]
3. Ge J, Sun S, Owens J, Galvez V, Gologorskaya O, Lai JC, et al. Development of a liver disease-specific large language model chat interface using retrieval augmented generation. medRxiv Preprint posted online on November 11, 2023 [FREE Full text] [doi: 10.1101/2023.11.10.23298364] [Medline: 37986764]

4. Xiong G, Jin Q, Lu Z, Zhang A. Benchmarking retrieval-augmented generation for medicine. arXiv Preprint posted online on February 20, 2024 [FREE Full text] [doi: 10.18653/v1/2024.findings-acl.372]
5. McNamara SL, Yi PH, Lotter W. The clinician-AI interface: intended use and explainability in FDA-cleared AI devices for medical image interpretation. NPJ Digital Med 2024;7(1):80 [FREE Full text] [doi: 10.1038/s41746-024-01080-1] [Medline: 38531952]
6. Ghosh S, Tyagi U, Kumar S, Manoch D. BioAug: conditional generation based data augmentation for low-resource biomedical NER. 2023 Presented at: SIGIR '23: The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 27, 2023; Taipei, Taiwan. [doi: 10.1145/3539618.3591957]
7. Somani S, Balla S, Peng AW, Dudum R, Jain S, Nasir K, et al. Contemporary attitudes and beliefs on coronary artery calcium from social media using artificial intelligence. NPJ Digital Med 2024;7(1):83 [FREE Full text] [doi: 10.1038/s41746-024-01077-w] [Medline: 38555387]
8. Teknium, theemozilla, karan4d, huemin_art. Nous-Hermes-2-Mistral-7B-DPO. Hugging Face. URL: https://huggingface.co/NousResearch/Nous-Hermes-2-Mistral-7B-DPO [accessed 2024-07-10]
9. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Leoni Aleman F, et al. GPT-4 technical report. arXiv Preprint posted online on March 15, 2023 [FREE Full text] [doi: 10.48550/arXiv.2303.08774]
10. Chaput M. Whoosh 2.7.4 documentation. Whoosh. URL: https://whoosh.readthedocs.io/en/latest/ [accessed 2024-07-10]
11. Lin CY. ROUGE: a package for automatic evaluation of summaries. 2004 Presented at: Text Summarization Branches Out; July 25-26, 2004; Barcelona, Spain p. 74-81 URL: https://aclanthology.org/W04-1013.pdf
12. Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: a method for automatic evaluation of machine translation. 2024 Presented at: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics; July 7-12, 2002; Philadelphia, PA URL: https://aclanthology.org/P02-1040.pdf [doi: 10.3115/1073083.1073135]
13. Coleman M, Liau TL. A computer readability formula designed for machine scoring. J Appl Psychol 1975;60(2):283-284. [doi: 10.1037/h0076540]
14. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
15. Subbiah M, Zhang SS, Chilton L, Zhang B, McKeown K. Reading subtext: evaluating large language models on short story summarization with writers. arXiv Preprint posted online on March 02, 2024 [FREE Full text] [doi: 10.1162/tacl_a_00702]
16. Hammane Z, Ben-Bouazza FE, Fennan A. SelfRewardRAG: enhancing medical reasoning with retrieval-augmented generation and self-evaluation in large language models. 2024 Presented at: International Conference on Intelligent Systems and Computer Vision (ISCV); August 12, 2024; Fez, Morocco URL: https://ieeexplore.ieee.org/abstract/document/10620139/ [doi: 10.1109/iscv60512.2024.10620139]
17. Alkhalaf M, Yu P, Yin M, Deng C. Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records. J Biomed Inform 2024;156:104662 [FREE Full text] [doi: 10.1016/j.jbi.2024.104662] [Medline: 38880236]
18. Soong D, Sridhar S, Si H, Wagner J, Sá ACC, Yu CY, et al. Improving accuracy of GPT-3/4 results on biomedical data using a retrieval-augmented language model. PLOS Digital Health 2024;3(8):e0000568. [doi: 10.1371/journal.pdig.0000568] [Medline: 39167594]
19. Jeong M, Sohn J, Sung M, Kang J. Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models. Bioinformatics 2024;40(Suppl 1):i119-i129 [FREE Full text] [doi: 10.1093/bioinformatics/btae238] [Medline: 38940167]

Abbreviations

BLEU: Bilingual Evaluation Understudy
CLI: Coleman-Liau Index
LLM: large language model
MQA: medical question answering
RAG: retrieval-augmented generation
ROUGE: Recall-Oriented Understudy for Gisting Evaluation

Edited by A Mavragani; submitted 06.09.24; peer-reviewed by G Xiong, M Jeong; comments to author 26.09.24; revised version received 30.10.24; accepted 05.12.24; published 06.01.25.

Please cite as:
Das S, Ge Y, Guo Y, Rajwal S, Hairston J, Powell J, Walker D, Peddireddy S, Lakamana S, Bozkurt S, Reyna M, Sameni R, Xiao Y, Kim S, Chandler R, Hernandez N, Mowery D, Wightman R, Love J, Spadaro A, Perrone J, Sarker A
Two-Layer Retrieval-Augmented Generation Framework for Low-Resource Medical Question Answering Using Reddit Data: Proof-of-Concept Study
J Med Internet Res 2025;27:e66220
URL: https://www.jmir.org/2025/1/e66220
doi:10.2196/66220
PMID:

Review

Bias Mitigation in Primary Health Care Artificial Intelligence Models: Scoping Review

Maxime Sasseville[1,2], PhD; Steven Ouellet[1], PhD; Caroline Rhéaume[2,3,4], MD, PhD; Malek Sahlia[5], MSc, MEng; Vincent Couture[1], PhD; Philippe Després[6], PhD; Jean-Sébastien Paquette[2,3], MD; David Darmon[7], PhD; Frédéric Bergeron[8], MSI; Marie-Pierre Gagnon[1,2], PhD

[1]Faculté des sciences infirmières, Université Laval, Québec, QC, Canada

[2]Vitam Research Center on Sustainable Health, Québec, QC, Canada

[3]Département de médecine familiale et de médecine d'urgence de la Faculté de médecine, Université Laval, Québec, QC, Canada

[4]Research Center of Quebec Heart and Lungs Institute, Québec, QC, Canada

[5]École Nationale des Sciences de l'Informatique, Université de La Manouba, La Manouba, Tunisia

[6]Département de physique, de génie physique et d'optique de la Faculté des sciences et de génie, Université Laval, Québec, QC, Canada

[7]Risques, Epidémiologie, Territoires, Informations, Education et Santé. Département d'enseignement et de recherche en médecine générale, Université Côte d'Azur, Nice, France

[8]Direction des services-conseils de la Bibliothèque, Université Laval, Québec, QC, Canada

Corresponding Author:
Maxime Sasseville, PhD
Faculté des sciences infirmières, Université Laval
1050 Av. de la Médecine
Québec, QC, G1V 0A6
Canada
Phone: 1 418 656 3356
Email: maxime.sasseville@fsi.ulaval.ca

Abstract

Background: Artificial intelligence (AI) predictive models in primary health care have the potential to enhance population health by rapidly and accurately identifying individuals who should receive care and health services. However, these models also carry the risk of perpetuating or amplifying existing biases toward diverse groups. We identified a gap in the current understanding of strategies used to assess and mitigate bias in primary health care algorithms related to individuals' personal or protected attributes.

Objective: This study aimed to describe the attempts, strategies, and methods used to mitigate bias in AI models within primary health care, to identify the diverse groups or protected attributes considered, and to evaluate the results of these approaches on both bias reduction and AI model performance.

Methods: We conducted a scoping review following Joanna Briggs Institute (JBI) guidelines, searching Medline (Ovid), CINAHL (EBSCO), PsycINFO (Ovid), and Web of Science databases for studies published between January 1, 2017, and November 15, 2022. Pairs of reviewers independently screened titles and abstracts, applied selection criteria, and performed full-text screening. Discrepancies regarding study inclusion were resolved by consensus. Following reporting standards for AI in health care, we extracted data on study objectives, model features, targeted diverse groups, mitigation strategies used, and results. Using the mixed methods appraisal tool, we appraised the quality of the studies.

Results: After removing 585 duplicates, we screened 1018 titles and abstracts. From the remaining 189 full-text articles, we included 17 studies. The most frequently investigated protected attributes were race (or ethnicity), examined in 12 of the 17 studies, and sex (often identified as gender), typically classified as "male versus female" in 10 of the studies. We categorized bias mitigation approaches into four clusters: (1) modifying existing AI models or datasets, (2) sourcing data from electronic health records, (3) developing tools with a "human-in-the-loop" approach, and (4) identifying ethical principles for informed decision-making. Algorithmic preprocessing methods, such as relabeling and reweighing data, along with natural language processing techniques that extract data from unstructured notes, showed the greatest potential for bias mitigation. Other methods aimed at enhancing model fairness included group recalibration and the application of the equalized odds metric. However, these approaches sometimes exacerbated prediction errors across groups or led to overall model miscalibrations.

Conclusions: The results suggest that biases toward diverse groups are more easily mitigated when data are open-sourced, multiple stakeholders are engaged, and during the algorithm's preprocessing stage. Further empirical studies that include a broader range of groups, such as Indigenous peoples in Canada, are needed to validate and expand upon these findings.

Trial Registration: OSF Registry osf.io/9ngz5/; https://osf.io/9ngz5/

International Registered Report Identifier (IRRID): RR2-10.2196/46684

(J Med Internet Res 2025;27:e60269) doi:10.2196/60269

KEYWORDS

artificial intelligence; AI; algorithms; expert system; decision support; bias; community health services; primary health care; health disparities; social equity; scoping review

Introduction

Developments in computer science have led to artificial intelligence (AI) models that learn from large datasets and can perform independent analysis [1-4]. Significant progress has been made in these tasks with the development of machine learning (ML). This branch of AI focuses on understanding, generating, and reasoning based on data without explicit human instructions [2,3] Such ML algorithms use datasets known as "training datasets" to capture the patterns required for clustering tasks or predictive modeling [3,4]. These models are now used in multiple contexts and industries to predict the likelihood of an event or to support human decision-making [4]. In health care, AI models applied in radiology can potentially detect and predict the progression of cancerous tumors accurately [5]. Algorithms can also be useful in community-based primary health care (CBPHC) for identifying individuals, such as heart failure or diabetes outpatients, who require specific health care services [6]. As defined by the Canadian Institutes of Health Research, CBPHC encompasses a comprehensive array of services aimed at community well-being, including primary prevention (such as public health), health promotion, disease prevention, diagnosis, treatment, and management of chronic and episodic illnesses, rehabilitation support, and end-of-life care [7].

Despite the potential benefits of AI, such as compensating for workforce shortage and maximizing access to CBPHC [6], algorithm biases toward diverse groups can hinder their application in health care settings. These biases may be perpetuated when protected attributes [1], as identified by the place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, and social capital (PROGRESS-Plus) framework [8], are underrepresented or misrepresented in the training data of algorithms [1,9]. Strategies aimed at identifying and mitigating bias, defined as a persistent inclination either in favor or toward something [9], in predictive models are in development and beginning to be empirically applied [10,11]. In computer science, attempts to achieve algorithmic fairness can involve which are (1) preprocessing, (2) in-processing, or even, (3) postprocessing strategies, such as those used in "out-of-the-box" commercial AI models [4]. Academic disciplines beyond computer science, such as medicine, management, and ethics, are also closely involved in addressing issues related to identifying potential bias toward diverse groups in AI models [1,3]. However, there remains a knowledge gap regarding which strategies and methods have been empirically applied to mitigate bias toward diverse groups in CBPHC algorithms [10,12].

To address this gap, we conducted a scoping review aimed at identifying and describing (1) the attempts made to mitigate bias in primary health care AI models, (2) which diverse groups or protected attributes have been considered, and (3) the results regarding bias attenuation and the overall performance of the models.

Methods

Search Strategy

We conducted a scoping review informed by the Joanna Briggs Institute (JBI) [13] and used the Population (or Participant), Concept, and Context Framework [14], as shown in Table 1.

Table 1. Population (or Participant), Concept, and Context framework used for the search strategy.

PCC[a] elements [14]	Definition (per JBI[b] Reviewer's Manual)	PCC elements applied in this review
Population	"Important characteristics of participants, including age and other qualifying criteria" (11.2.4)	Any diverse groups [8] based on their personal or protected attributes [1].
Concept	"The core concept examined by the scoping review should be clearly articulated to guide the scope and breadth of the inquiry. This may include details that pertain to elements that would be detailed in a standard systematic review, such as the "interventions" or "phenomena of interest" (11.2.4)	Strategies, attempts, or methods for assessing and mitigating bias in artificial intelligence.
Context	"May include...cultural factors such as geographic location or specific racial or gender-based interests. In some cases, context may also encompass details about the specific setting."	Community-based primary health care [7].

[a] PCC (Population [or Participant], Concept, and Context) framework [14].

[b] JBI: Joanna Briggs Institute.

Bias Mitigation in Primary Health Care Artificial Intelligence Models

Primary review questions are (1) What attempts have been made to mitigate bias in primary health care AI models? (2) Which diverse groups or protected attributes have been considered? and (3) What are the results regarding bias attenuation and model performance?

In November 2022, we developed a search strategy aligned with the main concepts of our primary review questions with an experienced librarian in 4 relevant databases (MEDLINE [Ovid], CINAHL [EBSCO], PsycInfo [Ovid], and Web of Science). The results of the search strategy in Web of Science were limited to the following 2 indexes: Science Citation Index Expanded and Emerging Sources Citation Index. We used 5 relevant articles to test the sensitivity of our search strategy, focusing on peer-reviewed publications from the past 5 years (between January 1, 2017, and November 15, 2022). The search strategies for each database can be found in Multimedia Appendix 1.

Data Collection

We imported all sources (n=1603) into the web-based collaborative tool Covidence (Veritas Health Innovation) [15], which automatically identified and removed 581 duplicates, with an additional 4 removed manually. The inclusion and exclusion criteria are presented in Table 2. During the title and abstract screening phase, 7 reviewers independently assessed the abstracts based on the selection criteria. We piloted the screening process on 50 sources that all reviewers independently assessed. Reviewers included a source if it met our inclusion criteria, such as featuring an AI predictive model in health, targeting primary health care populations, and presenting a strategy or method for reducing bias. All titles and abstracts were screened independently by at least 2 reviewers, with any discrepancies resolved through consensus involving all reviewers, including at least 1 senior researcher.

Table 2. Inclusion and exclusion criteria.

PCC (Population, Concept, and Context) elements [14]	Inclusion criteria	Exclusion criteria
Population	• Any populations targeted by CBPHC[a] interventions.	• Any populations targeted by hospital or specialized care interventions.
Concept	• All methods or strategies deployed to assess and mitigate bias toward diverse groups or protected attributes in AI models. • All mitigation methods or strategies deployed to promote and increase equity, diversity, and inclusion in CBPHC algorithms.	• Methods or strategies deployed to assess and mitigate bias in the AI model itself (eg, biased prediction of treatment effects), rather than bias related to individuals' characteristics or protected attributes. • Strategies, methods, or interventions that are not related to CBPHC. • CBPHC interventions that do not include any algorithm or AI system.
Context	• Include all CBPHC algorithms (AI) applications that can perpetuate or introduce potential biases toward diverse groups based on their characteristics or protected attributes.	• Algorithms used by primary health care providers for support in administrative tasks and operational aspects, rather than for clinical decisions.
Study design, study type, and time frame	• All empirical studies published in English or French between 2017 and 2022.	• Reviews, opinions, commentaries, editorial content, conference papers, communications, protocols, magazine articles, and so on.

[a] CBPHC: Community-based primary health care.

For the remaining articles assessed for eligibility at the full-text review stage, we searched for and obtained any missing full texts of selected references, then imported them into Covidence. Out of 5 reviewers independently applied the same selection criteria, and all reasons for exclusion were recorded in Covidence. All full texts underwent dual screening. As in the previous stage, any discrepancies regarding the included studies were resolved through consensus among all reviewers, including at least one senior researcher.

Data Extraction

One experienced reviewer performed the extraction of the included studies, and 2 senior researchers validated the data for all of them. We also hand-searched [16] and identified 2 relevant articles [17,18] related to 2 included studies [19,20], which were added to Covidence for extraction. Based on reporting standards for AI in health care [21], we extracted the following information (title of the paper, year of publication, lead author, and country), study objective, discipline and study design, AI model features, study population and setting, AI model architecture and evaluation, bias assessment method, strategy for deployment, diverse groups concerned, bias mitigation results, and the impact on AI model performance and accuracy.

Quality Assessment

One senior reviewer appraised the quality of the included studies by applying the Mixed-Methods Appraisal Tool (MMAT) [22,23] and at least one senior researcher validated each of them.

Data Synthesis

In accordance with the JBI recommendations [24], we synthesized data using structured narrative summaries around our review concepts (eg, model data source, model input, model output, diverse groups, or protected attributes), mitigation strategies deployed, and the results on bias mitigation and overall model performance. We reported our findings based on the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [25].

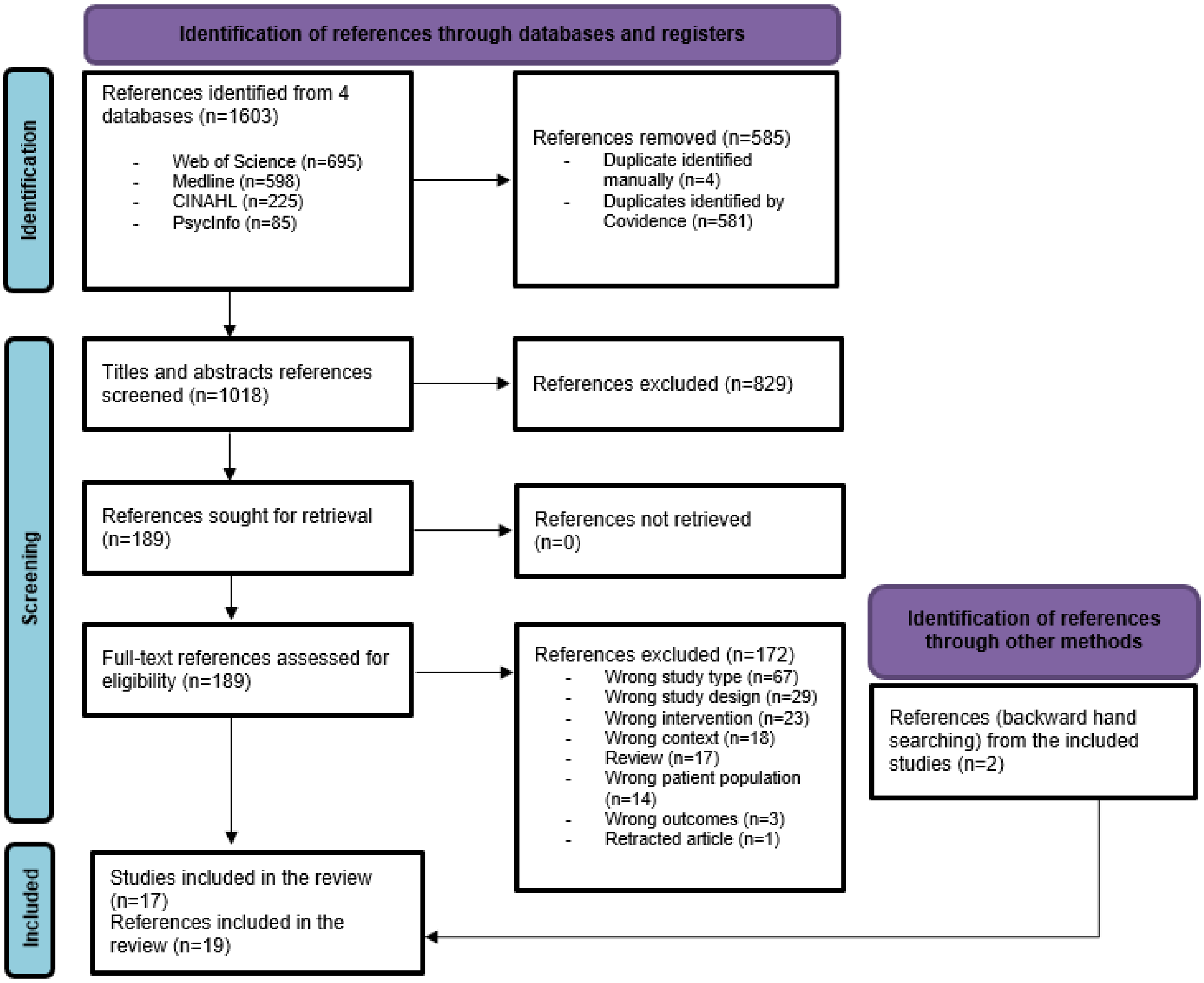
Ethical Considerations

We obtained approval from the ethics board of the “Comité d’éthique de la recherche sectoriel en Santé des Populations et Première Ligne du Centre Intégré Universitaire de Santé et de Services Sociaux de la Capitale-Nationale” for the Protecting and Engaging Vulnerable Populations in the Development of Predictive Models in Primary Health Care for Inclusive, Diverse and Equitable AI (PREMIA) project (#2023-2726).

Results

Out of a total of 1018 titles and abstracts, along with 189 full-text articles that underwent dual screening, 17 studies [19,20,26-40] met our eligibility criteria. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram is shown in Figure 1 [41].

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



The relatively high number of exclusions at the full-text review stage (172/189, 91%) can be attributed to our inclusive approach in the previous stage. For example, some reviews (17/189, 9%) and incorrect study types (67/189, 35%), such as editorials, commentaries, or conference papers, were excluded at this stage. Other exclusion reasons (88/189, 47%) included models that lacked AI components, models focusing on health care operational processes (eg, workflow modeling), studies targeting populations receiving specialized care (eg, hospitalized or cancer patients), interventions such as imaging research that were outside the scope of CBPHC, and methods for mitigating bias that were applied to the AI model itself (eg, biased predictions of treatment effects) rather than addressing biases related to diverse groups or personal attributes.

Overview of Included Studies

Of the 17 included studies published between 2019 and 2022, we identified 7 studies in the discipline of data science or informatics, 7 in medical informatics, 1 in medical ethics and informatics, 1 in medical ethics using a Delphi method, and 1 in management care ethics using a user-centered design. Most studies have been conducted in the United States (15/17, 88%), 1 in the United Kingdom, and 1 in Italy. The main characteristics of the included studies can be found in Multimedia Appendix 2.

Quality Assessment of Included Studies

Most studies had a quantitative descriptive study design (14/17, 82%), while 2 used a mixed methods design, and 1 used a qualitative design. All studies showed high quality, receiving scores of 3 or 4 stars (on a possibility of 5). All MMAT scores can be found in Multimedia Appendix 3.

Diverse Groups Considered

The most frequently studied protected attributes were race (or ethnicity), examined in 71% (12/17) of studies, and sex (defined as binary male versus female), considered in 59% (10/17) of studies. None of the studies distinguished between biological sex and socially constructed gender, and 5 of them incorrectly identified sex as gender. Race or ethnicity was most often categorized as White or Black, Black or non-Black or, in one study, as Asian, Black, White, and other.

Other protected attributes considered by the studies included age (7/17, 41%), socioeconomic status or its proxies, such as income, work class, education, health care insurance (5/17, 29%), place of residence (2/17, 12%), marital status (1/17, 6%), and disability status (1/17, 6%).

Categorization of Deployed Bias Mitigation Strategies

We identified considerable heterogeneity across the studies, which used various strategies and methods to assess and mitigate bias in algorithms impacting diverse groups. We categorized these efforts into four groups: (1) addressing bias in existing AI models or datasets, (2) mitigating biases from data sources such as electronic health records (EHRs), (3) developing tools that incorporate a "human-in-the-loop" approach, and (4) identifying ethical principles to guide informed decision-making.

Attempts in Existing AI Models or Datasets

We identified 7 studies that attempted to mitigate biases in existing AI models or datasets [19,20,27,28,35,37,39].

A debiasing attempt was made on an insurance coverage algorithm designed to identify individuals who could benefit from health resources according to their health needs [35]. Risk scores were initially calculated based on projected future costs rather than uncontrolled or unmanaged illnesses, disadvantaging Black patients. By changing the data labeling to focus on future illness rather than future costs, the percentage of Black patients who could benefit from health resources increased significantly [35].

Another cohort study [37] using a Medicaid enrollees' dataset showed that reweighing was more effective at reducing bias in postpartum depression risk scores between White and Black individuals compared with training without the race variable for comparison. Initially, it was found that the White individuals had higher rates of postpartum depression and mental health service use. However, after comparing postpartum depression rates between races based on population surveys, it became clear that the higher rates in White women might be due to disparities in the timely assessment, screening, and detection of symptoms in Black women [37].

A total of three other studies include (1) retraining models with data that incorporated health equity measures resulted in a slight decrease in performance for detecting abnormal electrocardiograms but significantly reduced gender, race and age biases [19]; (2) increasing diversity in the training data of a predictive pulmonary disease model improved its performance [27]; and (3) although a mental health assessment model achieved high accuracy, its performance was statistically higher and more accurate for men than for women [18]. The use of an algorithmic disparate remover, by adjusting the modeling data, significantly reduced this disparity while maintaining high accuracy [20].

Another attempt to assess bias involved replicating models predicting liver disease [39]. Importing an existing dataset reproduced predictive models with high accuracy but revealed a previously unobserved bias, with women experiencing a higher false negative rate.

We identified only 1 in-processing debiasing attempt [28]. Out of 2 algorithmic fairness strategies, group recalibration and equalized odds, were used to recalibrate a predictive model of cardiovascular diseases that was not initially adjusted for attributes such as sex or race. This resulted in an exacerbation of false positive and negative rates differences between groups, as well as overall model miscalibration.

Attempts in Data Sourcing

We identified 5 studies that attempted to mitigate biases in data sourcing [26,31,32,38,40].

Based on published synthetic datasets, such as the analysis of the American Time Use Survey dataset, using fairness metrics revealed potential discrepancies in representativeness between real and synthetic data across age, sex, and race [26].

Out of 4 other studies investigated EHRs datasets [31,32,38,40]. A natural language processing model was developed to extract vital sign features from unstructured notes, comparing risk scores with 2 convenience samples. This method reduced the missingness of vital signs by 31%, thereby mitigating possible discrimination toward diverse groups, such as Black men or Black women [32]. Based on data from a previous study, 2 ML models were trained to compare balanced error rates across different socioeconomic status levels and the incompleteness of EHRs data [31]. Asthmatic children with lower socioeconomic status exhibited larger balanced error rates than those with higher socioeconomic status and had more missing information regarding asthma care, severity, or undiagnosed asthma, despite meeting asthma criteria [31].

Potential bias based on place of residence in EHRs was examined by 2 studies [38,40]. Rebalancing class labels by adding zip-code level information to 19,367 EHRs during the preprocessing step showed no significant deviation in performance, indicating that bias can be mitigated through preprocessing [38]. Meanwhile, a simple 30-day readmission prediction model was developed, categorizing each patient as local (nearby) or not (far) [40]. The performance with and without this variable was assessed, revealing no significant differences. Considering that living locally only affects the observability of the outcome (eg, a patient may be readmitted to a different hospital), differential bias assessment cannot rely solely on observed data [40].

Attempts in Developing Tools With a "Human-in-the-Loop" Approach

We identified 3 studies that attempted to mitigate biases by incorporating a "human-in-the-loop" approach [29,30,36].

These studies led to the development of "human-in-the-loop" tools: (1) a visual tool for auditing and mitigating bias from tabular datasets, which was tested through experiments on 3 datasets with user participation and significantly reduced bias compared with another commercial debiasing toolkit [29]; (2) pragmatic tools developed for better use of risk scores with a Medicare members' dataset, allowing users to identify appropriate risk scores for each subgroup to achieve equality of opportunity [30]; and (3) a tool called "FairLens" capable of identifying and explaining biases, which was tested using a fictitious black box model serving as a decision support system [36]. Empirically validated by injecting biases into this fictitious decision support system, this tool outperformed other standard measures and enabled experts to identify problematic groups or affected patients, thereby allowing for the detection of potential misclassification [36].

Attempts at Identifying Ethical Principles for Informed Decision-Making

We identified 2 empirical studies that attempted to mitigate biases by identifying ethical principles for informed decision-making [33,34].

To assess the potential missingness of EHR data from phenotyping technology, a Delphi study was conducted to address ethical challenges and reach a consensus on the importance of privacy, transparency, consent, accountability, and fairness [33]. In addition, a user-centered design study was conducted to identify user requirements, mainly intended for health managers and clinicians, to support informed decision-making and confidence in using a hepatitis C severity illness predictive model prototype [34].

Discussion

Principal Findings

The reviewed studies illustrate a multifaceted approach to mitigating bias in primary care AI models. Strategies include retraining, reweighing, relabeling, adding more diversity, and attempting to replicate existing modeling data [19,20,27,35,37,39], as well as algorithmic recalibration applied to an existing prediction model [28]. Other strategies involve the development and application of fairness metrics to ensure equitable distributions in previously published databases [26], and the identification of missingness in EHRs datasets by rebalancing class labels or adding information [31,32,38]. Another group of strategies includes the introduction of visual interactive tools for human-in-the-loop bias auditing [29,30,36]. All these attempts cover a broad spectrum of interventions, ranging from data preprocessing and algorithmic modification to post hoc analysis, demonstrating the complexity and variety of approaches needed to address bias in AI models in primary health care.

The studies collectively address a wide range of protected attributes [1,8], including race or ethnicity [19,26,28-37], sex [19,20,26-31,36,39], age [19,26,27,29-31,36], socioeconomic status (SES) [27,29,31,33,36], and other demographic variables such as place of residence [38,40]. This underlines the recognition of the multifaceted nature of bias, which can intersect across various dimensions of identity and social determinants of health [9,42]. However, we have identified disparities in the number of protected attributes studied. Race (White vs Black) and sex (male vs female) are most frequently investigated, whereas other attributes, such as disability and gender, are underresearched or not studied at all.

Bias mitigation efforts reveal a nuanced landscape where attempts to reduce bias across protected attributes can result in complex trade-offs with model performance. For example, a decrease in overall model performance accompanied by significant reductions in bias was observed following the implementation of constrained optimization [19]. Similarly, improvements in calibration for specific groups came at the cost of increased disparities in false positive and false negative rates between groups [28]. Despite these trade-offs, the efforts have largely been successful in reducing bias, as evidenced by a study that achieved fairer distributions in synthetic data [26], and in another study where human-in-the-loop interventions significantly reduced bias while maintaining utility [29].

These empirical findings reinforce theoretical insights that emphasize the importance of health equity between protected and unprotected attributes [1,8]. To mitigate bias in AI health models, distributive justice options for ML have been proposed: (1) equal patient outcomes; (2) equal performance; and (3) equal allocation of resources [1]. Since these different types of fairness options are often incompatible, optimizing all these parameters seems challenging, as demonstrated by an identified study [28]. Trade-offs are essential, and a participatory process involving key stakeholders, including ethicists, clinicians, and marginalized populations, is strongly encouraged [1]. While striving to create ethically robust AI models, selected studies often reveal tension, as efforts to reduce bias can sometimes lead to a decrease in the model's overall performance. This presents a critical challenge: balancing the imperative of fairness with the need to maintain high accuracy and efficiency in algorithmic outputs.

Comparison With Previous Work

Initiatives focused on the fair use of AI in health care and the assessment of bias risk in AI predictive models have been published in recent years. Notable initiatives include Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) and Standard Protocol Items Recommendations for Interventional Trials-Artificial Intelligence (SPIRIT-AI) [43], which provide guidelines for the ethical presentation of the results of trials conducted with AI in the health care field. To assess the risk of bias in diagnostic and prognostic prediction model studies, the "Prediction Model Risk of Bias Assessment Tool" (PROBAST) [44] can be used. PROBAST consists of a list of signaling questions grouped into 4 categories: participants, predictors, outcomes, and analysis. This tool was used in a systematic scoping review to assess the quality of primary studies reporting applications of AI in CBPHC [45].

However, the objective of our scoping review differs; it is not to identify biases in the AI prediction models themselves, but rather to examine biases toward groups that are underrepresented or misrepresented in these AI models. An identified review has used and adapted PROBAST to assess related protected attributes, but the AI predictive models studied were hospital-based and not relevant to primary care [11]. We also identified a scoping review protocol that focused on bias toward diverse groups in AI systems in primary care; however, unless we are mistaken, the results of this protocol have never been published [10]. Another identified review aimed to assess age-related bias in AI but did not focus on primary health care [46]. Finally, we identified another systematic review investigating health inequities in primary care, but it adopted a system-wide perspective, focusing on aspects such as patient consultation and effects on health systems [47].

To our knowledge, no other published review has the objectives of identifying (1) the bias mitigation strategies or methods in primary health care, (2) the diverse groups that are underrepresented or misrepresented, and (3) the results of bias mitigation and AI model performance.

Strengths and Limitations

The strengths of this review include results that can be translated into recommendations for various stakeholders, such as AI developers, researchers, and decision makers. However, we acknowledge some limitations. First, we limited our search strategy to the last 5 years before November 2022 and focused on 4 databases, which may have excluded some relevant studies. Second, the extraction of studies and quality assessment were conducted only once, although all of them were validated by at least one senior researcher. Third, due to the heterogeneity of the studies, we were unable to combine results through a quantitative synthesis and remained at a narrative level of reporting. Finally, our review primarily identified research from a North American setting, which reduces its transferability to other continents.

Future Directions and Dissemination Plan

This scoping review serves as the initial phase of the iterative project “Protecting and Engaging Vulnerable Populations in the Development of Predictive Models in Primary Health Care for Inclusive, Diverse, and Equitable AI” (PREMIA).

Following the results of this review, we have developed a framework currently validated by a diverse group of experts, including clinicians, public health managers, primary care researchers, data scientists, and patient and citizen partners. This group is concentrating on existing AI predictive models and the bias mitigation strategies identified in our scoping review. Diverse populations, such as older adults, individuals with disabilities, and people from various racial and ethnic backgrounds, are actively involved in this second phase of PREMIA. We plan to prepare and submit a manuscript based on the findings of this Delphi study.

In addition, in recognition of the rapid advancements in this field, we plan to update this literature review in 2027 using a similar search strategy. This iterative approach will allow us to refine our framework and track the progress of bias mitigation in AI models within primary health care. Indigenous peoples in Canada represent a group historically underrepresented in health research, leading to inequities [3]. Since no other study has addressed bias related to Indigenous status, we collaborate with Indigenous representatives to develop methods for mitigating this bias in CBPHC algorithms.

Conclusion

This review identifies strategies and methods for mitigating bias in primary health care algorithms, considers diverse groups based on their personal or protected attributes, and examines the results of bias attenuation and model performance. The findings suggest that biases toward diverse groups can be more effectively mitigated when data are open-sourced, multiple stakeholders are involved, and during the preprocessing stage of algorithm development. More empirical studies are needed, with a focus on including participants who embrace greater diversity, such as nonbinary gender identities or Indigenous peoples in Canada.

Acknowledgments

The Protecting and Engaging Vulnerable Populations in the Development of Predictive Models in Primary Health Care for Inclusive, Diverse, and Equitable AI project is funded by the International Observatory on the Societal Impacts of AI and Digital Technology. The authors would like to thank Karine Gentelet for her contribution to the study’s design and for obtaining the funding.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1-3.

Authors' Contributions

MS, MPG, CR, VC, PD, JSP, and DD designed the study and obtained the funding. MS, MPG, SO, and FB developed the search strategy. MS, SO, MS, CR, MPG, VC, and FB participated in the screening of sources. MS, SO, and MPG conducted the data extraction. SO, MS, and MPG completed the first draft of the manuscript, and all authors participated in the revision and editing of the manuscript versions. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database Search Strategies.
[PDF File (Adobe PDF File), 84 KB - jmir_v27i1e60269_app1.pdf]

Multimedia Appendix 2

Characteristics of Included Studies.
[PDF File (Adobe PDF File), 328 KB - jmir_v27i1e60269_app2.pdf]

Multimedia Appendix 3
Quality assessment: MMAT (Mixed-Methods Appraisal Tool) scores.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v27i1e60269_app3.xlsx]

Multimedia Appendix 4
PRISMA-ScR checklist.
[DOCX File , 109 KB - jmir_v27i1e60269_app4.docx]

References

1. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169(12):866-872 [FREE Full text] [doi: 10.7326/M18-1990] [Medline: 30508424]
2. Qu Y, Wei C, Du P, Che W, Zhang C, Ouyang W, et al. Integration of cognitive tasks into artificial general intelligence test for large models. iScience 2024;27(4):109550 [FREE Full text] [doi: 10.1016/j.isci.2024.109550] [Medline: 38595796]
3. Gurevich E, El Hassan B, El Morr C. Equity within AI systems: what can health leaders expect? Healthc Manage Forum 2023;36(2):119-124 [FREE Full text] [doi: 10.1177/08404704221125368] [Medline: 36226507]
4. Alabdulmohsin I, Lucic M. A near-optimal algorithm for debiasing trained machine learning models. ArXiv. Preprint posted online on August 23, 2022 2022 [FREE Full text]
5. van Leeuwen KG, de Rooij M, Schalekamp S, van Ginneken B, Rutten MJCM. How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr Radiol 2022;52(11):2087-2093 [FREE Full text] [doi: 10.1007/s00247-021-05114-8] [Medline: 34117522]
6. Kang J, Hanif M, Mirza E, Khan MA, Malik M. Machine learning in primary care: potential to improve public health. J Med Eng Technol 2021;45(1):75-80. [doi: 10.1080/03091902.2020.1853839] [Medline: 33283565]
7. Canadian Institutes of Health Research. Community-based primary health care. URL: https://cihr-irsc.gc.ca/e/43626.html [accessed 2024-04-01]
8. Cochrane methods. PROGRESS-Plus. URL: https://methods.cochrane.org/equity/projects/evidence-equity/progress-plus [accessed 2024-04-01]
9. Delgado J, de Manuel A, Parra I, Moyano C, Rueda J, Guersenzvaig A, et al. Bias in algorithms of AI systems developed for COVID-19: a scoping review. J Bioeth Inq 2022;19(3):407-419 [FREE Full text] [doi: 10.1007/s11673-022-10200-z] [Medline: 35857214]
10. Wang JX, Somani S, Chen JH, Murray S, Sarkar U. Health equity in artificial intelligence and primary care research: protocol for a scoping review. JMIR Res Protoc 2021;10(9):e27799 [FREE Full text] [doi: 10.2196/27799] [Medline: 34533458]
11. Wang HE, Landers M, Adams R, Subbaswamy A, Kharrazi H, Gaskin DJ, et al. A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models. J Am Med Inform Assoc 2022;29(8):1323-1333 [FREE Full text] [doi: 10.1093/jamia/ocac065] [Medline: 35579328]
12. Sasseville M, Ouellet S, Rhéaume C, Couture V, Després P, Paquette JS, et al. Risk of bias mitigation for vulnerable and diverse groups in community-based primary health care artificial intelligence models: protocol for a rapid review. JMIR Res Protoc 2023;12:e46684 [FREE Full text] [doi: 10.2196/46684] [Medline: 37358896]
13. Peters MDJ, 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;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
14. Apply PCC. University of South Australia. URL: https://guides.library.unisa.edu.au/ScopingReviews/ApplyPCC [accessed 2024-04-01]
15. Veritas health innovation. Covidence. URL: https://www.covidence.org/ [accessed 2024-04-01]
16. Handsearching. Cochrane. URL: https://training.cochrane.org/resource/tsc-induction-mentoring-training-guide/5-handsearching [accessed 2024-04-01]
17. Reyna M, Sadr N, Gu A, Perez Alday EA, Liu C. Will two do? Varying dimensions in electrocardiography: the physioNet/computing in cardiology challenge 2021 v1.0.3. physionet.org. URL: https://physionet.org/content/challenge-2021/1.0.3/ [accessed 2024-01-11]
18. Singh VK, Long T. Automatic assessment of mental health using phone metadata. Proc. Assoc. Info. Sci. Tech 2019;55(1):450-459. [doi: 10.1002/pra2.2018.14505501049]
19. Perez Alday EA, Rad AB, Reyna MA, Sadr N, Gu A, Li Q, et al. Age, sex and race bias in automated arrhythmia detectors. J Electrocardiol 2022;74:5-9. [doi: 10.1016/j.jelectrocard.2022.07.007] [Medline: 35878534]
20. Park J, Arunachalam R, Silenzio V, Singh VK. Fairness in mobile phone-based mental health assessment algorithms: exploratory study. JMIR Form Res 2022;6(6):e34366 [FREE Full text] [doi: 10.2196/34366] [Medline: 35699997]

21. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc 2020;27(12):2011-2015 [FREE Full text] [doi: 10.1093/jamia/ocaa088] [Medline: 32594179]
22. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. EFI 2018;34(4):285-291. [doi: 10.3233/efi-180221]
23. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. Mixed methods appraisal tool (MMAT) version 2018: user guide. Mixed Methods Appraisal Tool. 2018. URL: https://content.iospress.com/articles/education-for-information/efi180221 [accessed 2024-04-01]
24. Pollock D, Peters MDJ, Khalil H, McInerney P, Alexander L, Tricco AC, de Moraes, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Bhanot K, Qi M, Erickson JS, Guyon I, Bennett KP. The problem of fairness in synthetic healthcare data. Entropy (Basel) 2021;23(9):1165 [FREE Full text] [doi: 10.3390/e23091165] [Medline: 34573790]
27. Fletcher RR, Nakeshimana A, Olubeko O. Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. Front Artif Intell 2020;3:561802 [FREE Full text] [doi: 10.3389/frai.2020.561802] [Medline: 33981989]
28. Foryciarz A, Pfohl SR, Patel B, Shah N. Evaluating algorithmic fairness in the presence of clinical guidelines: the case of atherosclerotic cardiovascular disease risk estimation. BMJ Health Care Inform 2022;29(1):e100460 [FREE Full text] [doi: 10.1136/bmjhci-2021-100460] [Medline: 35396247]
29. D-BIAS: a causality-based human-in-the-loop system for tackling algorithmic bias. IEEE Journals & Magazine | IEEE Xplore. URL: https://ieeexplore.ieee.org/document/9903601/authors#authors [accessed 2024-01-11]
30. Hane CA, Wasserman M. Designing equitable health care outreach programs from machine learning patient risk scores. Med Care Res Rev 2023;80(2):216-227 [FREE Full text] [doi: 10.1177/10775587221098831] [Medline: 35685000]
31. Juhn YJ, Ryu E, Wi C, King KS, Malik M, Romero-Brufau S, et al. Assessing socioeconomic bias in machine learning algorithms in health care: a case study of the HOUSES index. J Am Med Inform Assoc 2022;29(7):1142-1151 [FREE Full text] [doi: 10.1093/jamia/ocac052] [Medline: 35396996]
32. Khurshid S, Reeder C, Harrington LX, Singh P, Sarma G, Friedman SF, et al. Cohort design and natural language processing to reduce bias in electronic health records research. NPJ Digit. Med 2022;5(1):47. [doi: 10.1038/s41746-022-00590-0]
33. Martinez-Martin N, Greely HT, Cho MK. Ethical development of digital phenotyping tools for mental health applications: delphi study. JMIR Mhealth Uhealth 2021;9(7):e27343 [FREE Full text] [doi: 10.2196/27343] [Medline: 34319252]
34. Nong P, Raj M, Platt J. Integrating predictive models into care: facilitating informed decision-making and communicating equity issues. Am J Manag Care 2022;28(1):18-24 [FREE Full text] [doi: 10.37765/ajmc.2022.88812] [Medline: 35049257]
35. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
36. Panigutti C, Perotti A, Panisson A, Bajardi P, Pedreschi D. FairLens: auditing black-box clinical decision support systems. Inf Process Manag 2021;58(5):102657. [doi: 10.1016/j.ipm.2021.102657]
37. Park Y, Hu J, Singh M, Sylla I, Dankwa-Mullan I, Koski E, et al. Comparison of methods to reduce bias from clinical prediction models of postpartum depression. JAMA Netw Open 2021;4(4):e213909 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3909] [Medline: 33856478]
38. Seker E, Talburt JR, Greer ML. Preprocessing to address bias in healthcare data. Stud Health Technol Inform 2022;294:327-331. [doi: 10.3233/SHTI220468] [Medline: 35612086]
39. Straw I, Wu H. Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. BMJ Health Care Inform 2022;29(1):e100457 [FREE Full text] [doi: 10.1136/bmjhci-2021-100457] [Medline: 35470133]
40. Yan M, Pencina MJ, Boulware LE, Goldstein BA. Observability and its impact on differential bias for clinical prediction models. J Am Med Inform Assoc 2022;29(5):937-943 [FREE Full text] [doi: 10.1093/jamia/ocac019] [Medline: 35211742]
41. PRISMA Flow diagram. URL: https://www.prisma-statement.org/prisma-2020-flow-diagram [accessed 2024-04-12]
42. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health 2023;2(6):e0000278 [FREE Full text] [doi: 10.1371/journal.pdig.0000278] [Medline: 37347721]
43. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020;370:m3164 [FREE Full text] [doi: 10.1136/bmj.m3164] [Medline: 32909959]
44. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]

45. Abbasgholizadeh Rahimi S, Légaré F, Sharma G, Archambault P, Zomahoun HTV, Chandavong S, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal. J Med Internet Res 2021;23(9):e29839 [FREE Full text] [doi: 10.2196/29839] [Medline: 34477556]
46. Chu C, Donato-Woodger S, Khan SS, Nyrup R, Leslie K, Lyn A, et al. Age-related bias and artificial intelligence: a scoping review. Humanit Soc Sci Commun 2023;10(1) [FREE Full text] [doi: 10.1057/s41599-023-01999-y]
47. d'Elia A, Gabbay M, Rodgers S, Kierans C, Jones E, Durrani I, et al. Artificial intelligence and health inequities in primary care: a systematic scoping review and framework. Fam Med Community Health 2022;10(Suppl 1):e001670 [FREE Full text] [doi: 10.1136/fmch-2022-001670] [Medline: 36450391]

Abbreviations

AI: artificial intelligence
CBPHC: community-based primary health care
CONSORT-AI: Consolidated Standards of Reporting Trials-Artificial Intelligence
EHR: electronic health record
JBI: Joanna Briggs Institute
ML: machine learning
MMAT: Mixed-Methods Appraisal Tool
PREMIA: Protecting and Engaging Vulnerable Populations in the Development of Predictive Models in Primary Health Care for Inclusive, Diverse and Equitable AI
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
PROBAST: Prediction model Risk Of Bias Assessment Tool
PROGRESS: Place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, and social capital
SES: socioeconomic status
SPIRIT-AI: Standard Protocol Items Recommendations for Interventional Trials-Artificial Intelligence

Edited by A Coristine; submitted 06.05.24; peer-reviewed by J Bensemann, L He; comments to author 14.09.24; revised version received 26.09.24; accepted 07.11.24; published 07.01.25.

Please cite as:
Sasseville M, Ouellet S, Rhéaume C, Sahlia M, Couture V, Després P, Paquette JS, Darmon D, Bergeron F, Gagnon MP
Bias Mitigation in Primary Health Care Artificial Intelligence Models: Scoping Review
J Med Internet Res 2025;27:e60269
URL: https://www.jmir.org/2025/1/e60269
doi:10.2196/60269
PMID:

Original Paper

The Role of the Organization in Promoting Information Security–Related Behavior Among Resident Physicians in Hospitals in Germany: Cross-Sectional Questionnaire Study

Judith Kraushaar[1], Dr rer pol; Sabine Bohnet-Joschko[1], Prof Dr

[1] Chair of Healthcare Management and Innovation, Faculty of Management, Economics and Society, Witten/Herdecke University, Witten, Germany

Corresponding Author:
Judith Kraushaar, Dr rer pol
Chair of Healthcare Management and Innovation
Faculty of Management, Economics and Society
Witten/Herdecke University
Alfred-Herrhausen-Straße 50
Witten, 58455
Germany
Phone: 49 2302 926 47
Email: judith.kraushaar@uni-wh.de

Abstract

Background: Nowadays, optimal patient care should be based on data-driven decisions. In the course of digitization, hospitals, in particular, are becoming complex organizations with an enormously high density of digital information. Ensuring information security is, therefore, essential and has become a major challenge. Researchers have shown that—in addition to technological and regulatory measures—it is also necessary for all employees to follow security policies and consciously use information technology (compliance), because noncompliance can lead to security breaches with far-reaching consequences for the organization. There is little empirical research on information security–related behavior in hospitals and its organizational antecedents.

Objective: This study aimed to explore the impact of specific job demands and resources on resident physicians' information security–related compliance in hospitals through the mediating role of work engagement and information security–related awareness.

Methods: We used a cross-sectional, survey-based study design to collect relevant data from our target population, namely resident physicians in hospitals. For data analysis, we applied structural equation modeling. Our research model consisted of a total of 7 job demands and resources as exogenous variables, 2 mediators, and information security–related compliance as the endogenous variable.

Results: Overall, data from 281 participating physicians were included in the analyses. Both mediators—work engagement and awareness—had a significant positive effect on information security–related compliance (β=.208, *P*=.001 vs β=.552, *P*<.001). Quality of leadership was found to be the only resource with a significant indirect effect on physicians' compliance, mediated by work engagement (β=.086, *P*=.03). Furthermore, awareness mediated the relationships between information security–related communication and information security–related compliance (β=.192, *P*<.001), as well as between further education and training and the endogenous variable (β=.096, *P*=.02). Contrary to our hypothesis, IT resources had a negative effect on compliance, mediated by awareness (β=–.114, *P*=.02).

Conclusions: This study provides new insights into how a high standard of information security compliance among resident physicians could be achieved through strengthening physicians' security work engagement and awareness. Hospital management is required to establish an information security culture that is informative and motivating and that raises awareness. Particular attention should be paid to the quality of leadership, further education and training, as well as clear communication.

(J Med Internet Res 2025;27:e46257) doi:10.2196/46257

KEYWORDS

information security; compliance; work engagement; awareness; leadership; communication; education and training; security; privacy; structural equation modeling; resident; fellow; medical education; continuing education; professional development

Introduction

Information Security–Related Issues in Hospitals

Nowadays, direct, easy, and quick access to and transmission of information are essential factors for health professionals to provide optimal patient care based on data-driven diagnosis and treatment. In the course of digitization, hospitals, in particular, are becoming complex organizations with an enormously high density of such information, which comes from a wide variety of sources and is available in different information systems. In order to make the decision-making process in hospitals more efficient, increasingly more (medical) devices are being digitally connected to hospital networks, providing health care professionals with data from multiple sources at a central access point. In a recently published study, the average number of active medical devices in large university hospitals in Germany was 25,150, with 4500 (17.9%) of them being digitally connected, a rate that is expected to increase rapidly in the near future [1]. Despite their major benefits in terms of the enhanced quality and efficiency of patient care, the use and high variety of connected devices poses potential information security risks to hospitals. Many of the devices used in hospitals are not only connected to local networks but also directly or indirectly connected to the internet or other external systems. For example, patient data can be transitioned to other stakeholders outside the local network or a cloud where they will be further processed [2,3]. Adding to this, with the increasing usage of personal mobile devices for work purposes by health care professionals, the issue becomes even more complex and difficult to handle for hospital management [4-6].

Information security in hospitals refers to the state of full functionality of all information systems, processes, and components and, therefore, the protection of all information required for optimal patient care. It must be guaranteed at all times. Continuous monitoring, as well as rapid responses to breaches and attacks, are essential. Due to the immense impact of poor information security on patient care, as well as the critical standard of security due to years of underinvestment and neglect of this issue, hospitals around the world have recently become targets of cyberattacks, with massive increases each year [7-10]. A study from 2020 shows how vulnerable the German hospital landscape is to ransomware attacks: the authors analyzed the publicly visible system attack surface of hospitals located in Germany and found that more than 1 in 3 hospitals analyzed had vulnerabilities, including those that were part of the critical infrastructure at the time [11]. In times of scarce financial and human resources, ensuring information security is an enormous challenge [4]. In addition to technological and regulatory measures, research literature also highlights and discusses organizational measures to enhance information security in hospitals [2,8,9]. The reason for this is that a large proportion of information security breaches still occur internally as a result of employee misconduct [2,9,12]. German politicians have already recognized that a high level of information security in hospitals can only be achieved with holistic measures at a human, technological, and organizational level. In addition to providing the technological basis, employees must be made aware of security-relevant issues and obliged and motivated to comply with policies. The hospital management must revise information and communication processes and create appropriate support structures (eg, responsibilities, IT, and training) [13].

Empirical research on information security–related behavior of employees at work and its supporting organizational and individual factors is a relatively new field. Several researchers have found that in addition to the technological equipment of the organization to ensure information security, it is also necessary for the employees to follow security policies and consciously use information technology (compliance), because noncompliance can lead to security breaches with far-reaching consequences for the organization.

To explain information security–related behavior within the organization, various behavioral theories and frameworks have been used, including the Theory of Planned Behavior (TPB), Organizational Behavior (OB), and the Job Demands-Resources Model (JDRM) [14]. Hu et al [15] used the TPB to investigate the role of top management, organizational culture, and individual cognitive beliefs on information security–related behavior among alumni of Management Information Systems and Master of Business Administration programs, finding significant effects. D'Arcy and Greene [16] combined different theories for their research model and examined the influence of security culture, job satisfaction, and perceived organizational support on security compliance intentions among computer-using professionals, finding positive effects on security culture and job satisfaction. Solomon and Brown [17] used the TPB and showed relationships between organizational culture, information security culture—as an organizational subculture—and compliance. They further argued that goal orientation among employees has a stronger influence on compliance than rule orientation.

There is little empirical research on information security behavior in hospitals. Yeng et al [18] recently conducted a study in a paperless hospital in Ghana and assessed the security behavior of health care staff. In their study, the authors covered different individual, psychosocial, cultural, and work factors and found several significant correlations with participants' security behavior [18]. In their case studies, Hedström et al [19] showed that employees in health care organizations are exposed to different value conflicts, for example, health care values versus information security values, which they have to resolve quickly for each situation during their practice. This poses security risks that need to be considered by hospital management [19]. In our first article on information security published in 2022, we examined specific security-related practices of physicians when using smartphones and medical apps in everyday clinical practice. Among the apps examined, some were found to pose a certain risk to data protection, information security, and patient safety. A large majority of the participating physicians who communicated by their smartphone did not use General Data Protection Regulation (GDPR)–compliant messenger services and paid little attention to some security-relevant criteria (such as manufacturer information, information on data protection, and information security) when selecting medical apps for professional use [20]. Both the papers from 2022 described above and the following paper are based on the same dataset. While our first paper focused on specific security-related practices, this paper builds on this and examines the underlying constructs of this behavior.

Due to the particular threat of cyberattacks, the ongoing digitization, increasing regulatory requirements and the existing research gap on information security–related behavior in hospitals, the aim of this study was to explore the impact of job demands and job resources on resident physicians' information security–related compliance.

Theoretical Framework

To develop our research model, we used elements from the Extended JDRM as well as the OB research field.

JDRM Framework

The JDRM by Demerouti et al [21] postulates that working conditions can be divided into 2 categories—demands and resources—and are associated with different job-related outcomes. Demerouti et al [21] originally intended to use the JDRM to explain the development of burnout: in one of their studies, they confirmed the links between high demands and exhaustion and between a lack of resources and disengagement [21]. Three years later, a revised version was presented by Schaufeli and Bakker [22]. Their model included burnout and engagement as mediators between different predictors and different possible consequences. Since then, the JDRM has served as a theoretical basis in many studies in industrial and organizational psychology [23]. The JDRM has proven to be flexible, with researchers adapting it to their specific study context regarding the outcomes, demands, and resources to be considered, as well as the presumed mediators and moderators [24]. Pham et al [25] used the JDRM for their qualitative research on the information security–related behavior of employees in Vietnam, finding 3 security resources, 3 security demands, and 2 personal resources that affected security engagement and security compliance burnout as the mediators of security compliance [25].

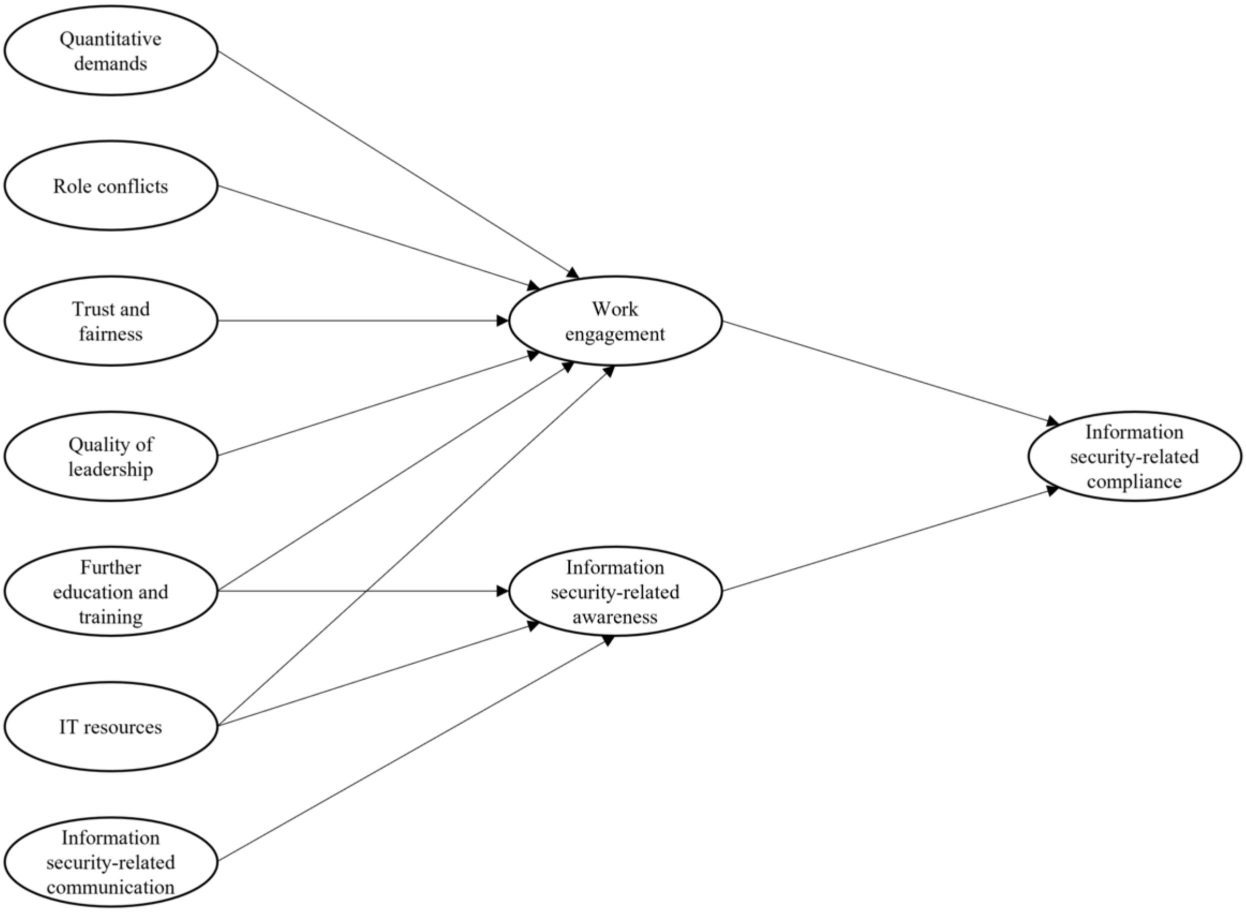
OB Framework

OB research has been around for about 100 years now [26]. There are 3 levels at which OB can be studied: the individual, the group, and the organizational level [27]. By analyzing information security–related compliance, our study focuses on how resident physicians behave as individuals in the organization and which job-related antecedents promote or inhibit compliance. Amankwa et al [28] chose—among others—concepts from OB research to explain information security–related compliance, showing that supportive organizational culture and end-user involvement had a significant effect on employees' attitudes toward compliance, which in turn had a positive effect on their behavioral intentions.

Research Model and Hypotheses

Our initial research model is presented in Figure 1. The relevant constructs, definitions, and corresponding items of the questionnaire can be found in Multimedia Appendix 1.

Figure 1. Research model.



As the first mediator, we have chosen work engagement ("WENG"). In a revised version of the JDRM presented by Schaufeli and Bakker [22], work engagement is considered to be a mediator of the relationship between job resources and job-related outcomes. The authors explained their findings through an underlying motivational process. Next to quality of leadership ("LEAD"), education and training ("EDUC"), and IT resources ("ITRE"), we chose trust and fairness ("TRFA") as the fourth potential resource. Unlike Schaufeli and Bakker [22], we also assumed a negative effect of certain job demands—quantitative demands ("QUAD") and role conflicts ("ROLC")—on work engagement. Here, we wanted to refer to the specific working conditions of physicians in hospitals, which are in general associated with high quantitative demands, but also—as Hedström et al [19] showed—different value conflicts.

Awareness ("AWAR"), as the second mediator, is one of the most commonly investigated variables related to information security–related behavior [29-32]. It is mostly considered to play a mediating role between the exogenous variables and the endogenous compliance variable ("COMP"). In our research model, it also acts as a mediator and reports on physicians' awareness of the risks, threats, policies and responsibilities related to information security in their hospital. In addition to information security–related communication ("COMM"), for which a relationship with information security–related behavior has already been shown in several studies [16,17,28,32], we added EDUC as well as ITRE as potential exogenous variables in our model. Here, the assumption was that both job resources the provision of adequate IT resources and good education and training could lead to greater information security–related awareness. A summary of our research hypotheses is provided in Table 1.

Table 1. Research hypotheses.

#	Hypotheses
H1a	Work engagement is positively related to information security–related compliance.
H1b	Information security–related awareness is positively related to information security–related compliance.
H2a	There is a negative relationship between quantitative demands and information security–related compliance mediated by work engagement.
H2b	There is a negative relationship between role conflicts and information security–related compliance mediated by work engagement.
H2c	There is a positive relationship between trust and fairness and information security–related compliance mediated by work engagement.
H2d	There is a positive relationship between quality of leadership and information security–related compliance mediated by work engagement.
H2e	There is a positive relationship between further education and training and information security–related compliance mediated by work engagement.
H2f	There is a positive relationship between IT resources and information security–related compliance mediated by work engagement.
H3a	There is a positive relationship between further education and training and information security–related compliance mediated by awareness.
H3b	There is a positive relationship between IT resources and information security–related compliance mediated by awareness.
H3c	There is a positive relationship between information security–related communication and information security–related compliance mediated by awareness.

Methods

Study Design

We used a cross-sectional, survey-based study design. A structured questionnaire in German was developed and designed with the free online survey tool LimeSurvey, whereby the first page contained information on the target group, the research project, the content of the questionnaire, and the estimated processing time. After accepting the privacy policy, participants were taken to the second page with demographic questions. These were followed by the main part including seven sections: (1) working conditions; (2) resilience; (3) job satisfaction and work engagement; (4) IT resources, information security, and data protection; (5) information security–related awareness and compliance; (6) technical affinity and innovative work behavior; and (7) mobile device usage. At the end of the survey, participants had the opportunity to share their comments with us. The sections relevant to this report are explained below.

- Working conditions: To assess working conditions, we used the following scales from the Copenhagen Psychosocial Questionnaire (COPSOQ): Quantitative Demands, Predictability, Role Conflicts, Quality of Leadership, Social Support, Feedback, Sense of Community, Trust and Fairness, and Appreciation. The COPSOQ is an internationally established instrument to measure psychosocial work factors, with good to very good validity and reliability for most of its scales [33]. In Germany, the third version of the questionnaire, which we used in our study, was published in 2019 [34]. In contrast to the original questionnaire, we divided the Social Support and Feedback scales into 2 subscales each (supervisors versus colleagues) to separate social support and feedback from supervisors from social support and feedback from colleagues, which could be rated differently, especially in hierarchical organizations. In addition, we used 2 scales (Uncertainty, and Further Education and Training) and 2 individual questions on working hours and shifts from the German instrument for stress-related job analysis for hospital

physicians (ISAK) [35,36]. We also included 4 self-developed items regarding IT resources in the hospital because we could not find a suitable scale in the research literature.
- Information security–related awareness and compliance: In addition to state-of-the-art technical information security solutions, employees of an organization should be aware of the importance of information security and trained accordingly to behave in a compliant manner [31,37]. The items that we used to assess information security–related communication, awareness, self-efficacy, top management commitment, and compliance are based on the works of Hu et al [15], D'Arcy and Greene [16], Karlsson et al [30], and Solomon and Brown [17]. We adapted the items to the clinical situation. Overall, this resulted in 4 items on communication and awareness, 2 on self-efficacy, 1 on top management commitment, and 4 on compliance.

The Work Engagement scale is also part of the COPSOQ and comprises 3 items. All English scales and single items were translated into German, checked independently by 2 bilinguals, and then adapted based on their comments. We used 5-point Likert scales, spanning from "To a very high degree" to "To a very low degree," from "Not true" to "Completely true," from "Always" to "Never/almost never," and from "Strongly disagree" to "Strongly agree," respectively.

We followed the Checklist for Reporting Results of Internet E-Surveys (CHERRIES; Multimedia Appendix 2) [38]. To ensure content validity, the survey was reviewed by faculty members and statisticians and modified accordingly. It was then piloted with a group of residents, who highlighted and took notes on any remaining ambiguities, which we corrected in the final questionnaire. We also tested the usability and technical functionality of the questionnaire before fielding it.

Determining the sample size for structural equation modeling (SEM) is difficult due to its flexibility. There are no generalized guidelines regarding sample size requirements [39]. Therefore, we used the rule of thumb of N≥200 [40].

Data Collection

Data were collected between March and June 2022. Our target group comprised physicians who are currently undergoing medical specialist training/residency training in hospitals in Germany (henceforth referred to as "residents" or "resident physicians"). We focused on residents because they predominantly belong to Generation Y and, compared with older generations, have been used to using digital technologies and especially mobile devices since childhood or youth; therefore, they represent a homogeneous target group in terms of this characteristic. An invitation with a link to the questionnaire was sent directly to the residents by email or social media channels, or indirectly by our contacts in the medical field. Important contacts were chief physicians, senior physicians, university professors, hospital managers, alumni networks, as well as presidents of the German medical societies. In addition, we asked medical experts with significant influence on social media platforms to share the link. The Hartmannbund, an important association of physicians in Germany, forwarded the link to its resident members.

Data Analysis

In total, we examined 2 job demands (QUAD and ROLC) and 5 job resources (TRFA, LEAD, EDUC, ITRE, and COMM) as exposure variables, 2 mediators (WENG and AWAR), and 1 outcome variable (COMP). All variables were measured using Likert scales and are interval scaled. A summary of our research hypotheses is provided in Table 1. To be statistically precise, we would like to highlight that the hypotheses are always intended to analyze the unique effects of the variables in the model and not the bivariate correlations.

For the analysis of our collected data, we applied SEM using the free software environment for statistical computing and graphics R (R Foundation for Statistical Computing) and the *lavaan* package. SEM refers to a family of data analysis techniques for complex relationships between multiple variables in a single research model and has become popular in the social and behavioral sciences [41,42]. Based on our research model and hypotheses, we followed the widely adopted 2-step approach to SEM by Anderson and Gerbing [43], in which the assessment of the measurement model is followed by the analysis of the structural model including tests of the research hypotheses. While the measurement model describes how the latent variables are measured by the indicator variables, the structural model shows how the latent variables are related. In the first step, a confirmatory factor analysis was performed to determine whether the initially proposed measurement model fits the data or needs to be respecified accordingly. Various fit indices exist for evaluating the global model fit. In accordance with Hu and Bentler [44], we used the standardized root mean squared residual (SRMR), the comparative fit index (CFI), and the root mean squared error of approximation (RMSEA). The authors recommended the following thresholds for fit indices that characterize a good global fit: SRMR<.08, CFI≥.95, and RMSEA<.06 [44]. Modification indices were also analyzed to determine the local fit. Here, we only made changes that were theoretically sensible in terms of our model. In addition, the results obtained were assessed regarding reliability, validity, and plausibility. If respecification was indicated, the entire evaluation process was repeated until a good model fit was achieved. In the second step, the relationships between the latent variables are analyzed by again evaluating the model fit and afterward interpreting the path coefficients β and the coefficient of determination R^2. To calculate the results, we used the robust WLSMV (weighted least squares means and variance adjusted) estimator. To ensure model identification, we only used factors with 3 or more items in the initial measurement model.

Ethical Considerations

Ethical review and approval were waived for this study. We are thus following the guidelines of the German Research Foundation (DFG) and the German Data Forum (RatSWD). The RatSWD is an independent body of empirically working scientists and representatives of the most important public institutions for data collection in Germany. It recommends a careful self-examination of research ethics in order to decide whether an ethical review by a committee is necessary for the planned research project. After careful self-examination, we were able to classify our research project as unobjectionable in terms of research ethics: neither patients nor other vulnerable groups took part in the study. Also, our study was not likely to trigger powerful emotions or cause severe psychological stress or traumatic experiences. It did not involve physical risks to the participants or result in physical pain. Furthermore, the participants were informed of the study. Potential risks of participation (such as social risks, risks of criminal or civil liability, financial loss, professional disadvantages, or damage to reputation; risk due to a difficult security situation in the study room) did not exist. Participation in the study did not involve any type of deception [45,46]. No compensation was provided.

Results

Sample Characteristics

A total of 611 people entered the survey, of whom 349 completed it. An exact statement on the response rate cannot be made because we do not have information on the number of residents who received the questionnaire indirectly through our contacts. Data from 6 participants had to be excluded due to conspicuous response patterns (n=4), work in a hospital abroad (n=1), and specific information in the comment section (n=1). Furthermore, only data from participating physicians who had a clear understanding of information security as distinct from data protection were included. This was checked with the help of a single comprehension item. Overall, data from 281 participants could be used for our analyses.

Table 2 shows the sociodemographic characteristics of the study participants. A large majority of the participants were female (173/281, 61.6%). The two age groups with the highest frequency were 31-35 years (114/281, 40.6%) and 26-30 years (108/281, 38.4%). There was a total of 16 specialties represented by at least two participants. Most participants were part of a residency program for internal medicine (62/281, 22.1%), followed by anesthesiology (47/281, 16.7%), surgery (43/281, 15.3%), and pediatrics and adolescent medicine (30/281, 10.7%). Participants were distributed across all residency levels, with the fewest residents in their fourth year (43/281, 15.3%) and most residents in their fifth year or above (82/281, 29.2%). The majority of the participants worked in a public hospital (174/281, 61.9%), while approximately one-fifth (62/281, 22.1%) worked in a nonprofit hospital and 14.2% (40/281) in a private hospital. Most of the participants were employed in a university/teaching hospital (230/281, 81.9%). The size of the hospitals (measured by the number of beds) in which the residents underwent their training varied, whereby most of them worked in a hospital with more than 800 beds (109/281, 38.8%), followed by hospitals with 300-800 beds (104/281, 37%). Fewer than one-fifth (52/281, 18.5%) of the residents had a job in a hospital with less than 300 beds.

Table 2. Sample characteristics (n=281).

Variable	Value, n (%)
Gender	
Female	173 (61.6)
Male	108 (38.4)
Other	0 (0)
Age group (years)	
21-25	7 (2.5)
26-30	108 (38.4)
31-35	114 (40.6)
36-40	30 (10.7)
40 or older	22 (7.8)
Specialty	
Internal medicine	62 (22.1)
Anesthesiology	47 (16.7)
Surgery	43 (15.3)
Pediatrics and adolescent medicine	30 (10.7)
Other	99 (35.2)
Residency level	
First year	59 (21)
Second year	47 (16.7)
Third year	50 (17.8)
Fourth year	43 (15.3)
Fifth year or higher	82 (29.2)
Hospital sponsorship	
Private	40 (14.2)
Public	174 (61.9)
Nonprofit	62 (22.1)
I don’t know	5 (1.8)
University or teaching hospital	
Yes	230 (81.9)
No	48 (17.1)
I don’t know	3 (1.1)
Hospital size (beds)	
Fewer than 300	52 (18.5)
300-800	104 (37)
More than 800	109 (38.8)
I don’t know	16 (5.7)

Evaluation of the Measurement Model

We analyzed the model fit, the reliability, and the validity of the measurement model and respecified it accordingly until we received sufficient results. We began with an assessment of the global fit of our original measurement model. Table 3 shows the measurement model fit indices for the initial and the final models. It revealed that the CFI was below the recommended threshold, whereas the RMSEA and SRMR already showed acceptable values.

Table 3. Fit indices of the measurement model (N=281).

Model	Chi-square (*df*)	*P* value	CFI[a]	RMSEA[b]	SRMR[c]
Initial measurement model	1217.7 (657)	<.001	.933	.055	.070
Respecified final measurement model	794.3 (514)	<.001	.965	.044	.055
Recommended thresholds	—[d]	≥.05	≥.95	<.06	<.08

[a] CFI: comparative fit index.

[b] RMSEA: root mean squared error of approximation.

[c] SRMR: standardized root mean squared residual.

[d] Not applicable.

Looking at Cronbach α of the latent variables of the initial measurement model, we received a poor value (<0.70) for the latent variable AWAR. Here, we decided to drop the first two items (AWAR1 and AWAR2) because, compared to the third and fourth items, they seem to be too special and an own latent variable, asking about concrete information security–related knowledge. One reason could be that at the time of the survey, information security policies and officers were not yet implemented in most hospitals and could, therefore, not be known to the physicians. After removing those 2 items, the Spearman-Brown coefficient was acceptable.

We further respecified the model according to the modification indices and added an error covariance between QUAD1 and QUAD2. We then dropped items with loadings <.50 (TRFA1 and ITRE1). Although the χ^2 goodness-of-fit statistic was still significant (*P*<.001), these respecifications resulted in global fit indices above the thresholds, as shown in Table 3, meaning that the measurement model can be considered to be an acceptable approximation of the empirical data. According to the Fornell-Larcker [47] criterion, we also compared the square root of the average variance extracted (AVE) for each variable with the corresponding correlations of all other constructs [47]. Since the square root of the AVE was higher than the correlations on the respective horizontal and vertical lines, as shown in Table 4, the discriminant validity of the variables can be confirmed. In addition, the AVE of each variable should be higher than .50 to achieve acceptable convergent validity. All AVEs were above this threshold. Afterward, we looked again at Cronbach α and the composite reliability of each latent variable. All values were higher than the recommended threshold of .70, indicating acceptable reliability of the latent variables. In summary, we obtained acceptable model fit, reliability, and validity for the measurement model. A summary of the quality indicators of the measurement model is presented in Table 5.

Table 4. Mean, SD, and correlations after item exclusion (N=281).

Latent variable	Mean (SD)	1	2	3	4	5	6	7	8	9	10
1. QUAD	3.672 (0.688)										
r		**.778**[a]									
P value											
2. ROLC	3.172 (0.781)										
r		.422	**.746**								
P value		<.001									
3. TRFA	2.961 (0.688)										
r		-.225	-.393	**.732**							
P value		<.001	<.001								
4. LEAD	2.897 (0.857)										
r		-.341	-.391	.348	**.822**						
P value		<.001	<.001	<.001							
5. EDUC	2.898 (0.897)										
r		-.389	-.463	.300	.627	**.820**					
P value		<.001	<.001	<.001	<.001						
6. ITRE	2.367 (0.910)										
r		-.166	-.231	.398	.138	.141	**.813**				
P value		.002	<.001	<.001	.013	.005					
7. COMM	2.686 (0.901)										
r		-.155	-.199	.313	.173	.254	.350	**.792**			
P value		.005	<.001	<.001	.001	<.001	<.001				
8. WENG	3.471 (0.674)										
r		-.135	-.261	.184	.379	.277	.086	.118	**.808**		
P value		.033	<.001	.001	<.001	<.001	.226	.027			
9. AWAR	3.835 (0.971)										
r		-.087	-.076	.076	.188	.182	-.004	.256	.121	**.904**	
P value		.141	.244	.197	.002	.001	.797	<.001	.036		
10. COMP	3.563 (0.731)										
r		-.080	-.196	.175	.201	.239	.084	.337	.219	.502	**.776**
P value		.234	<.001	.002	<.001	<.001	.129	<.001	<.001	<.001	

[a] Bold values on the diagonal are the square root of the corresponding average variance extracted (AVE).

Table 5. Measurement model quality indicators (n=281).

Latent variable	Loadings	Cronbach α	Composite reliability	AVE[a]
Quantitative demands		0.854	.883	.605
QUAD1	.644	—[b]	—	—
QUAD2	.710	—	—	—
QUAD3	.899	—	—	—
QUAD4	.888	—	—	—
QUAD5	.715	—	—	—
Role conflicts		0.743	.789	.557
ROLC1	.827	—	—	—
ROLC2	.711	—	—	—
ROLC3	.693	—	—	—
Trust and fairness		0.720	.774	.536
TRFA1	Dropped	—	—	—
TRFA2	.689	—	—	—
TRFA3	.835	—	—	—
TRFA4	.661	—	—	—
Quality of leadership		0.861	.892	.676
LEAD1	.884	—	—	—
LEAD2	.872	—	—	—
LEAD3	.806	—	—	—
LEAD4	.716	—	—	—
Further education and training		0.853	.891	.672
EDUC1	.784	—	—	—
EDUC2	.910	—	—	—
EDUC3	.721	—	—	—
EDUC4	.852	—	—	—
IT resources		0.776	.851	.661
ITRE1	Dropped	—	—	—
ITRE2	.831	—	—	—
ITRE3	.934	—	—	—
ITRE4	.647	—	—	—
Information security–related communication		0.826	.870	.627
COMM1	.761	—	—	—
COMM2	.731	—	—	—
COMM3	.788	—	—	—
COMM4	.880	—	—	—
Work engagement		0.777	.846	.652
WENG1	.658	—	—	—
WENG2	.969	—	—	—
WENG3	.765	—	—	—
Information security–related awareness		0.855[c]	.899	.817
AWAR1	Dropped	—	—	—

Latent variable	Loadings	Cronbach α	Composite reliability	AVE[a]
AWAR2	Dropped	—	—	—
AWAR3	.866	—	—	—
AWAR4	.940	—	—	—
Information security–related compliance		0.789	.853	.602
COMP1	.896	—	—	—
COMP2	.801	—	—	—
COMP3	.842	—	—	—
COMP4	.504	—	—	—

[a]AVE: average variance extracted.

[b]Not applicable.

[c]Spearman-Brown coefficient.

Analysis of the Structural Model and Hypotheses Testing

The initial structural model already showed a good global fit (Table 6). Nevertheless, we followed a proposed modification index and added a direct effect of COMM on COMP, resulting in the final structural model. The R^2 value for the endogenous variable COMP was .512, indicating that 51.2% of the variance of this variable can be explained by our structural model (Figure 2). The R^2 values for the mediators were .229 (WENG) and .154 (AWAR).

Figure 2 also presents the significant direct effects and standardized path coefficients (β). A summary of the hypotheses tested is provided in Multimedia Appendix 3. The results show a significant relationship between LEAD and WENG (β=.414, *P*=.001). There were no significant unique effects of the other exogenous variables tested (QUAD, ROLC, TRFA, EDUC, and ITRE) on WENG (all *P*>.05). Furthermore, we found a significant positive relationship between EDUC and AWAR (β=.173, *P*=.01), as well as between COMM and AWAR (β=.349, *P*<.001). In addition, there was a significant negative relationship between ITRE and AWAR (β=–.206, *P*=.01). Both mediators WENG (β=.208, *P*=.001) and AWAR (β=.552, *P*<.001) were significantly associated with the endogenous variable COMP, which supports H1a and H1b.

In order to test the mediation hypotheses, we calculated the partial indirect effects. A summary of the results is also presented in Multimedia Appendix 3. The results showed that WENG significantly mediated the relationship between LEAD and COMP (*P*=.03), supporting hypothesis H2d. The positive indirect effect resulted from the positive direct effects between LEAD and WENG and between WENG and COMP. AWAR mediated the relationships between the exogenous variables EDUC (*P*=.02), ITRE (*P*= .02), and COMM (*P*<.001) and the endogenous variable COMP, which supports H3a and H3c. Since we expected a positive indirect effect of ITRE on COMP, H3b could not be confirmed. The positive indirect effects of EDUC and COMM resulted from the positive direct effects involved.

Table 6. Fit indices of the structural model (n=281).

Model	Chi-square (*df*)	*P* value	CFI[a]	RMSEA[b]	SRMR[c]
Initial structural model	799.1 (527)	<.001	.966	.043	.059
Respecified final structural model	769.5 (526)	<.001	.970	.041	.056
Recommended thresholds	—[d]	≥.05	≥.95	<.06	<.08

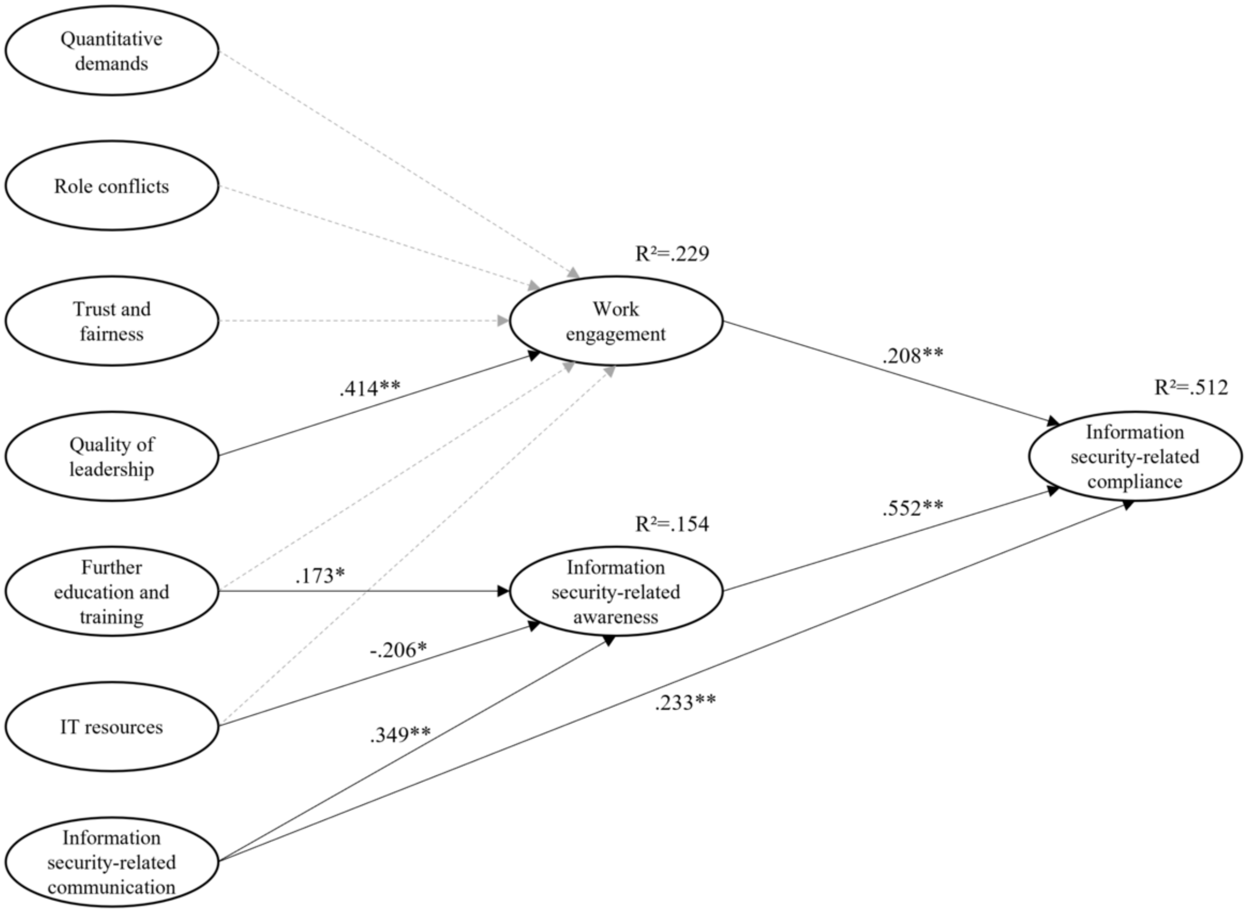
[a]CFI: comparative fit index.

[b]RMSEA: root mean squared error of approximation.

[c]SRMR: standardized root mean squared residual.

[d]Not applicable.

Figure 2. Structural model and standardized direct effects. *P<.05, **P<.01. Dashed gray arrows indicate statistically nonsignificant effects.



Discussion

Principal Findings and Comparison With Previous Work

With this study, we could statistically prove the influence of resident physicians' job resources and job demands on their information security–related compliance through the mediating roles of work engagement and awareness. This is one of the first studies to empirically analyze organizational antecedents of physicians' information security–related behavior in hospitals. Intensified research activities in this field are of great importance given that increasing digitization in hospitals and the associated risks from cybersecurity attacks require—in addition to technical solutions—well-considered organizational measures to enhance information security.

It was shown from the results that the information security–related compliance of resident physicians is significantly related to their work engagement. The assumed positive effect from our hypothesis was, therefore, confirmed. Pham et al [25] already described this relationship in their explorative study using interview analyses, which could be now supported by our results. In line with the findings of the study by Schaufeli and Bakker [22] and their revised version of the JDRM, in which they stated that "engagement is exclusively predicted by available job resources," we did not find unique relationships between the job demands involved (quantitative demands and role conflicts) and work engagement. Both demands also had no significant indirect effect on compliance. It can be, therefore, assumed that decreasing job demands would not be the most effective strategy in enhancing physicians' information security–related compliance, which would need to be further investigated by subsequent studies. Instead, our study highlights the importance of the job resource "quality of leadership," which had a significant indirect effect on information security–related compliance mediated by work engagement. Based on that, we can conclude that supervisors have the power to increase resident physicians' compliance through the motivational path: a good relationship with their direct supervisors, who are mostly senior physicians, increases resident physicians' work engagement, which leads to better compliance with the hospital's information security measures. According to our study, this relationship can be supported by the supervisor's efforts in the areas of staff development, job satisfaction, work planning, and conflict-solving. In light of the increasing digitization, this should be accompanied by digital and change management competencies [48].

Besides work engagement, awareness was the second mediator in our model, and based on our results, it is also significantly related to information security compliance. Here, our findings are in line with the results of several other studies, analyzing the role of awareness as a mediator in various industries [22,31,32]. We found 2 job resources with a unique positive effect on awareness and also a significant indirect effect on compliance: information security–related communication and further education and training. Communication is a highly effective factor in improving information security–related compliance, as has been shown in other studies [16,17]. Regular, comprehensible information security–related communication increases physicians' awareness of potential information security risks and threats. Here, we see a particular responsibility on the part of the hospital management. A communication strategy should be developed that makes all employees aware of the risks and threats and their particular information security responsibilities. In addition, information security–related communication had a direct effect on compliance. It could be assumed here that good communication also triggers automatic mechanisms for resolving critical situations. Furthermore, our results show that awareness can be created not only by communication but also by good education and training. Modern training concepts take a holistic approach and use various methods to raise awareness, for example, through personalized phishing simulations for training purposes and classic learning sessions with subsequent skills tests and personalized follow-up training based on individual results. Contrary to our hypothesis, IT resources are negatively related to awareness. One possible explanation could be that digitization and technologization increase confidence in the security of the tools used. On the other hand, if only outdated IT resources are available, physicians might be more sensitive to possible security risks.

Overall, our study provides initial insights into how the hospital as an organization could influence the individual information security–related behavior of resident physicians during clinical practice. In our opinion, the core element is the creation of an information security culture as a subculture of the organizational culture, which increases both the work engagement and the awareness of resident physicians through good leadership, information, and training.

Limitations

A few limitations should be taken into account when considering and evaluating our results, one of which pertains to the study design. First, since our study was conducted in Germany, it is not possible to directly apply the results to other countries. However, our study is intended to create incentives to conduct studies in hospitals in other countries to explore organizational factors for improving information security–related behavior. Second, when using a survey, comprehension problems on the part of the participants cannot be identified and addressed. However, we believe that the approach was the most suitable for the aim of the study and the selected target group as—for example—interviews would have resulted in a much smaller sample, which might have reduced the validity of the results. Third, the cross-sectional design does not allow us to form causal relationships between organizational antecedents, work engagement, awareness, and compliance, but only correlations. Any alterations in physicians' behavior due to changes in organizational factors can only be assessed with a longitudinal design. Nevertheless, we assume causal relationships based on logical considerations.

Even though it was an anonymous questionnaire, the risk of social desirability bias remains, with participants trying to be much more positive about their job-related attitudes and behavior. Furthermore, the fact that we cannot precisely determine the response rate represents another limitation. Since we do not have information on the number of residents who received the questionnaire indirectly through our contacts, for example, through chief physicians or the hospital management, we are not able to make an exact statement on the number of residents being invited, which is the basis for calculating the response rate. However, based on the information available to us, we estimate that fewer than 10% of those who received the questionnaire actually responded. If there are systematic differences between the responders and nonresponders, the results of our survey may not be representative of the target population. This so-called nonresponse bias may threaten the external validity of our study by reducing the representativeness of the results. However, previous research suggests that physician surveys are less susceptible to nonresponse bias than general population studies because they are a more homogeneous study population [49]. Furthermore, it is not always the case that a low response rate automatically reduces the representativeness, which is why the response rate should not be considered in isolation [50]. We believe the main reason for the low response rate was the heavy workload, which did not allow physicians the time to participate. Another indicator of the representativeness of a study is the sampling method [51]. The approach chosen for the data collection, in which residents were invited through different channels, may have led to a selection bias. For example, it is possible that digitally active and networked physicians were primarily addressed. The high proportion of physicians who received the questionnaire through the Hartmannbund may also contribute to the selection bias. Our study may have appealed to physicians with a higher average digital affinity than in the target population. To sum up, statistical conclusions on the entire target population should, therefore, always be drawn, taking into account the supposedly limited representativeness.

Future studies could use a longitudinal design to investigate the specific introduction of information security–related measures and the associated change in behavior. Furthermore, objective criteria, such as the number of breaches, could also be used as a measure of compliance instead of subjective self-reports. In addition, qualitative studies in the form of interviews are suitable for recording specific activities, opinions, perceptions, and concerns of physicians. It would also be interesting to conduct a study among senior physicians who have been working in hospitals for significantly longer and may need other measures to ensure that they behave in accordance with information security. For example, existing habits may be more difficult to change by adjusting resources and demands, and the resource "quality of leadership" may play a weaker role than for resident physicians.

Conclusions

A high standard of information security can only be achieved holistically through coordinated technical and organizational measures. In hospitals, this includes achieving a great level of work engagement among physicians through good leadership, but also creating awareness of the risks and threats relating to information security through communication and training. Hospital management is required to establish an information security culture that is informative and motivating and that raises awareness.

Acknowledgments

This study was actively supported by the Hartmannbund. The APF is paid for under the ATLAS project "Innovation and digital transformation in healthcare" funded by the State of North Rhine-Westphalia, Germany (grant ITG-1-1). The authors declare that they did not use generative artificial intelligence (AI) in any portion of the manuscript writing.

Data Availability

The datasets used and/or analyzed in the context of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JK designed the study, and SB-J made important contributions during this phase. With the kind support of the Hartmannbund, JK recruited most of the participants. SB-J helped distribute the questionnaire through her professional network. JK analyzed and interpreted the data. JK wrote the manuscript. SB-J was a major contributor to the revision of the manuscript. Both authors have approved the manuscript before submission. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Constructs, definitions, and corresponding items of the questionnaire.
[DOCX File , 38 KB - jmir_v27i1e46257_app1.docx]

Multimedia Appendix 2
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 24 KB - jmir_v27i1e46257_app2.docx]

Multimedia Appendix 3
Summary of the hypotheses tested.
[DOCX File , 16 KB - jmir_v27i1e46257_app3.docx]

References

1. Willing M, Dresen C, Haverkamp U, Schinzel S. Analyzing medical device connectivity and its effect on cyber security in german hospitals. BMC Med Inform Decis Mak 2020;20(1):246 [FREE Full text] [doi: 10.1186/s12911-020-01259-y] [Medline: 32993623]
2. Argaw ST, Bempong NE, Eshaya-Chauvin B, Flahault A. The state of research on cyberattacks against hospitals and available best practice recommendations: a scoping review. BMC Med Inform Decis Mak 2019;19(1):10 [FREE Full text] [doi: 10.1186/s12911-018-0724-5] [Medline: 30634962]
3. Yaqoob T, Abbas H, Atiquzzaman M. Security vulnerabilities, attacks, countermeasures, and regulations of networked medical devices—a review. IEEE Commun. Surv. Tutorials 2019;21(4):3723-3768. [doi: 10.1109/COMST.2019.2914094]
4. Jalali MS, Kaiser JP. Cybersecurity in hospitals: a systematic, organizational perspective. J Med Internet Res 2018;20(5):e10059 [FREE Full text] [doi: 10.2196/10059] [Medline: 29807882]
5. Wani TA, Mendoza A, Gray K, Smolenaers F. Status of bring-your-own-device (BYOD) security practices in Australian hospitals – a national survey. Health Policy and Technology 2022;11(3):100627. [doi: 10.1016/j.hlpt.2022.100627]
6. Kraushaar J, Bohnet-Joschko S. Prevalence and patterns of mobile device usage among physicians in clinical practice: a systematic review. Health Informatics J 2023;29(2):14604582231169296 [FREE Full text] [doi: 10.1177/14604582231169296] [Medline: 37063054]
7. Millard WB. Where bits and bytes meet flesh and blood. Annals of Emergency Medicine 2017;70(3):A17-A21. [doi: 10.1016/j.annemergmed.2017.07.008]
8. Argaw ST, Troncoso-Pastoriza JR, Lacey D, Florin MV, Calcavecchia F, Anderson D, et al. Cybersecurity of hospitals: discussing the challenges and working towards mitigating the risks. BMC Med Inform Decis Mak 2020;20(1):146 [FREE Full text] [doi: 10.1186/s12911-020-01161-7] [Medline: 32620167]
9. Kruse CS, Frederick B, Jacobson T, Monticone DK. Cybersecurity in healthcare: a systematic review of modern threats and trends. Technol Health Care 2017;25(1):1-10. [doi: 10.3233/THC-161263] [Medline: 27689562]

10. Tully J, Selzer J, Phillips JP, O'Connor P, Dameff C. Healthcare challenges in the era of cybersecurity. Health Secur 2020;18(3):228-231. [doi: 10.1089/hs.2019.0123] [Medline: 32559153]
11. Klick J, Koch R, Brandstetter T. Epidemic? The attack surface of German hospitals during the COVID-19 pandemic. : IEEE; 2021 Presented at: Proceedings of the 13th International Conference on Cyber Conflict (CyCon); May 25-28, 2021; Tallin, Estonia p. 73-94. [doi: 10.23919/cycon51939.2021.9468304]
12. Pollini A, Callari TC, Tedeschi A, Ruscio D, Save L, Chiarugi F, et al. Leveraging human factors in cybersecurity: an integrated methodological approach. Cogn Technol Work 2022;24(2):371-390 [FREE Full text] [doi: 10.1007/s10111-021-00683-y] [Medline: 34149309]
13. Branchenspezifischer Sicherheitsstandard „Medizinische Versorgung“. Deutsche Krankenhaus Gesellschaft. 2022 Dec 8. URL: https://www.dkgev.de/fileadmin/default/Mediapool/2_Themen/2.1_Digitalisierung_Daten/2.1.4._IT-Sicherheit_und_technischer_Datenschutz/2.1.4.1._IT-Sicherheit_im_Krankenhaus/Branchenspezifischer_Sicherheitsstandard_Medizinische_Versorgung_v1.2_Stand_2022-12-08.pdf [accessed 2024-11-10]
14. Kuppusamy P, Samy GN, Maarop N, Magalingam P, Kamaruddin N, Shanmugam B, et al. Systematic literature review of information security compliance behaviour theories. J. Phys.: Conf. Ser 2020;1551(1):012005. [doi: 10.1088/1742-6596/1551/1/012005]
15. Hu Q, Dinev T, Hart P, Cooke D. Managing employee compliance with information security policies: the critical role of top management and organizational culture*. Decision Sciences 2012;43(4):615-660. [doi: 10.1111/j.1540-5915.2012.00361.x]
16. D'Arcy J, Greene G. Security culture and the employment relationship as drivers of employees' security compliance. Info Mngmnt & Comp Security 2014;22(5):474-489. [doi: 10.1108/imcs-08-2013-0057]
17. Solomon G, Brown ITJ. The influence of organisational culture and information security culture on employee compliance behaviour. JEIM 2020;34(4):1203-1228. [doi: 10.1108/jeim-08-2019-0217]
18. Yeng PK, Fauzi MA, Yang B. A comprehensive assessment of human factors in cyber security compliance toward enhancing the security practice of healthcare staff in paperless hospitals. Information 2022;13(7):335. [doi: 10.3390/info13070335]
19. Hedström K, Kolkowska E, Karlsson F, Allen JP. Value conflicts for information security management. The Journal of Strategic Information Systems 2011;20(4):373-384. [doi: 10.1016/j.jsis.2011.06.001]
20. Kraushaar J, Bohnet-Joschko S. Smartphone use and security challenges in hospitals: a survey among resident physicians in Germany. Int J Environ Res Public Health 2022;19(24):16546 [FREE Full text] [doi: 10.3390/ijerph192416546] [Medline: 36554426]
21. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol 2001;86(3):499-512. [Medline: 11419809]
22. Schaufeli WB, Bakker AB. Job demands, job resources, and their relationship with burnout and engagement: a multi‐sample study. J Organ Behavior 2004;25(3):293-315. [doi: 10.1002/job.248]
23. Schaufeli WB, Taris TW. A critical review of the job demands-resources model: implications for improving work and health. In: Bauer GF, Hämmig O, editors. Bridging Occupational, Organizational and Public Health. Dordrecht, the Netherlands: Springer; 2014:43-68.
24. Taris TW, Schaufeli WB. The job demands-resources model. In: Clarke S, Probst TM, Guldenmund F, Passmore J, editors. The Wiley Blackwell Handbook of the Psychology of Occupational Safety and Workplace Health. Hoboken, NJ: Wiley; 2015:155-180.
25. Pham HC, El-Den J, Richardson J. Stress-based security compliance model – an exploratory study. ICS 2016;24(4):326-347. [doi: 10.1108/ics-10-2014-0067]
26. Borkowski N, Meese KA. Organizational Behavior in Health Care. Fourth Edition. Burlington, MA: Jones et Bartlett Learning; 2021.
27. Nahavandi A, Denhardt RB, Denhardt JV, Aristigueta MP. Organizational Behavior (Sage Texts). New Delhi, India: SAGE Publications; 2015.
28. Amankwa E, Loock M, Kritzinger E. Establishing information security policy compliance culture in organizations. ICS 2018;26(4):420-436. [doi: 10.1108/ics-09-2017-0063]
29. Li L, Xu L, He W, Chen Y, Chen H. Research and practical issues of enterprise information systems. In: Tjoa AM, Xu L DA, Raffai M, Novak NM, editors. Cyber Security AwarenessIts Impact on Employee's Behavior. Cham, Switzerland: Springer International Publishing; 2016:103-111.
30. Karlsson F, Karlsson M, Åström J. Measuring employees’ compliance – the importance of value pluralism. ICS 2017;25(3):279-299. [doi: 10.1108/ics-11-2016-0084]
31. Hwang I, Wakefield R, Kim S, Kim T. Security awareness: the first step in information security compliance behavior. Journal of Computer Information Systems 2019;61(4):345-356. [doi: 10.1080/08874417.2019.1650676]
32. Koohang A, Anderson J, Nord JH, Paliszkiewicz J. Building an awareness-centered information security policy compliance model. IMDS 2019;120(1):231-247. [doi: 10.1108/imds-07-2019-0412]
33. Burr H, Berthelsen H, Moncada S, Nübling M, Dupret E, Demiral Y, international COPSOQ Network. The third version of the Copenhagen psychosocial questionnaire. Saf Health Work 2019;10(4):482-503 [FREE Full text] [doi: 10.1016/j.shaw.2019.10.002] [Medline: 31890332]

34. Lincke HJ, Vomstein M, Lindner A, Nolle I, Häberle N, Haug A, et al. COPSOQ III in Germany: validation of a standard instrument to measure psychosocial factors at work. J Occup Med Toxicol 2021;16(1):50 [FREE Full text] [doi: 10.1186/s12995-021-00331-1] [Medline: 34784940]
35. Keller M, Bamberg E, Kersten M, Nienhaus A. Entwicklung eines instruments zur stressbezogenen arbeitsanalyse für klinikärztinnen und -ärzte (ISAK). Z Arb Wiss 2010;64:337-353. [doi: 10.1037/t70200-000]
36. Keller M, Bamberg E, Kersten M, Nienhaus A. Validierung des Instruments zur stressbezogenen Arbeitsanalyse für Klinikärztinnen und -ärzte (ISAK). Zeitschrift für Arbeits- und Organisationspsychologie A&O 2013;57(1):3-21. [doi: 10.1026/0932-4089/a000096]
37. Bulgurcu B, Cavusoglu H, Benbasat I. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS Quarterly 2010;34(3):523-548. [doi: 10.2307/25750690]
38. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
39. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas 2013;76(6):913-934 [FREE Full text] [doi: 10.1177/0013164413495237] [Medline: 25705052]
40. Kyriazos TA. Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. PSYCH 2018;09(08):2207-2230. [doi: 10.4236/psych.2018.98126]
41. Bollen KA, Noble MD. Structural equation models and the quantification of behavior. Proc Natl Acad Sci U S A 2011;108 Suppl 3(Suppl 3):15639-15646 [FREE Full text] [doi: 10.1073/pnas.1010661108] [Medline: 21730136]
42. Kline RB, St C. Principles and Practice of Structural Equation Modeling. New York City, NY: Guilford Press; 2022.
43. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin 1988;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
44. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
45. When do I need a statement by an ethics committee? Deutsche Forschungsgemeinschaft. 2023. URL: https://www.konsortswd.de/wp-content/uploads/RatSWD_Output9_Forschungsethik.pdf [accessed 2024-09-21]
46. Rat für Sozial- und Wirtschaftsdaten (RatSWD). Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften. Output Series 2017;9(5) [FREE Full text]
47. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 1981;18(1):39-50. [doi: 10.2307/3151312]
48. Balling S, Bohnet-Joschko S. Ärztliche führung im krankenhaus der zukunft. Gesundheitsökonomie & Qualitätsmanagement 2022;28(02):89-94. [doi: 10.1055/a-1920-3609]
49. Kellerman SE, Herold J. Physician response to surveys. A review of the literature. Am J Prev Med 2001;20(1):61-67. [doi: 10.1016/s0749-3797(00)00258-0] [Medline: 11137777]
50. Schouten B, Cobben F, Bethlehem J. Indicators for the representativeness of survey response. Survey Methodol 2008;35(1):101-113 [FREE Full text]
51. Ortmanns V, Schneider S. Can we assess survey representativeness of cross-national surveys using the education variable? Survey Research Methods 2016;10(3):189-210. [doi: 10.18148/srm/2016.v10i3.6608]

Abbreviations

AVE: average variance extracted
BYOD: bring your own device
CFI: comparative fit index
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COPSOQ: Copenhagen Psychosocial Questionnaire
GDPR: General Data Protection Regulation
JDRM: Job Demands-Resources Model
OB: Organizational Behavior
RMSEA: root mean squared error of approximation
SEM: structural equation modeling
SRMR: standardized root mean squared residual
TPB: Theory of Planned Behavior
WLSMV: weighted least squares means and variance adjusted

Edited by T de Azevedo Cardoso; submitted 03.02.23; peer-reviewed by B von Solms, I Schiering, J Klein; comments to author 29.01.24; revised version received 30.03.24; accepted 11.10.24; published 07.01.25.

Please cite as:
Kraushaar J, Bohnet-Joschko S
The Role of the Organization in Promoting Information Security–Related Behavior Among Resident Physicians in Hospitals in Germany: Cross-Sectional Questionnaire Study
J Med Internet Res 2025;27:e46257
URL: https://www.jmir.org/2025/1/e46257
doi:10.2196/46257
PMID:

Original Paper

Examining Intersectionality and Barriers to the Uptake of Video Consultations Among Older Adults From Disadvantaged Backgrounds With Limited English Proficiency: Qualitative Narrative Interview Study

Laiba Husain[1], DPhil, MPH, BSc; Trisha Greenhalgh[1], PhD

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

Corresponding Author:
Laiba Husain, DPhil, MPH, BSc
Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building, Radcliffe Observatory Quarter
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865289300
Email: laiba.husain@phc.ox.ac.uk

Abstract

Background: The rapid shift to video consultation services during the COVID-19 pandemic has raised concerns about exacerbating existing health inequities, particularly for disadvantaged populations. Intersectionality theory provides a valuable framework for understanding how multiple dimensions of disadvantage interact to shape health experiences and outcomes.

Objective: This study aims to explore how multiple dimensions of disadvantage—specifically older age, limited English proficiency, and low socioeconomic status—intersect to shape experiences with digital health services, focusing on video consultations.

Methods: Following familiarization visits and interviews with service providers, 17 older people with multiple markers of disadvantage (older age, low socioeconomic status, and limited English proficiency) were recruited in the Redbridge borough of London. Data collection included narrative interviews and ethnographic observations during home visits. Field notes captured participants' living conditions, family dynamics, and technological arrangements. Guided by intersectionality theory and digital capital concepts, interviews explored participants' experiences accessing health care remotely. Intersectional narrative analysis was used to identify key themes and examine how different forms of disadvantage interact. We developed theoretically informed narrative portraits and user personas to synthesize findings.

Results: Analysis revealed that the digitalization of health care can exacerbate existing inequities, erode trust, compound oppression, and reduce patient agency for multiply disadvantaged patient populations. Examining intersectionality illuminated how age, language proficiency, and socioeconomic status interact to create unique barriers and experiences. Key themes included the following: weakened presence in digital interactions, erosion of therapeutic relationships, shift from relational to distributed continuity, increased complexity leading to disorientation, engagement shaped by previous experiences of discrimination, and reduced patient agency.

Conclusions: This study provides critical insights into how the digitalization of health care can deepen disparities for older patients with low income and limited English proficiency. By applying intersectionality theory to digital health disparities, our findings underscore the need for multifaceted approaches to digital health equity that address the complex interplay of disadvantage. Recommendations include co-designing inclusive digital services, strengthening relational continuity, and developing targeted support to preserve agency and trust for marginalized groups in an increasingly digital health care landscape.

(J Med Internet Res 2025;27:e65690) doi:10.2196/65690

KEYWORDS

digital health disparities; video consultations; intersectionality; health inequity; digital capital; mobile phone

Introduction

Overview

While existing research has explored digital health disparities [1-5] and the application of intersectionality theory in health care [6-8] separately, there remains a critical gap in understanding how multiple dimensions of disadvantage intersect specifically in the context of digital health services. Previous studies have largely focused on single-axis analyses of digital exclusion [9-11]. For example, Ramsetty and Adams [9] examined socioeconomic status as a determinant of digital health access through quantitative analysis of technology ownership and internet connectivity rates. Eberly et al [10] used a large-scale quantitative study to analyze telemedicine access solely through the lens of age, finding lower video visit completion rates among older adults. Similarly, Donaghy et al [11] investigated the acceptance of video consultations primarily through the single dimension of technological literacy using survey data. While these quantitative studies provide valuable insights into individual factors, they have not adequately examined the compounded effects of age, language proficiency, and socioeconomic status on experiences with digital health technologies, particularly video consultations. Furthermore, there is a lack of in-depth, qualitative research that captures the lived experiences of multiply disadvantaged patients navigating these digital health spaces [12]. This study aimed to address this gap by applying an intersectional lens to explore how various forms of disadvantage interact and manifest in the context of video consultations, providing a more nuanced and comprehensive understanding of digital health disparities among susceptible populations.

We adopted an intersectional perspective from feminist studies to highlight the intersection and entanglement between digital technology, structural stratifications, and ingrained tendencies of “othering” in societies. This approach allowed us to move beyond simplistic notions of digital divisions to examine how digital technology is implicated in complex and intersectional systems of power. Drawing on narrative interviews with older, low-income individuals with limited English proficiency, we examined intersectionality and how video consultations can exacerbate existing inequities for multiply disadvantaged patient populations.

Our analysis revealed that digital health disparities operate at the intersection of multiple fracture lines of difference that mediate various spaces of inclusion and exclusion. We argue that addressing digital health disparities requires moving beyond single-axis analyses to consider how different aspects of disadvantage intersect in individual lives. This paper contributes to information systems (ISs) literature by providing a richer theorization of digital inequity, highlighting the need for intersectional approaches to digital health equity. We propose a research agenda that calls for IS scholars to reconceptualize actors beyond simplistic notions of “users,” to consider positioning rather than contextualizing, and to examine how digital health technologies are intertwined with producing and reproducing social orders and stratifications. Our findings have important implications for policy and practice in designing and implementing more inclusive and equitable digital health services.

Background and Literature Review

Overview

In the United Kingdom, video consultation uptake has increased significantly since the COVID-19 pandemic. National Health Service (NHS) England reported that before the pandemic, <1% of primary care appointments used video, but this number rose sharply during early 2020 to approximately 13% as part of the NHS Long Term Plan to expand digital-first care [13]. By 2023, about 9% of consultations in general practice continued to use video, reflecting an enduring shift toward remote health care delivery despite a return to face-to-face appointments as the dominant mode (65%) [14,15]. However, the use of video consultations varies significantly across demographic groups. Research indicates that older patients (aged >65 years) are less likely to use video services, with adoption rates around 7%, compared to 20% among younger adults aged between 18 and 34 years. Similarly, individuals from lower socioeconomic backgrounds and those with limited English proficiency face notable barriers, demonstrating uptake rates <10% [14,16].

Digital Health Inequalities and the Digital Divide

The IS literature has long engaged with the concept of the digital divide, traditionally focusing on accessibility, literacy, and adoption of digital technologies [17]. This discourse often conceptualizes individuals as “users” of technology, assigning them to specific group categories such as “the excluded” or based on binary divisions of “haves” and “have-nots” [18]. However, this notion of the digital divide fails to account for the multifaceted and compounded nature of digital inequality [12,19]. Recent scholarship has recognized that digital exclusion is a complex and dynamic phenomenon influenced by various factors beyond access to technology, including age, gender, education, and socioeconomic status [20-25]. As health care becomes increasingly digitalized, there is a growing need to understand how these broader digital inequalities manifest in health contexts. The concept of digital health equity has emerged as a critical area of study, examining how social determinants of health intersect with digital access and skills to shape health outcomes. This evolving field calls for more nuanced, intersectional approaches to understanding and addressing disparities in digital health access and use.

Intersectionality Theory

To address the limitations of single-axis analyses of digital exclusion, we turn to the concept of intersectionality from feminist studies. Originally proposed by Crenshaw [26] to expose the marginalization of Black women under both sexism and racism, intersectionality stands against the tendency in critical social theorizing to treat individuals in independent categories. It emphasizes that systems of oppression are inherently bound together, creating singular social experiences for people who bear the force of multiple systems [27]. While IS research has explored the relationship between IT and identity [28,29], most studies focus on the individual or group level, investigating how IT mediates or shapes identities. However, an intersectional perspective views subjectivity as emerging from differential experiences produced by multiple and intersecting power structures [26]. This approach moves beyond a behavioral, individualistic sense of identity to one of the "social positioning" of individuals within social structures [30]. Applying intersectionality to digital health inequalities allows us to examine how various dimensions of disadvantage—such as age, language proficiency, and socioeconomic status—interact and compound to shape individuals' experiences with digital health services.

Digital Capital Theory

The concept of digital capital by Ragnedda and Ruiu [31], an extension of the cultural capital theory by Bourdieu [32], offers another valuable lens for examining digital health disparities. Bourdieu [32] applied the idea of capital to signify the internal (eg, abilities and attitudes) and external (possessions and attributes) resources that people mobilize to achieve their goals in social life. He highlighted cultural capital as a form of capital that can be accumulated and transformed into other forms of capitals. Digital capital is made up of both digital competencies and digital technologies, which Ragnedda and Ruiu [31] argue is a form of capital in its own right and is essential for building up social, economic, and cultural resources in the digital world we live in today. Disparities involving digital skills originate in inequalities of access but are mediated by orientations that can only be understood in relation to total life contexts (eg, education, income bracket, age, location, and social support all influence a person's access to digital technologies and the level of digital skills they can acquire) [33]. Digital capital is a relatively new concept that scholars have begun to explore empirically through various methodological approaches [34]. Digital capital may be estimated, for example, at the individual level by assessing a person's digital literacy and skills, at the organizational level by measures of digital infrastructure (including the digital competence of personnel), and at the locality level in terms of the quality of the area's IT infrastructure.

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

Recent studies have begun to explore how digital capital interacts with other forms of capital to influence health outcomes and access to digital health services [35]. However, there remains a need for more in-depth, theoretically informed research on how digital capital intersects with other dimensions of disadvantage in shaping experiences with digital health technologies.

In the context of this study, intersectionality works as an overall guiding principle for understanding how people's lives and characteristics stem from and lead to multiple axes of disadvantage, while digital capital theory helps us understand how these axes of disadvantage play out in terms of access to and use of digital resources.

Research Gap and Objectives

Our previous narrative review [12] highlighted that while existing literature recognizes the multifaceted nature of digital inequality, there is a critical lack of in-depth, theoretically informed studies examining how different dimensions of disadvantage combine to affect digital health disparities. The review found the available literature on digital health disparities, particularly in relation to video consultations, to be sparse and primarily descriptive rather than explanatory. Most research has focused on identifying barriers and enablers without adequately exploring the complex interplay of factors contributing to these disparities.

Importantly, our review revealed no theoretically informed studies that examined how different dimensions of disadvantage combined to affect digital health disparities. This gap in the literature limits our understanding of how multiple disadvantages intersect and compound to shape individuals' experiences with digital health services.

Building on these findings and responding to the recommendations of our narrative review, this study aims to address these critical gaps. By doing so, we seek to move beyond the descriptive accounts that have dominated the field and provide a richer, more nuanced theorization of digital health disparities. This approach allows us to explore the complex ways in which different aspects of disadvantage interact, compound, and manifest in the context of digital health services. Our goal is to contribute to the development of more inclusive and equitable digital health services by offering insights into the lived experiences of multiply disadvantaged individuals, as called for in our previous work.

This study represents a direct response to the research agenda proposed in our narrative review, aiming to deepen our understanding of digital health disparities and inform more effective, equitable strategies for digital health implementation.

Methods

Overview

This study used a qualitative, interpretive approach to explore the intersecting effects of age, socioeconomic status, and limited English proficiency on experiences with digital health services. The focus on age, language proficiency, and socioeconomic status was informed by our aforementioned narrative review [12], indicating these as key predictors of lower digital service uptake and barriers to access. We adopted narrative inquiry [36] as our primary methodology, which aligns with our aim to center the voices and experiences of marginalized patients. The study was conducted in Redbridge, one of London's most diverse boroughs. It has >65% of its residents from Black and minority ethnic communities, predominantly Asian (42%). Over 90 different languages are spoken in the borough, with nearly a quarter of residents having a first language other than English. The borough includes several areas ranked among the 20% most deprived in England, with an unemployment rate (8%) exceeding London's average. This demographic and socioeconomic profile made Redbridge an ideal setting to examine intersecting dimensions of disadvantage in digital health care access. Participants were recruited through the Redbridge Respiratory Service within the North East London NHS Foundation Trust, community organizations, and snowball sampling.

Inclusion criteria for participants were as follows: (1) aged ≥65 years, (2) limited English proficiency (self-reported or identified by health care providers), (3) living within an Index of Multiple Deprivation decile of 1 to 5, (4) having attempted ≥1 video consultation, and (5) residing within the Redbridge borough.

As this was a doctoral research project, the primary investigator, LH, conducted all interviews and primary analysis, with regular supervision and analytical discussions with her supervisory team to challenge interpretations and biases throughout the research process. Semistructured narrative interviews were conducted with 17 participants between July 2022 and January 2023. All (17/17 100%) study participants had access to health care through the NHS despite being multiply disadvantaged. The interview guide was designed to elicit rich narratives about participants' experiences with digital health services, particularly video consultations. Questions explored their overall health care journey; experiences with digital technologies; and perceptions of how their age, language abilities, and financial situation affected their access to and use of digital health services. As a multilingual researcher fluent in Hindi, Urdu, and English, LH conducted all interviews directly in participants' preferred languages. Participants could freely switch between languages during interviews, which many did. This linguistic flexibility allowed participants to express themselves more fully and comfortably when discussing their experiences. All non-English segments were translated to English during transcription by LH, while "broken English" was transcribed verbatim when the meaning was clear. In addition to interviews, LH conducted ethnographic observations with 9 (53%) of the 17 participants, including home visits and participation in daily routines. These observations provided valuable contextual insights into participants' living conditions, family environments, and technological exposure. The ethnographic data were analyzed alongside interview transcripts to provide richer context to participants' narratives and to understand how their home environment and daily routines influenced their experiences with digital health care.

LH used thematic narrative analysis [37] using an intersectionality lens to identify key themes while preserving the integrity of individual stories. The analysis process involved the following: (1) familiarization with the data through repeated reading of transcripts, (2) open coding to identify initial themes and patterns, (3) development of a coding framework informed by intersectionality theory, (4) axial coding to explore relationships between themes, and (5) selective coding to refine and integrate themes into a coherent narrative.

To enhance the intersectional analysis, a modified version of the Equity Design Collaborative's meta-empathy mapping approach was incorporated [38]. This approach focuses on understanding users' needs through transformative empathy, moving beyond surface-level observations to deeply understand how systemic barriers and power structures affect user experiences. This modified approach shaped the analysis in 3 key ways.

First, it guided us to examine both immediate barriers (ie, language difficulties) and deeper structural challenges (ie, how health care systems may inadvertently privilege certain cultural norms). Second, it helped identify power dynamics in health care relationships. Third, it informed how outputs such as user personas (discussed in subsequent sections) were structured to capture both individual circumstances and systemic influences on participants' experiences. This methodological lens helped ensure findings reflected not only individual experiences but also broader systemic factors affecting remote care access.

Several strategies were used to ensure the trustworthiness of findings [39], including prolonged engagement with participants through multiple interactions, triangulation of data sources, member checking with participants to verify our interpretations, and maintaining a reflexive journal to document the decision-making process and potential biases.

Particular attention was paid to ensuring that participants fully understood the nature of the research and their rights, given potential language barriers and vulnerabilities. As researchers, we also acknowledge our own positionalities and how they may influence the research process. LH is a female Muslim of South Asian descent, which facilitated trust building with many participants but also required ongoing reflexivity to avoid assumptions based on shared cultural backgrounds. To synthesize findings, LH developed 4 [40] theoretically informed narrative portraits and user personas that captured key intersecting dimensions of disadvantage, including how age, language barriers, socioeconomic status, cultural factors, and health conditions combined to shape participants' experiences with digital health care. These are reported in detail in our previous paper [40] and serve as complementary outputs that distill key themes into accessible archetypal stories, balancing the need to honor individual perspectives with extracting cross-cutting insights about the interplay of technology and disadvantage.

Ethical Considerations

The study received ethics approval from the NHS Research Ethics Committee and the University of Oxford's Central University Research Ethics Committee. Informed consent was obtained from all participants, and pseudonyms were used to protect their identities. Ethical approval was obtained from East Midlands—Leicester South Research Ethics Committee and UK Health Research Authority (September 2021, reference number: 21/EM/0170; Integrated Research Application System project ID 300719).

Results

Overview

The analysis revealed 8 key themes that illuminate how multiple dimensions of disadvantage intersect to shape experiences with digital health services, particularly video consultations. These themes highlight the complex ways in which the digitalization of health care can exacerbate existing inequities, erode trust, compound oppression, and reduce patient agency for multiply disadvantaged patient populations.

Digital Interactions May Have a Weak Presence

The concept of "absent presence," originally developed by Gergen [41] to describe technology-induced distraction in face-to-face interactions, takes on new significance in the context of digital health care. The analysis found that participants consistently reported a sense of diminished presence in digital health care interactions, particularly in video consultations. This weakness of presence manifested in 3 key ways: delayed responses, mechanical intonation, and perceived motionlessness.

Delayed responses were frequently noted by participants as a sign of disengagement. Maneshi, an Indian immigrant aged 83 years, articulated this experience as follows:

> *If my doctor is doing twenty other things on his side of the screen while I'm talking, even if he is technically listening, it just doesn't feel like a genuine conversation to me...When I go to the GP [general practitioner] in person, those distractions aren't there. I know the GP is looking at me and having a conversation with me.*

This account illustrates how the perceived divided attention of health care providers during video consultations can erode the sense of a genuine, engaging interaction. The lack of immediate responsiveness disrupts the natural rhythm of conversation, leading to a feeling of disconnection.

Mechanical intonation was another aspect that contributed to the sense of weak presence. Fowzia, a recently widowed immigrant from Pakistan, expressed this concern as follows:

> *The absence of eye contact and non-verbal cues makes you feel...I don't know it's off putting. It feels like I'm talking to a computer program instead of a real doctor. They used to, the doctors, they would bring a lot of comfort and assurance; now, it's replaced with a sense of disinterest. It just makes you think you know? Do they even care?*

Participants described feeling as though they were interacting with a "computer program" rather than a real person, noting the flat, unemotional tone often used by health care providers during video consultations. Recent research in general practice has shown that video-mediated communication affects sensory conditions in clinical interactions, technical issues such as delayed facial expressions and sound can disrupt natural conversation flow, and the screen interface limits health care providers' ability to read and respond to patient cues [42]. This lack of vocal vitality, as conceptualized by Stern [43], can significantly impact the patient's perception of the provider's engagement and empathy.

The third manifestation of weak presence was perceived motionlessness. Abed, a recently retired repairman aged 69 years, highlighted this issue as follows:

> *His movements don't signal any interest. I mean I guess that could just be because I'm seeing his face and can't tell over video call how much he's actually moving. But I'm not seeing hand movements or his head moving much you know.*

In face-to-face interactions, body language and subtle movements play a crucial role in conveying attention and engagement. However, in video consultations, the limited visual field and potential technical constraints can result in an appearance of stillness that patients interpret as a lack of involvement or interest. Basmah, a Bangladeshi immigrant aged 73 years with arthritis and limited mobility, noted the following:

> *Back in the day, a doctor's touch and a comforting pat on the back meant so much. Now, it's different, you just hear typing with those hands.*

This triad of delayed responses, mechanical intonation, and perceived motionlessness collectively contributed to a sense of "absent presence" in digital health care interactions. Therefore, many (11/17, 65%) participants reported feeling disconnected and unheard during video consultations, potentially impacting the quality of care and patient satisfaction. Mohammad, a Sri Lankan immigrant aged 70 years with multiple chronic conditions and living with his extended family, starkly put the following:

> *Facial expression and body language is so important. When you talk to people you can see what's wrong and things like that and some of the things you don't know whether when the GP say things, how they mean it because I can't tell over remote. I don't think the whole system is fracturing, I think it's completely collapsing.*

While all patients may experience reduced presence in video consultations, this sense of disconnection is particularly problematic for patients with limited English proficiency, who rely heavily on nonverbal cues and physical presence to support communication and understanding. The technical barriers to presence can additionally compound existing language barriers, potentially leading to misunderstandings or missed clinical information.

Digital Encounters May Weaken Relationships

Analysis revealed that digital encounters, particularly when there was no previous in-person relationship, often weakened the therapeutic relationship between patients and health care providers. Participants consistently expressed difficulty in forming bonds through screens, highlighting the importance of relational foundations for ethical care.

Rajpreet, a first-generation Indian woman experiencing economic hardship and multiple chronic conditions, articulated this challenge as follows:

> *Doing this by video makes it harder. If I met her in person, maybe we could connect more...but I know women like her, that's not to say in a bad way, just that she doesn't really struggle with the same things that I do, it's very different when two women come*

> *from...I don't know how to say...just that...she's of a different social class...if that makes sense.*

This quote illustrates how the lack of physical presence can exacerbate perceived social and cultural differences, making it more difficult to establish a connection and mutual understanding.

The importance of preexisting relationships was emphasized by many (14/17, 82%) participants. Fatima, an Afghan woman aged 81 years with diabetes and heart conditions, who relies on her children for translation, explained the following:

> *I think it [video consultation] was easier because I had met Dr. Samari before. So I already had that initial relationship with her. I've seen her a few times. I felt comfortable with her. I think it would have been a bit more awkward if it had been like a first meeting.*

Conversely, participants who had video consultations with unfamiliar clinicians often reported fewer positive experiences. Ramnik, an Indian immigrant aged 78 years with hypertension and diabetes, who lives alone and relies on community support services, explained the following:

> *You stay in the online waiting room being all confused and then they let you into the call. And its someone you've never seen before. And he just wants yes or no. Then it's finished. You can go. That's it. You know? So I don't know. I don't like it. I don't. This is why I don't want to see the GP unless it's really bad. When I did it [video consultation] with my own specialist it wasn't like this. That time was good because he knew me and I knew him.*

This account highlights how the lack of a prior relationship can lead to a sense of disconnection and dissatisfaction with the consultation process.

Some (5/17, 29%) participants stressed the need for occasional in-person visits to establish and maintain a connection:

> *It'd be nice if they could see a person like myself every three months or normal patients at least once every six months. Yeah, so you know, so then you know them. So that builds up some, like, friendship as well. Now there's no friendship. But that's why, you know, it's important we get to know who the person is. Here there's no chance to do that.*

This suggestion underscores the perceived value of face-to-face interactions in building and sustaining therapeutic relationships.

The findings indicate that even when participants had successive video consultations with the same health care provider, they rarely developed a sense of building a strong and positive therapeutic relationship. This contrasts with face-to-face environments, where the patient-provider relationship typically strengthens with each encounter. Some (4/17, 23%) participants even described a deterioration in their relationship over repeated remote encounters, unless preceded by face-to-face meetings.

The weakening of therapeutic relationships through video consultations has consequences for marginalized older adults who often rely on trusted health care provider relationships to maintain engagement with health services. Limited English speakers face additional barriers to relationship building in digital settings, where language difficulties are compounded by technical constraints.

The Shift From Relational to Distributed Continuity

The analysis also revealed a significant shift from relational to distributed continuity as digitalization increased after the COVID-19 pandemic. This transition often left patients feeling lost, unsupported, and struggling to navigate their care effectively. The loss of relational continuity was particularly pronounced for marginalized older adults who had previously relied on long-standing relationships with health care providers.

Tasneem, a Bengali immigrant aged 75 years, expressed this loss as follows:

> *Dr. Talib has seen me through so much. I could speak to her about anything, and she really listened and understood me.... [Now] I don't even know my doctor's name. How can I trust someone I don't even know?*

This quote encapsulates the profound impact of losing a trusted health care relationship and the challenge of building trust in a system of distributed care.

The fragmentation of care was a recurring theme among participants. Arjun, a retiree aged 72 years, described his frustration with the depersonalization of care as follows:

> *It's becoming less and less personal. It's like you are not a person there you are just a face. For example, the company I work for we went through the same sort of process, I joined in '87 with BT and then in 1993 they brought this employee individual identification numbers they're called right, so you're given a nine digit number and then after that whenever you wanted to talk to the HR department or pay group or something like that, that's the thing that you gave them and that's it. So you only become a number, in this case you're only becoming a face to the GP, always a new number, new face, new GP, it's not who you are or what you are.*

This shift to distributed continuity often resulted in communication breakdowns and potential risks to patient care. Priya, a Punjabi woman aged 71 years with diabetes, expressed her concerns as follows:

> *Ordering repeat medicine with the GP always main problem. They are, you know, taking so long and even sometimes they don't know your condition there because communication, everything is so broken. They just forget everything. Like you are the new person to them. You know that's the problem with the GP always.*
>
> *They don't know my history or me. I feel scared they'll make a mistake with the dosage but what can I do?*

The loss of community connections was another significant aspect of this shift. Another participant highlighted how digital triaging erased the familiarity and efficiency of previous care arrangements:

> *[B]efore, you know, they knew everything already. They knew you. And they had my history. Whereas. Where I am now, as I said, the first five minutes of any session are taken up with me explaining who I am, what my situation is. And then, you know, explaining what the problem is, why I'm seeing them all the rest. And this is repeating after filling out a whole econsult first. It's a right shame.*

The fragmentation of care extended beyond primary care to specialist services. Najma, a Pakistani immigrant aged 66 years with limited English proficiency and respiratory conditions, living in council housing, described her frustration with revolving specialists and poor communication:

> *Now my specialist has also now changed. Three of them I went through. At the moment one, I think one is they know she's the main one. She was good, but now I don't know which one. Last two weeks ago they giving me one medicine which is a high dose they asked me to stop and they my daughter she sent a lot of e-mail to her but she don't reply anything here so this one [the new specialist] I think she's not so good. Communication is not good. You keep chasing them, you know?*

These experiences highlighted how the shift to distributed continuity has disrupted the holistic, coordinated care that many patients, especially those from marginalized communities, relied upon. The fragmentation of care across digital platforms has created new challenges in maintaining consistent, personalized health care relationships and effective communication between health care providers and patients.

Digital Interactions May Compound Oppression

Findings revealed that digital health services often compounded existing forms of oppression and discrimination, particularly for participants with limited English proficiency and low digital literacy. These individuals faced multiple, intersecting barriers in accessing and navigating digital health platforms, which exacerbated their existing challenges in health care settings.

Rajpreet, a first-generation Indian woman aged 74 years living in council housing with multiple chronic conditions, articulated this challenge:

> *I told her, don't mind me saying this Dr. Kaur, but you don't really know what it's like, you come from a different background or world whatever you want to call it [laughs], you know medically sure, but over video you can't grasp it, you won't understand because you don't have it either, my condition, what surrounds it, that sort of thing...*

This quote illustrates how digital interactions can amplify cultural and experiential gaps between patients and health care providers, making it more challenging to discuss sensitive health issues.

Language barriers were particularly problematic in digital settings. Samiyah, a Pakistani grandmother aged 69 years, with limited English proficiency and chronic pain conditions who recently moved in with her son's family, noted the following:

> *It's already hard to explain my symptoms in English, but over video, it's even worse. I can't use gestures or show them exactly where it hurts. Sometimes I feel like they don't understand me at all.*

Low digital literacy compounded these challenges. Tasneem shared the following:

> *I struggle with technology, and now I have to figure out how to use these apps just to see my doctor. It makes me feel stupid and left behind. Sometimes I just give up.*

The shift to video consultations also highlighted existing inequalities in access to technology. Mohan, living on a basic pension in social housing with diabetes and heart disease, explained the following:

> *They tell us to do video calls, but I don't have a smartphone or good internet. It's like they're saying healthcare is only for people who can afford fancy gadgets.*

Participants from marginalized groups often found themselves at the intersection of multiple disadvantages—language barriers, cultural differences, low digital literacy, and limited access to technology—all of which were amplified with video consultations.

Digitalization May Erode Trust in Health Providers and Systems

Findings revealed that the shift to video consultations also frequently eroded participants' trust in health care providers and systems. This erosion of trust was often rooted in a sense of depersonalization and lack of continuity in care, particularly for marginalized and susceptible populations. Research has also documented how marginalized communities may have lower trust in health care providers due to several factors, including historical medical mistreatment, documented disparities in quality of care, and systemic barriers to culturally and linguistically appropriate services [44,45]. In this study, this broader context of health care mistrust appeared to be exacerbated by the shift to video consultations, particularly when patients could not establish consistent relationships with health care providers.

Hasan, with his multiple chronic conditions and previous negative health care experiences, articulated this sentiment strongly in the following manner:

> *You don't wanna be a part of this system. Like you can't trust these people in the NHS... Because I need the GP, but to be honest, I don't trust that like you know, she's not very good and she she doesn't really care, right? That's the impression I get so I try not to go but then it gets worse.*

This quote illustrates how the perceived lack of care and attention in digital interactions can lead to a cycle of disengagement and worsening health outcomes.

The impersonal nature of video consultations was a recurring theme. As Fowzia explained the following:

> *The personal connection is not the same. I miss the stuff that didn't require words, stuff you could just see and feel and the comfort of being physically there in the same room. It's harder to build that trust through a screen.*

This highlighted the importance of nonverbal cues and physical presence in building trust, which many felt was lacking in digital interactions.

The erosion of long-standing relationships with health care providers was particularly distressing for some participants. Fatima shared the following:

> *Thing is you can't even think about trusting the GP now even if you wanted to this way [over video consults]. These things take a lot of time, beta, my old GP, I knew him for 10 years, every small flu, back ache, little cold he knew it all. I don't even know if I have another 5 years left in me and if I keep seeing a different GP over a different platform where does that leave me, beta?*

This account underscores how the fragmentation of care across multiple health care providers and platforms can disrupt the accumulation of shared knowledge and understanding that forms the basis of trust in health care relationships.

The erosion of trust sometimes led to nonadherence to medical advice. Zainab, a Pakistani immigrant aged 73 years with heart disease, who had previously experienced dismissive treatment from health care providers, explained the following:

> *I mean why should I take it [the medication], ok yeah they prescribe it but they don't even bother listening to me, click clacketing away at their keyboard, not even looking at the screen, they think they know what's wrong with me just like that? I don't trust it one bit course I'm not gonna take it.*

For some, the distrust in the health care system led to anxiety and avoidance. Hamida, a Bangladeshi mother aged 75 years living with her daughter's family and experiencing both physical health conditions and growing anxiety about health care interactions, shared the following:

> *My daughter thinks I have anxiety because I don't want to deal with the NHS. I don't know, maybe I do. But to me it's more about the fact that they can't be trusted.*

However, it is important to note that not all participants experienced an erosion of trust. In a notable disconfirming case, 1 participant reported high levels of trust in video consultations due to a preexisting, long-term relationship with her general practitioner. This suggests that strong, preexisting therapeutic relationships may buffer against the potential erosion of trust in digital interactions.

These findings highlight the complex relationship between digitalization and trust in health care. Video consultations can risk eroding the personal connections and continuity of care that many patients, particularly those from marginalized groups, rely on to build trust with their health care providers. This erosion of trust can have serious implications for patient engagement, adherence to treatment, and overall health outcomes.

Digitalization Increases Complexity, Which May Lead to Disorientation

The analysis highlighted that the introduction of multiple digital platforms and access points often led to increased complexity and disorientation for participants. Many (12/17, 70%) struggled to navigate the various digital pathways and processes, leading to frustration and, in some cases, disengagement from health care services.

Mukesh, a man aged 85 years with cognitive challenges, articulated this complexity in the following manner:

> *The health system as a whole there is a lack of communication and sharing between the different functions. I mean I've got access to a long COVID clinic. Then I also deal with my GP and there should be some information exchange between the two. Not me filling them both in on my own. And then with the hospital, they're not connected in a way that they can get anything from my GP to the hospital and it has been very frustrating.*

This quote highlights how the fragmentation of digital systems can place an additional burden on patients, particularly those with complex or multiple health conditions.

The sense of being overwhelmed by digital options was a common theme. Ramnik expressed the following:

> *I felt like I was drowning in all the options, do I call the practice, do I do this econsult thing, do I use the NHS app, do I first check the website, but then the website it too complicated anyways and then I'm back to square one. It's overwhelming, especially when you're already struggling with other things, this is the last thing I should have to worry about.*

This account illustrates how the proliferation of digital access points, while intended to improve accessibility, can paradoxically create barriers for some patients.

The disorientation experienced in digital health care settings was often compounded by language barriers and limited digital literacy. As Samiyah shared the following:

> *I wouldn't know the first thing about doing a video call. He [son] set it up for me, clicked some stuff, and had it up and running and I just sat there. They did all the talking without me, but it probably was for the best anyways because I don't know if I could've even said what I needed to properly, the language issue, the screen issue, just looking at it all was too much for me.*

This quote underscores how digital health care can inadvertently exclude patients who lack the necessary language skills or technological proficiency, potentially exacerbating existing health inequalities.

The complexity of digital systems also led to challenges in maintaining continuity of care. Mohammad noted the following:

> *Look I've got more health issues than I can count on both of my hands, navigating through different services and providers, half remote, half in person, half on the phone, I just feel like I'm lost in a maze. I'm constantly juggling between appointments with different specialists, trying to piece together the whole story for myself and for the providers too because they themselves don't know the full story. It's exhausting and overwhelming.*

This account highlights how the fragmentation of care across various digital and in-person platforms can create a significant cognitive and emotional burden for patients, particularly those managing multiple health conditions.

The disorientation caused by digital complexity was often exacerbated by socioeconomic factors. As Maneshi explained the following:

> *Not having the money to go private adds another layer of hard to the mix. I can't afford the luxury of choosing the most convenient healthcare option. No. Instead, I'm forced to navigate through a patchwork of NHS resources that takes weeks, months, even years and it's just getting worse because they're trying to move things online now and my brain is already just scattered from long-covid first and then the mess that is the NHS trying to be something they aren't so this is just the cherry on top.*

This quote illustrates how the digitalization of health care, when not adequately supported or implemented, can compound existing health inequalities and create additional barriers for those already struggling to access care.

Engagement With Digital Services May be Shaped by Previous Experiences of Racism and Discrimination

Findings revealed that participants' willingness to engage with digital health services was often profoundly influenced by their previous experiences of racism and discrimination within the health care system. These past negative experiences created a foundation of mistrust that often extended to new digital health initiatives.

Hamida, with her growing anxiety, articulated this heightened vigilance as follows:

> *So I'm literally now I'm very vigilant. I will check every single medication. I will read the leaflet 3 times, I will Google it because I don't trust these people they already prescribed the wrong one [medication] to me once before.*

This quote illustrates how past negative experiences can lead to a deep-seated mistrust that influences future interactions with health care services, including digital platforms.

The intersection of racial identity and socioeconomic status in shaping health care experiences was highlighted by Abed in the following manner:

> *Well, yes. At the end of the day, yes, I do think that if I was a white person from a rich background, instead of brown and poor, I might have been treated differently.*

This perception of differential treatment based on race and class extended to video consultations, with participants expressing concern that these biases would persist in digital interactions.

Some (9/17, 53%) participants described a resigned acceptance of interpersonal racism from health care staff:

> *You know, sometimes you can tell they're a little bit... racist, but it's OK, you know, it's not a big deal.*
>
> *And sometimes you get certain doctors who are racist. But it's whatever, I'm used to that.*

These statements reveal a troubling normalization of discriminatory treatment, which may influence patients' expectations and engagement with digital health services.

Language barriers were identified as a particular challenge in video consultations, often intersecting with perceptions of racial discrimination:

> *They make it such that you know she's not given one [appointment for his wife]. If you know what I mean. It's I think it's easier for them to treat people who, for example, who they can see is very much quite different. For example, if you have really great English and you're able to communicate, you're able to get your points across, you won't deal with it [racism] as much as you will for example, if, it's just ok.*

This quote highlights how language proficiency can intersect with racial bias to create additional barriers to accessing health care, including digital services.

Some (6/17, 35%) participants, like Mukesh, expressed concern that video consultations might amplify existing biases:

> *Not too long ago my wife she had a video appointment because she had an issue with the hand. So the doctor like any questions we asked, he was just reluctant to answer and he had the, you know, the sarcasm.... So you just know where some doctors, you don't know whether it's racist or not, but like because we call it that because we feel that way.*

This account suggests that the physical distance in video consultations may exacerbate perceptions of dismissive or discriminatory treatment.

The cumulative effect of these experiences led participants like Tasneem to express extreme distrust in the health care system:

> *This is what I mean. And that's why I can't trust these people with your life. You can't. You can't. You you'd rather die than trust some idiot with your life because you're gonna die anyway.*

This level of distrust poses significant challenges to the adoption and effective use of digital health services among marginalized communities.

Digital Interactions May Reduce Patient Agency

The study revealed that the cumulative effect of the previously discussed factors often resulted in reduced patient agency. Many

(14/17, 82%) participants, particularly those facing multiple disadvantages, felt disempowered and unable to effectively advocate for themselves in digital health care interactions:

> *I'm not very happy with the new GP, but I'm scared to change. The next one might be even worse.*

This quote illustrates how the lack of options and fear of further negative experiences can trap patients in unsatisfactory care arrangements, reducing their ability to seek better alternatives.

The complexity of digital systems often led to confusion and reliance on others, as Najma described the following:

> *I can't tell you how many times I've tried using the website or app and just ended up more confused. Which link was I supposed to click? How do I even describe my symptoms properly in writing? I don't know half these medical terms they're asking. My English is no good. Eventually I just give up and tell my son to figure it out for me instead.*

This account highlights how language barriers and limited digital literacy can significantly undermine patients' ability to independently navigate their care.

The lack of physical presence in video consultations was also cited as a factor reducing patient agency. Mukesh, who experiences complex neurological symptoms, shared the following:

> *I mean doctors have it tough now seeing us all through video calls. But it almost feels pointless for me. He's just staring at notes on some other screen barely listening. I can't really show him what's going on with my body in a genuine way. And to be quite honest it just feels like he's already decided before properly hearing me out.*

This quote underscores how the limitations of video consultations can leave patients feeling unheard and unable to effectively communicate their concerns.

The reliance on family members further eroded the personal agency of participants like Samiyah, who explained the following:

> *At my age, it's impossible to keep track of the different numbers to call or steps to do appointments online. My children handle everything now—they email test results, book consultations, order medications for me. I feel so helpless relying entirely on them, but I don't really have much of a choice.*

This dependence on others for health care management significantly diminishes patients' autonomy and control over their own care.

Some participants, like Mohammad, felt entirely excluded from digital health care due to lack of access to necessary technology:

> *We don't have good internet like that or one of those fancy smartphones they keep saying to use, you see that phone? [points to older Samsung phone with cracked screen] That's what I have and it only works when we can pay for the you know [data]. My nephew always talking about this gadget and that, we just can't afford it. I know they started some phone video service during corona, but it wasn't for me. It feels like they are saying either we have to use this stuff or else we don't deserve to get treatment.*

This quote highlights how socioeconomic factors can create barriers to accessing digital health services, further reducing patient agency.

These 8 themes collectively illustrate the complex and intersecting ways in which the digitalization of health care can exacerbate existing inequities for multiply disadvantaged patients.

Discussion

Principal Findings

Overview

This study provides critical insights into how multiple dimensions of disadvantage intersect to shape experiences with digital health services, particularly video consultations, among older, low-income, individuals with limited English proficiency. The findings revealed a complex interplay of factors that contribute to digital health disparities, extending beyond notions of access and skills to encompass issues of presence, trust, continuity, oppression, complexity, and agency. The analysis revealed that the digitalization of health care can exacerbate existing inequities, erode trust, compound oppression, and reduce patient agency for multiply disadvantaged patient populations. Video consultations often created dynamics of "absent presence," where patients perceived health care providers as distracted or disengaged, leading to a sense of disconnection. The lack of previous in-person rapport negatively shaped patients' perceptions of subsequent video consultations with unfamiliar clinicians, which led to weakened relationships. The shift eroded established continuities of care for some, with the fragmentation of relational continuity becoming apparent. Digital interactions compounded experiences of oppression for patients navigating multiple, intersecting forms of structural disadvantage, such as age, ethnicity, and socioeconomic status. The complexity of navigating multiple digital platforms and pathways led to profound disorientation and fragmentation of care, especially for those with limited digital literacy or language proficiency. Experiences of racism and discrimination within health care settings shaped patients' engagement with digital services, often leading to disengagement and mistrust. The cumulative effect of these factors resulted in a significant reduction in patient agency, particularly for marginalized individuals, undermining their ability to effectively navigate their care and make informed decisions.

The experiences of our participants suggest that digital health technologies, rather than being neutral tools, act as both mirrors and magnifiers of existing social inequalities. This aligns with the concept of the "digital poorhouse" by Eubanks [46], where technology reinforces and exacerbates existing patterns of marginalization. In this study, the shift to video consultations not only reflected existing disparities in health care access but also often amplified them, creating new barriers for those already struggling to navigate the health care system. This finding challenges the often-optimistic rhetoric surrounding digital health innovations [47,48]. While proponents argue that digital technologies can democratize access to health care [49,50], our results suggest a more nuanced reality. For multiply disadvantaged patients, the digitalization of health care services can create a cascade of exclusionary experiences, from difficulties in accessing technology to challenges in effectively communicating health needs in a digital environment.

Extending Intersectionality Theory in Digital Health Contexts

This research extends intersectionality literature by applying its insights to the study of digital health disparities, revealing specific mechanisms through which the increasing digitization of health care creates new forms of inequity and exclusion. The findings highlight how digital access disparities effectively excluded individuals with low digital literacy from video consultation, while limited English proficiency significantly reduced older adults' engagement with health services in the digital space. Importantly, we found that digital competency and digital access do not always go hand in hand, demonstrating how digitalization, while improving health care access for some, simultaneously creates new vectors of exclusion that intersect with and exacerbate existing social inequities. Moreover, while earlier intersectionality research has focused primarily on traditional axes of oppression such as race, class, and gender, this study highlights the emergence of new vectors of disadvantage related to digital access, literacy, capital, and competency. This underscores the need for an expanded understanding of intersectionality that accounts for the growing centrality of digital technologies in shaping health outcomes and experiences.

This study shows that digital health disparities do not operate along independent axes of division but often overlap, interlink, and interact, demonstrating how patterns of dominance and vulnerability intersect to shape people's experiences with digital health care. We argue that inequity and digital health exclusion are relational and occur along multiple fracture lines, which differentiate people's spaces of opportunities, well-being, and level of agency. These disparities are produced and reinforced through complex social relationships and interactions within health care systems and broader societal structures. The COVID-19 pandemic may have brought out new instantiations and shed light on what was less visible before; however, the roots of digital health inequity are deeply entrenched in systems of power and social order.

The Constellation of Challenges for Multiply Disadvantaged Patients

The findings from this study revealed that video consultations created a constellation of challenges for older patients with multiple disadvantages. Participants described feeling profoundly disoriented in digital spaces, struggling to navigate unfamiliar platforms and processes. This disorientation was compounded by disrupted continuity of care, as they cycled between health care providers in fragmented digital encounters. The remote modality also engendered a weak sense of presence and connection—many (13/17, 76%) felt their health care providers were not fully attentive or did not understand their needs. Crucially, the digital interfaces exacerbated feelings of disempowerment and loss of agency. Participants felt adrift, unable to steer the direction of their care. Intersecting barriers reinforced the following: limited economic, social, and linguistic resources; social and cultural isolation; low digital, health, and health care system literacy; and physical impairments of illness and age. Some (9/17, 53%) perceived the digital challenges as yet another form of discrimination. These intersecting factors fed into a cycle of growing disengagement and mistrust toward individual providers and the health care system as a whole.

Digital Capital and Health Equity

Our results also contribute to the emerging literature on digital capital [35] by illustrating how disparities in digital competencies and access intersect with other forms of disadvantage to shape health outcomes. The struggles of our participants to navigate complex digital health systems reflect not only a lack of technical skills but also a broader deficiency in the social and cultural capital needed to effectively engage with digitalized health care. This finding aligns with the theory of capital conversion by Bourdieu [32], suggesting that disadvantages in 1 domain (eg, socioeconomic status) can translate into disadvantages in another (eg, digital health access).

Trust, Presence, and the Digitalization of Care Relationships

Our findings on the erosion of trust and the sense of "absent presence" in video consultations raise important questions about the nature of care relationships in digital environments. Drawing on the work by Giddens [30] on trust in modern societies, we can interpret these experiences as reflective of the disembedding of social relations that occurs with increased digitalization. The loss of physical copresence in health care interactions appears to disrupt established mechanisms for building and maintaining trust, particularly for patients who may already have reasons to distrust health care institutions.

This erosion of trust and presence challenges dominant narratives about the efficiency and convenience of digital health services. While video consultations may offer logistical benefits, our findings suggest they may come at the cost of the relational aspects of care that are particularly important for susceptible patients. This aligns with the critique of the logic of choice in health care by Mol [51], suggesting that the move toward digital health services may prioritize a transactional model of care over a relational one.

Compounding Oppression and Reduced Agency

Perhaps the most concerning finding is that digital health interactions can compound existing forms of oppression and reduce patient agency. This aligns with critical perspectives on technology that view it not as a neutral tool but as a social force that can reinforce existing power structures [52]. In the context of health care, where power imbalances between health care providers and patients are already pronounced [53], the addition of digital interfaces appears to further tilt the scales against marginalized patients. The reduction in patient agency observed in this study has important implications for patient-centered care and shared decision-making, which are increasingly recognized as crucial elements of high-quality health care [54].

Our findings suggest that current approaches to digital health may be undermining these important principles for certain patient populations.

Methodological Contributions

This research also makes significant methodological contributions to intersectionality literature by demonstrating the value of qualitative and narrative-based approaches to studying the lived experiences of marginalized groups. While quantitative approaches have dominated much of the existing research on digital health disparities, our research draws on feminist and critical race theories that emphasize the importance of storytelling and counternarratives as forms of epistemic resistance [55,56]. By using in-depth narrative interviews and persona development as key methodological tools [40], we provide a more nuanced and contextualized understanding of how digital health disparities are experienced and navigated by multiply disadvantaged individuals.

Implications for Practice and Policy

Viewed through the lens of health equity and social justice, our findings suggest that the rapid digitalization of health care risks exacerbating existing health disparities. This aligns with the concept of “digital redlining” proposed by Gilliard and Culik [57], where digital systems create new forms of discriminatory exclusion. The compounding of oppression and reduction in patient agency experienced by our participants raise serious concerns about the potential for digital health technologies to undermine principles of equity and patient-centered care.

However, these findings also point to potential avenues for intervention. Through reconditioning the intersectional nature of digital health disparities, policy makers and health care providers can develop more flexible, targeted approaches to support susceptible patients. This might involve not only addressing technical barriers to access but also working to build the broader forms of capital needed to effectively navigate digital health systems. For example, offering a range of communication options, including in-person visits, and providing additional support for patients navigating digital systems.

In addition, our results underscore the importance of maintaining and strengthening relational continuity in health care, even as care becomes increasingly digitalized. This might involve strategies to ensure patients can maintain relationships with preferred health care providers across digital and in-person interactions. Third, our findings point to the need for greater attention to issues of structural competency in the design and implementation of digital health systems. This could involve training health care providers in culturally competent digital communication and the development of digital health interfaces that are more inclusive and culturally sensitive. Finally, our results suggest that efforts to address digital health disparities must go beyond simply providing access and skills training to address deeper structural inequalities. This aligns with calls for a “digital determinants of health” framework that recognizes the broader social, economic, and political factors shaping digital health equity [58].

Limitations and Future Research

While this study offers valuable insights, several methodological limitations warrant consideration. The relatively small sample size (N=17) and geographical confinement to Redbridge, London, limit the generalizability of our findings. Our reliance on narrative interviews may have introduced recall and social desirability biases, potentially skewing participants’ accounts of their digital health experiences. The inclusion criterion of having attempted ≥1 video consultation may have inadvertently excluded those facing the most severe barriers to digital health access. In addition, our study focused primarily on video consultations, potentially overlooking other forms of digital health interventions.

Looking ahead, several key areas warrant further investigation to address these limitations and expand our understanding of digital health equity. Longitudinal studies are needed to track the long-term impacts of digital health services on marginalized populations, providing insights beyond the snapshot our study offers. Intervention studies should evaluate targeted approaches to address identified barriers, such as building digital capital and fostering trust, which could help overcome some of the access issues noted in our limitations.

To address the geographical limitations of our study, comparative analysis across different health care systems and cultural contexts could identify transferable principles and context-specific challenges in promoting digital health equity. This broader perspective would enhance the generalizability of findings and inform more universally applicable strategies.

From a policy perspective, there is a pressing need to examine how existing health policies and digital strategies impact health disparities. The study analysis could provide valuable context for understanding the systemic factors influencing digital health equity, beyond the individual experiences captured in our study.

Finally, to complement this study’s qualitative insights and address the limitations of our small sample size, future work should focus on developing and validating quantitative measures for intersectional digital health disparities. This would enable population-level tracking and more comprehensive evaluations of digital health interventions, providing a broader evidence base to complement in-depth qualitative studies like our study.

These research directions will contribute to a more nuanced understanding of digital health equity and inform evidence-based strategies for inclusive health care digitalization. By addressing the limitations of this study and expanding the scope of the investigation, future research can build a more comprehensive picture of the challenges and opportunities in promoting equitable access to digital health services.

Conclusions

This study provides critical insights into how the digitalization of health care can deepen disparities for older, low-income, individuals with limited English proficiency. By applying an intersectional lens to the study of digital health inequalities, our research reveals the complex, overlapping, and mutually reinforcing nature of digital exclusion. Our findings underscore the need for intersectional approaches to digital health equity that address the multifaceted nature of disadvantage.

This study makes several unique contributions to the field. First, it extends the application of intersectionality theory to digital health disparities, demonstrating how multiple dimensions of disadvantage interact to shape experiences with digital health services, particularly video consultations. This approach has revealed nuanced insights into how different forms of marginalization compound to create unique barriers to accessing and benefiting from digital health innovations.

Second, our development of theoretically informed user personas [40], grounded in intersectionality and digital capital theories, offers a novel methodological approach for representing the complex lived experiences of multiply disadvantaged patients. These personas provide a powerful tool for humanizing research findings and informing patient-centered service design in digital health contexts.

Finally, by centering the voices of marginalized patients, our research exposes how the rapid shift to video consultations has inadvertently exacerbated existing inequities and eroded trust for some susceptible groups. This challenges prevailing narratives about the universally positive impact of digital health innovations and highlights the need for more nuanced, context-sensitive approaches to digital health implementation.

Furthermore, our findings on the erosion of trust, the sense of “absent presence” in digital consultations, the shift from relational to distributed continuity of care, the weakening of patient-provider relationships, the compounding of oppression, the increased complexity leading to disorientation, the influence of previous experiences of discrimination on engagement, and the reduction in patient agency contribute new insights to the ongoing discourse on the impact of health care digitalization.

In conclusion, as health care continues to digitalize, it is imperative that we remain vigilant to the unintended consequences of technological change and work to ensure that the benefits of innovation are equitably distributed. This will require a fundamental rethinking of how we design, deploy, and evaluate digital health interventions, as well as a renewed commitment to the principles of social justice and health equity.

Acknowledgments

This was an empirical doctoral study by LH, based in UK primary care and linked to the Remote by Default 2 study [59]; the PhD is supported by the Health Foundation’s grant to the University of Cambridge for the Healthcare Improvement Studies Institute. The study was overseen by the independent advisory group of the Remote by Default 2 study.

Conflicts of Interest

None declared.

References

1. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
2. 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]
3. Gann B. Combating digital health inequality in the time of coronavirus. J Consum Health Internet 2020 Sep 01;24(3):278-284. [doi: 10.1080/15398285.2020.1791670]
4. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the digital divide: age and race disparities in use of an inpatient portal. Telemed J E Health 2020 May 01;26(5):603-613 [FREE Full text] [doi: 10.1089/tmj.2019.0065] [Medline: 31313977]
5. Ferguson JM, Goldstein KM, Zullig LL, Zulman DM. Gender differences in adoption and frequency of virtual primary care among men and women veterans. J Womens Health (Larchmt) 2024 Jun;33(6):749-757. [doi: 10.1089/jwh.2023.0781] [Medline: 38629443]
6. Bowleg L. The problem with the phrase women and minorities: intersectionality-an important theoretical framework for public health. Am J Public Health 2012 Jul;102(7):1267-1273. [doi: 10.2105/AJPH.2012.300750] [Medline: 22594719]
7. Weber L, Fore ME. Race, ethnicity, and health: an intersectional approach. In: Vera H, Feagin JR, editors. Handbooks of the Sociology of Racial and Ethnic Relations. Boston, MA: Springer; 2007.
8. Heard E, Fitzgerald L, Wigginton B, Mutch A. Applying intersectionality theory in health promotion research and practice. Health Promot Int 2020 Aug 01;35(4):866-876. [doi: 10.1093/heapro/daz080] [Medline: 31390472]
9. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
10. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SA, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]

11. 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]
12. Husain L, Greenhalgh T, Hughes G, Finlay T, Wherton J. Desperately seeking intersectionality in digital health disparity research: narrative review to inform a richer theorization of multiple disadvantage. J Med Internet Res 2022 Dec 07;24(12):e42358 [FREE Full text] [doi: 10.2196/42358] [Medline: 36383632]
13. NHS long term plan. National Health Service. URL: https://www.longtermplan.nhs.uk/ [accessed 2024-12-23]
14. Remote consulting. NHS England. 2023 Jun 12. URL: https://www.england.nhs.uk/long-read/remote-consulting/ [accessed 2024-11-17]
15. Online and video consultation undertaken in general practice data collection. NHS England. URL: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/online-and-video-consultation-undertaken-in-general-practice [accessed 2024-11-17]
16. Parker RF, Figures EL, Paddison CA, Matheson JI, Blane DN, Ford JA. Inequalities in general practice remote consultations: a systematic review. BJGP Open 2021 Jun;5(3):BJGPO.2021.0040 [FREE Full text] [doi: 10.3399/BJGPO.2021.0040] [Medline: 33712502]
17. van Dijk J, Hacker K. The digital divide as a complex and dynamic phenomenon. Inf Soc 2003 Sep;19(4):315-326. [doi: 10.1080/01972240309487]
18. Qureshi S. Creating a better world with information and communication technologies: health equity. Inf Technol Dev 2016 Jan 22;22(1):1-14. [doi: 10.1080/02681102.2015.1121585]
19. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
20. Lenhart A, Horrigan JB. Re-visualizing the digital divide as a digital spectrum. IT Soc 2003;1(5):23-39.
21. Shahid S, Hogeveen S, Sky P, Chandra S, Budhwani S, de Silva R, et al. Health equity related challenges and experiences during the rapid implementation of virtual care during COVID-19: a multiple case study. Int J Equity Health 2023 Mar 11;22(1):44 [FREE Full text] [doi: 10.1186/s12939-023-01849-y] [Medline: 36906566]
22. Hoffman LC. Reconnecting the patient: why telehealth policy solutions must consider the deepening digital divide. Indiana Health Law Rev 2022 Jul 08;19(2):351-385. [doi: 10.18060/26409]
23. Bathija P, Krupinski EA, Rodriguez JA, Sklar T. Achieving digital health equity by personalizing the patient experience. Telemed Rep 2023 Jun 01;4(1):166-173 [FREE Full text] [doi: 10.1089/tmr.2023.0018] [Medline: 37405125]
24. Zahir A, Yip D, Garcia C, Smith AN, Dhatt Z, Duke M, et al. "I needed for you to see what I'm talking about": experiences with telehealth among homeless-experienced older adults. Gerontol Geriatr Med 2023 May 06;9:23337214231172650 [FREE Full text] [doi: 10.1177/23337214231172650] [Medline: 37168020]
25. Choi NG, DiNitto DM, Marti CN, Choi BY. Telehealth use among older adults during COVID-19: associations with sociodemographic and health characteristics, technology device ownership, and technology learning. J Appl Gerontol 2022 Mar;41(3):600-609 [FREE Full text] [doi: 10.1177/07334648211047347] [Medline: 34608821]
26. Crenshaw K. Demarginalizing the intersection of race and sex: a Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ Chic Leg Forum 1989;1989(1):139-167. [doi: 10.4324/9780429500480-5]
27. Carastathis A. The concept of intersectionality in feminist theory. Philos Compass 2014 Apr 07;9(5):304-314. [doi: 10.1111/phc3.12129]
28. Carter M, Grover V. Me, my self, and I(T): conceptualizing information technology identity and its implications. MIS Q 2015;39(4):931-957. [doi: 10.25300/misq/2015/39.4.9]
29. Whitley EA, Gal U, Kjaergaard A. Who do you think you are? A review of the complex interplay between information systems, identification and identity. Eur J Inf Syst 2019 Jan 28;23(1):17-35. [doi: 10.1057/ejis.2013.34]
30. Giddens A. The Constitution of Society: Outline of the Theory of Structuration. Berkeley, CA: University of California Press; 1984.
31. Ragnedda M, Ruiu ML. Digital Capital: A Bourdieusian Perspective on the Digital Divide. Bingley, UK: Emerald Publishing Limited; 2020.
32. Bourdieu P. The field of cultural production, or: the economic world reversed. Poetics 1983 Nov;12(4-5):311-356. [doi: 10.1016/0304-422x(83)90012-8]
33. Robinson L. A taste for the necessary: a Bourdieuian approach to digital inequality. Inf Commun Soc 2009 Jun;12(4):488-507. [doi: 10.1080/13691180902857678]
34. Ragnedda M, Ruiu ML, Addeo F. Measuring digital capital: an empirical investigation. New Media Soc 2019 Aug 27;22(5):793-816. [doi: 10.1177/1461444819869604]
35. Bannykh G, Kostina S. Measuring digital capital: methodological approaches. KnE Soc Sci 2022 Feb 04:96-103. [doi: 10.18502/kss.v7i2.10282]
36. Clandinin DJ, Connelly FM. Narrative Inquiry: Experience and Story in Qualitative Research. Hoboken, NJ: John Wiley & Sons; 2000.
37. Riessman CK. Narrative Analysis. Thousand Oaks, CA: SAGE Publications; 1993.
38. equityXdesign. URL: https://www.equityxdesign.xyz/ [accessed 2024-12-23]

39. Lincoln YS, Guba EG. Naturalistic Inquiry. Thousand Oaks, CA: SAGE Publications; 1985.
40. Husain L, Finlay T, Husain A, Wherton J, Hughes G, Greenhalgh T. Developing user personas to capture intersecting dimensions of disadvantage in older patients who are marginalised: a qualitative study. Br J Gen Pract 2024 Jan 19;74(741):e250-e257. [doi: 10.3399/bjgp.2023.0412]
41. Gergen KJ. The challenge of absent presence. In: Katz JE, Aakhus M, editors. Perpetual Contact: Mobile Communication, Private Talk, Public Performance. Cambridge, UK: Cambridge University Press; 2002:227-241.
42. Kofod FG, Assing Hvidt E, Arreskov AB, Guassora AD. Interpersonal contact and altered sensory conditions in video consultation - a qualitative interview study in Danish general practice. Scand J Prim Health Care 2024 Dec 16;42(4):650-658 [FREE Full text] [doi: 10.1080/02813432.2024.2376744] [Medline: 39012270]
43. Stern DN. Forms of Vitality: Exploring Dynamic Experience in Psychology, the Arts, Psychotherapy, and Development. Oxford, UK: Oxford University Press; 2010.
44. Bazargan M, Cobb S, Assari S. Discrimination and medical mistrust in a racially and ethnically diverse sample of California adults. Ann Fam Med 2021;19(1):4-15 [FREE Full text] [doi: 10.1370/afm.2632] [Medline: 33431385]
45. Roberts KM, Trejo AN. Provider, heal thy system: an examination of institutionally racist healthcare regulatory practices and structures. Contemp Fam Ther 2022 Jan 28;44(1):4-15 [FREE Full text] [doi: 10.1007/s10591-021-09630-5] [Medline: 35106026]
46. Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York, NY: St. Martin's Publishing Group; 2018.
47. Kim HS, Kwon IH, Cha WC. Future and development direction of digital healthcare. Healthc Inform Res 2021 Apr;27(2):95-101 [FREE Full text] [doi: 10.4258/hir.2021.27.2.95] [Medline: 34015874]
48. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012;2(4):e001574 [FREE Full text] [doi: 10.1136/bmjopen-2012-001574] [Medline: 22815469]
49. Latifi R, Doarn CR, Merrell RC. Telemedicine, Telehealth and Telepresence: Principles, Strategies, Applications, and New Directions. Cham, Switzerland: Springer; 2020.
50. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY: Basic Books; 2019.
51. Pichelstorfer A. The Logic of Care: Health and the Problem of Patient Choice. Milton Park, UK: Routledge; 2008.
52. Feenberg A. Critical theory of technology: an overview. Tailor Biotechnol 2005;1(1):47-64. [doi: 10.5040/9798400670268.ch-001]
53. Nimmon L, Stenfors-Hayes T. The "handling" of power in the physician-patient encounter: perceptions from experienced physicians. BMC Med Educ 2016 Apr 18;16(1):114 [FREE Full text] [doi: 10.1186/s12909-016-0634-0] [Medline: 27091146]
54. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
55. Solórzano DG, Yosso TJ. Critical race methodology: counter-storytelling as an analytical framework for education research. Qual Inquiry 2002 Feb 01;8(1):23-44. [doi: 10.1177/107780040200800103]
56. Medina J. The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations. Oxford, UK: Oxford University Press; 2013.
57. Gilliard C. Quotes from the episode. Teaching in Higher Ed. URL: https://teachinginhighered.com/podcast/digital-redlining-privacy/ [accessed 2024-04-29]
58. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
59. Greenhalgh T, Shaw SE, Alvarez Nishio A, Booth A, Byng R, Clarke A, et al. Protocol: remote care as the 'new normal'? Multi-site case study in UK general practice. NIHR Open Res 2022 Aug 8;2:46 [FREE Full text] [doi: 10.3310/nihropenres.13289.1] [Medline: 37881300]

Abbreviations

IS: information system
NHS: National Health Service

Edited by K Williams; submitted 22.08.24; peer-reviewed by J Woo, J Walumbe; comments to author 07.11.24; revised version received 18.11.24; accepted 26.11.24; published 06.01.25.

Please cite as:
Husain L, Greenhalgh T
Examining Intersectionality and Barriers to the Uptake of Video Consultations Among Older Adults From Disadvantaged Backgrounds With Limited English Proficiency: Qualitative Narrative Interview Study
J Med Internet Res 2025;27:e65690
URL: https://www.jmir.org/2025/1/e65690
doi:10.2196/65690
PMID:

Original Paper

Digital Transformation of Rheumatology Care in Germany: Cross-Sectional National Survey

Susann May[1], MPH; Robert Darkow[2], Dr rer med; Johannes Knitza[3,4], MHBA, MD, PhD; Katharina Boy[1], BSc; Philipp Klemm[5], MD; Martin Heinze[1,6], MD; Nicolas Vuillerme[3,7,8], PhD; Pascal Petit[3], PhD; Patricia Steffens-Korbanka[9], MFA; Heike Kladny[10], MFA; Johannes Hornig[9], MD; Peer Aries[11], MD; Martin Welcker[10], MD; Felix Muehlensiepen[1,3], MPH, Dr rer med, PhD

[1]Center for Health Services Research, Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Rüdersdorf bei Berlin, Germany

[2]FH JOANNEUM Gesellschaft mbH, Graz, Austria

[3]AGEIS, Université Grenoble Alpes, Grenoble, France

[4]Institute for Digital Medicine, University Hospital of Giessen and Marburg, Philipps University Marburg, Marburg, Germany

[5]Department of Rheumatology, Immunology, Osteology, Kerckhoff-Klinik GmbH, Bad Nauheim, Germany

[6]University Clinic for Psychiatry and Psychotherapy, Brandenburg Medical School, Immanuel Hospital Rüdersdorf, Rüdersdorf bei Berlin, Germany

[7]Institut Universitaire de France, Paris, France

[8]LabCom Telecom4Health, Orange Labs & Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP-UGA, Grenoble, France

[9]Rheumapraxis an der Hase, Osnabrück, Germany

[10]MVZ für Rheumatologie Dr. Martin Welcker GmbH, Planegg, Germany

[11]Immunologikum Hamburg, Rheumatologie & Klinische Immunologie, Hamburg, Germany

Corresponding Author:
Felix Muehlensiepen, MPH, Dr rer med, PhD
Center for Health Services Research
Faculty of Health Sciences Brandenburg
Brandenburg Medical School Theodor Fontane
Seebad 82/83
Rüdersdorf bei Berlin, 15562
Germany
Phone: 49 1636055860
Email: Felix.Muehlensiepen@mhb-fontane.de

Abstract

Background: In recent years, health care has undergone a rapid and unprecedented digital transformation. In many fields of specialty care, such as rheumatology, this shift is driven by the growing number of patients and limited resources, leading to increased use of digital health technologies (DHTs) to maintain high-quality clinical care. Previous studies examined user acceptance of individual DHTs in rheumatology, such as telemedicine, video consultations, and mHealth. However, it is essential to conduct cross-technology and continuous analyses of user acceptance and DHT use to maximize the benefits for all relevant stakeholders.

Objective: This study aimed to explore the current acceptance, use, and preferences regarding DHTs among patients in rheumatology care in Germany.

Methods: Rheumatology patients from 3 clinics in Germany were surveyed to understand their perspectives on DHTs. The survey included main themes, including acceptance, preferences, COVID-19's impact, potential, and barriers related to DHTs. The data were analyzed using descriptive statistics and correlation analysis.

Results: Out of 337 participants, 53% (179/337) reported using DHTs. Specific technologies included wearables (72/337, 21%), mHealth apps (71/337, 21%), digital therapeutics (32/337, 9%), electronic prescriptions (30/337, 9%), video consultations (15/337, 4%), and at-home blood self-sampling (3/337, 1%). Nearly two-thirds (220/337, 65%) found DHTs useful, and 69% (233/337) held a generally positive attitude toward DHTs. Attitudes shifted positively during the COVID-19 pandemic for 40% (135/337) of participants. Higher education was more prevalent among DHT users (114/179, 63.7%) compared with nonusers (42/151, 27.8%; *P*=.02). The main potential benefits identified were location-independent use (244/337, 72%) and time-independent use

(216/337, 64%). Key barriers included insufficient user knowledge (165/337, 49%) and limited information on DHTs (134/337, 40%).

Conclusions: Patient acceptance and use of DHTs in rheumatology is increasing in Germany. A prospective, standardized monitoring of digital transformation in rheumatology care is highly needed.

(J Med Internet Res 2025;27:e52601) doi:10.2196/52601

KEYWORDS

telemedicine; digital health technologies; rheumatology; COVID-19; questionnaire; telehealth; eHealth; digital health; survey; rheumatism; Germany; Europe; national; use; experience; attitude; opinion; perception; perspective; acceptance; preference; correlation

Introduction

Health care is undergoing a multidisciplinary digital transformation, which refers to “a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies” [1].

Within the realm of rheumatology care, a notable surge in digital technologies has transpired in recent years [2,3]. This occurrence stems from the existing global disparity between the increasing number of patients with rheumatic and musculoskeletal disease (RMD) [4] and the static or even decreasing availability of human resources in the field of rheumatology care [5,6].

Consequently, Miloslavsky and Bolster [7] have discerned the amplified use of telemedicine as a prospective remedy to ensure the continuity of rheumatological care in the future [7]. Concurrently, and accelerated by the COVID-19 pandemic, the European Alliance of Associations for Rheumatology (EULAR) has recently published points to consider for remote care in RMD [8].

In Germany, a country struggling with a shrinking and aging population and a shortage of health care professionals, digital health technologies (DHTs) in rheumatology care have proven effective in improving rheumatology care [2,9]. Prominent among the prevailing DHTs are video consultations, sensors, wearables, digital health applications, and digital therapeutics with varied objectives (symptom checkers, ePRO [electronic patient-reported outcome] documentation, or patient education), social media platforms or messenger platforms [10], which can also be combined with home-based self-sampling [11-13].

Past studies have assessed patients’ and physicians’ acceptance of DHTs in the domain of rheumatology within Germany. These investigations primarily centered on individual digital modalities, such as telemedicine [13,14], mHealth [15], and video consultations [16,17]. Following the onset of the COVID-19 pandemic, the influence of this global crisis on the adoption of DHTs was also scrutinized [3,17]. It remains essential to continuously analyze the digital transformation in rheumatology to maximize the benefits for all relevant stakeholders. In particular, the patient perspective is critical to ensuring that DHTs actually improve health care, meets patient needs, and builds patient trust in the use of DHTs. Thus, the aim of this study was to explore the current acceptance, use, and preferences regarding DHTs among patients in rheumatology care in Germany.

Methods

Overview

To explore the patients’ perspectives and preferences regarding digital health systems in rheumatology care, we recruited rheumatology patients in 3 outpatient clinics. Inclusion criteria were being aged ≥18 years, literate in German, having the physical and mental ability to fill out a paper-pencil questionnaire, and having a diagnosis of RMD. The survey was conducted between February 2023 and April 2023. The 3 clinics were sent 125 questionnaires with the request to distribute the questionnaires to eligible patients.

The questionnaire was created based on a literature review and the results of a qualitative study with health care providers (HCPs) and nurses. The questionnaire addressed the following topics: acceptance of DHTs, use of DHTs, preferences for DHTs, the impact of COVID-19 on digital transformation, potential benefits, and barriers regarding DHTs, and sociodemographic characteristics. The survey was pilot-tested with 10 patients to detect necessary formatting and wording changes. Minor revisions were made accordingly. Please refer to Multimedia Appendix 1 for the German version of the questionnaire and Multimedia Appendix 2 for the translated version in English.

Data were analyzed using quantitative descriptive analysis and correlation analysis supported by SPSS (version 23.0; IBM). For correlation analyses, the data were converted into scores, and the Spearman rank correlation coefficient was calculated. The α error distribution was chosen to be 2-sided, as an undirected correlation is assumed to be open to hypotheses. These correlations had a correlation coefficient higher than 0.29.

Ethical Considerations

This study was approved by the ethics committee of Brandenburg Medical School (E-02-20211028). All patients provided informed consent. The rheumatology clinics received an expense allowance of €150 (US $ 158) for the distribution and return of the questionnaires. The study data are anonymous. The questionnaires were digitized and then stored at the Center for Mental Health at Brandenburg Medical School.

Results

Patient Characteristics

In total, 337 patients with RMD completed the questionnaire. The number of patients rejecting participation was not measured. The study sample's demographics are shown in Table 1. The mean age was 52.5 (SD 14.2) years; 64% (219/337) were female, 32% (108/337) were male, 1% were (3/337) nonbinary, and 2% (7/337) did not give any information on gender. On average, participants had been receiving rheumatologic treatment for a median 8 years (mean 9.35 years, SD 7.63; range: 1 month-399 months), and the median time with the diagnosis was 8.9 years (mean 10.2 years, SD 8.33; range: 3 months-579 months).

Table 1. Characteristics of participants.

Characteristics	Participants (N=337), n (%)
Sex	
Female	219 (65)
Male	108 (32)
Nonbinary	3 (1)
No data	7 (2)
Diagnosis	
Axial spondyloarthritis	13 (4)
Fibromyalgia	7 (2)
Psoriatic arthritis	40 (12)
Rheumatoid arthritis	182 (54)
Sjörgren syndrome	10 (3)
Spondyloarthritis	21 (6)
Systemic lupus erythematosus	9 (3)
Other	39 (12)
No data	16 (5)
Education level	
>10 years of school completed	156 (46)
9-10 years of school completed	163 (48)
Student	2 (1)
No formal schooling	2 (1)
No data	14 (4)
Vocational training	
Apprenticeship	188 (56)
Bachelor's, master's, or magister's degree, diploma	109 (3)
Started an apprenticeship	6 (2)
Without apprenticeship	11 (3)
No data	23 (7)
Place of residence (number of inhabitants)	
Rural region (<5000)	70 (21)
Small city (>5000-20,000)	73 (22)
Medium city (>20,000-100,000)	63 (19)
Large city (>100,000-1,000,000)	28 (8)
Metropolis city (>1,000,000)	93 (28)
No data	10 (3)

Attitudes Toward DHTs

About half of the patients used DHTs overall (179/337, 53%) (Table 2). Overall, most patients (220/337, 65%) rated DHTs as useful, while a third remained neutral in their assessment of usefulness. Almost two-thirds of the patients (233/337, 69%) were positive or rather positive toward DHTs. Among all participants, 55% (185/377) stated that their attitude toward DHTs had not changed due to the COVID-19 pandemic. However, 40% (135/337) mentioned that their attitude had become more positive.

Table 2. Attitudes toward digital health technologies.

Items	Participants (N=337), n (%)
Do you use digital health technologies?	
Yes	179 (53)
No	151 (45)
No data	7 (2)
Digital health technologies are useful.	
Strongly agree	99 (29)
Agree	121 (36)
Neutral	99 (29)
Disagree	5 (1)
Strongly disagree	7 (2)
No data	6 (2)
How do you rate your attitude towards digital health services?	
Positive	118 (35)
Rather positive	115 (34)
Neutral	77 (23)
Rather negative	17 (5)
Negative	4 (1)
No data	6 (2)
Did you change your attitude due to the COVID-19 pandemic?	
Yes, the attitude has become more positive	135 (40)
Yes, the attitude has become more negative	8 (2)
No	185 (55)
No data	9 (3)

Use of DHTs

Almost half of the patients (143/337, 42%) were already using email for communication with physicians before the COVID-19 pandemic (Table 3). Currently, this value has increased to 58% (195/337). The number of patients using video consultations was very low: 4% (15/337). However, 14% (48/337) stated that they would use video consultations in the future. We found that 9% (32/337) were using reimbursed digital health applications (German: Digitale Gesundheitsanwendung; DiGA). Overall, 21% (71/337) were using other mobile health apps. Only 1% (3/337) had used blood self-sampling. Almost a quarter (78/337, 23%) stated that they wanted to use electronic prescriptions in the future.

Table 3. Digital health technology use.

	Used before COVID-19, n (%)	I currently use, n (%)	I will use in the future, n (%)	I am not interested, n (%)	I don't use, n (%)	I don't know, n (%)
Video consultation	7 (2)	15 (4)	48 (14)	14 (4)	191 (57)	53 (16)
Digital therapeutics (DiGA[a])	14 (4)	32 (9)	38 (11)	10 (3)	154 (46)	86 (26)
mHealth apps (not DiGA)	19 (6)	71 (21)	40 (12)	10 (3)	152 (45)	54 (16)
Wearables	19 (6)	72 (21)	21 (6)	19 (6)	175 (52)	43 (13)
Blood self-sampling at home	2 (1)	3 (1)	9 (3)	50 (15)	187 (55)	99 (29)
Electronic prescription	7 (2)	30 (9)	78 (23)	5 (1)	147 (44)	75 (22)

[a] DiGA: digital health application.

Potential Benefits of and Barriers to DHTs

Independence in terms of location (244/337, 72%) and time (216/337, 64%) and generally increased flexibility (170/337, 50%) were cited most often by patients as potential benefits of digital health. The detailed results are displayed in Table 4.

The biggest barrier at present is the lack of information. Thus, 49% (165/337) stated that they do not have sufficient knowledge, and 40% (134/337) stated that they are not sufficiently informed about DHTs. A third (118/337, 35%) stated that technical equipment is a barrier to using the tools.

Table 4. Potential benefits and barriers of digital health technologies.

Items	Participants (N=337), n (%)
Potential benefits	
Location independent use	244 (72)
Time-independent use	216 (64)
Detailed documentation of the course of the disease	139 (41)
Cost savings	134 (40)
More options for accessing information, diagnostics, and therapy	125 (37)
Accessibility	85 (25)
More flexibility	170 (50)
Better preparation for the physician-patient consultation	139 (41)
Needs-based care	68 (20)
No potential benefits	32 (9)
Barriers	
Limited information about digital health services	134 (40)
Insufficient evidence of the benefits of the offers	39 (12)
Poor quality of current offers	32 (9)
Gaps in data protection	92 (27)
Lack of user-friendliness	84 (25)
Lack of accessibility	12 (4)
High costs	12 (4)
Lack of technical equipment (eg, poor Internet connection, old end devices)	118 (35)
Lack of knowledge among users	165 (49)
No need because satisfied with the current analogue solutions	57 (17)

Correlation Analyses

While correlation analysis revealed some relationships between content items (Table 5), no significant correlations between the content items and sociodemographic data were observed.

Table 5. Results of the correlation analyses.

Content items	Spearman ρ	*P* values (2-tailed)
Current use		
Positive attitude	0.458	<.001
Number of months diagnosed		
Number of months in rheumatology treatment	0.777	<.001
Potential benefits		
Location-independent use		
Detailed documentation of disease progression	0.355	<.001
Accessibility	0.328	<.001
More flexibility	0.331	<.001
No	–0.525	<.001
Detailed documentation of disease progression		
More options for accessing information, diagnostics, and therapy	0.330	<.001
More flexibility	0.336	<.001
Better preparation for the doctor-patient consultation	0.400	<.001
Needs-based care	0.330	<.001
Cost savings		
Accessibility	0.380	<.001
More flexibility	0.320	<.001
Needs-based care	0.302	<.001
More possibilities to access information, diagnostics, therapy		
Accessibility	0.346	<.001
More flexibility	0.405	<.001
Better preparation for the doctor-patient consultation	0.305	<.001
Accessibility		
More flexibility	0.357	<.001
Needs-based care	0.338	<.001
More flexibility		
Needs-based care	0.380	<.001
None	–0.307	<.001
Better preparation for the doctor-patient consultation		
Needs-based care	0.360	<.001
Barriers		
Lack of technical equipment		
Lack of knowledge among users	0.307	<.001

Discussion

To explore the current patient acceptance, use, and preferences for DHTs in rheumatology care, we performed a paper-pencil questionnaire survey among patients with RMD in Germany.

Principal Findings

More than half (179/337, 53%) of 337 patients reported that they use DHTs. Overall, 21% (72/337) used wearables, 21% (71/337) used mHealth apps, 9% (32/337) used digital therapeutics (DiGA), 9% (30/337) used electronic prescriptions, 4% (15/337) used video consultations, and 1% (3/337) used at-home blood self-sampling. Nearly two-thirds of the patients with RMD (220/337, 65%) rated DHTs as useful. While 69% (233/337) reported a generally positive attitude toward DHTs, about 40% (135/337) mentioned their attitudes became more positive due to the COVID-19 pandemic. The main potential benefits of DHTs reported by the patients were location-independent use (244/337, 72%) and time-independent use (216/337, 64%). The main barriers included insufficient knowledge among users (165/337, 49%) and limited information about digital health services (134/337, 40%).

Comparison With Previous Work

Our findings are aligned with previous studies that have examined patients' acceptance of DHTs in rheumatology [3,14-19]. Comparing our results with previous findings, clear trends emerge that further underscore the growing acceptance of DHTs among patients with RMD.

In a survey conducted from September 2019 to December 2019 among 766 German patients with RMD [18], only 51% (364/718) of participants indicated familiarity with the term "telemedicine." A mere 30% (210/690) expressed intentions to try telemedicine in the future, and a total of 21% (139/663) expressed a desire for their rheumatologist to offer telemedicine. In this study, conducted from February 2023 to April 2023, that is, after the COVID-19 period, 53% (179/337) of participants reported using DHTs. It should be noted, however, that while the terms telemedicine and DHTs are closely related, they are not synonymous.

In the 2019 survey, 0.3% of patients with RMD reported having experienced a video consultation with a physician. The current results reveal a 4% (15/337) use rate of video consultations, still representing a modest figure, suggesting that video consultations in rheumatology care remain an exceptional practice. These findings corroborate the outcomes observed by Richter et al [17]. In a survey of rheumatologists in Germany, 27% (55/205) reported offering video consultations during COVID-19 lockdown phases, with the frequency of provided video conferences diminishing as the pandemic progressed.

Knitza et al [15] reported that in 2018/2019, most patients with RMD (68%) believed that using medical apps could be beneficial for their own health [15]. However, out of 193 patients, only 8 (4%) were currently using medical apps. In the fall of 2020, Kernder et al [3] explored patient and rheumatologist attitudes toward digital technologies, particularly digital health applications. Even a higher rate of patients (222/299, 74%) and also rheumatologists (98/129, 76%) believed that digital health apps were valuable for managing RMDs. Compared with Knitza et al [15], our results reveal that digital health app use increased notably (71/337, 21%). In our study, a distinction was made between certified and prescribed digital therapeutics (DiGA) and other nonreimbursed digital health apps. Interestingly, patients reported less use of prescribed DiGAs (32/337, 9%) compared with other noncertified mHealth apps (21%). This discrepancy could be attributed to the absence of DiGAs explicitly tailored for RMDs, whereas non-DiGA mHealth apps fill this gap. A first pilot study evaluating DiGA use in rheumatology [20] revealed high patient acceptance and some clinical benefits, yet poor adherence as a major limiting barrier.

According to Kernder et al [3], 38% (112/299) of patients reported a positive change in attitude due to COVID-19, comparable to 40% (135/337) in this study. The most commonly cited advantages of DHAs were their independence from time and place, which were also expressed by participants in our study.

Our study also inquired about the use and acceptance of blood self-sampling, a prospect that has recently gained significance in future rheumatological care [11-13]. However, our results suggest that self-sampling currently remains largely confined to research settings.

In a recent secondary data analysis [21], we demonstrated that, specifically, older patients with RMD residing in rural areas, who could potentially benefit from telemedicine, currently lack the motivation to embrace it and appear to require additional support. However, these relationships were not confirmed by this study. Nevertheless, our data reaffirm the profound relevance of knowledge in the use of digital technologies in rheumatology care. In line with earlier findings [3,14,19,21], our participants identified "lack of knowledge among users" as the main barrier to use. Besides the patient perspective, the viewpoint of HCPs is also pertinent to the implementation of digital technologies. Lack of knowledge among rheumatologists was previously identified as a major barrier to implementing ePROs in routine care [22]. Conveniently, recent surveys also indicated a positive inclination of HCPs toward digital technologies in rheumatology [3,14,16,19].

Limitations

Our study has several limitations. Despite our best efforts, it is possible that our survey did not capture all important emerging technologies relevant to rheumatology. In addition, the terminology used in this survey for DHTs may have caused confusion among participants, potentially leading to information bias. In addition, our survey reached a selected population from 3 outpatient clinics. This also applies to the selection of study participants, as we assume that individuals with a specific interest in digital health were more likely to have completed our questionnaire. A strength of previous surveys [3,16,17,19,22] is that the survey was paper-pen–based (instead of digital) to minimize selection bias.

Finally, it is essential to acknowledge that numerous research groups are focusing on measuring the use and acceptance in the domain of digital rheumatology. However, these studies often use varying approaches and terminologies and refer to different study populations. Therefore, the comparability of data across studies for the purpose of continuous monitoring of the digital transformation in rheumatology remains limited.

Implications

Considering the heterogeneity regarding digital rheumatology surveys and fast transformation, we advocate for a standardized, regular, survey-based monitoring of the digital transformation in rheumatology from the perspectives of both patients with RMD and HCPs. Ideally, this monitoring should be conducted on an international level, including dedicated societies such as EULAR and the Digital Rheumatology Network. Furthermore, the specific digital technologies, their nomenclature, and other questionnaire content should be harmonized through Delphi studies in collaboration with an international expert board involving input from patients.

The results depict high acceptance regarding DHTs, which is currently limited mainly by a lack of knowledge. Dedicated education for patients with RMD and the treating HCPs is necessary to foster implementation in routine clinical practice.

Conclusions

The digital transformation in rheumatology care in Germany is progressing. Patient acceptance and use is increasing. This provides hope that, despite the rising burden of disease and stagnating human resources, digital health can continue to ensure high-quality care for patients with RMD in the future. A prospective, standardized monitoring of digital transformation in rheumatology care is highly needed.

Acknowledgments

The authors would like to thank all participants of this study. We also owe special gratitude to the 3 outpatient clinics and their teams that supported this study and distributed the questionnaires.

This study was conducted in cooperation with RheumaDatenRhePort GbR (RHADAR) and uses data from a health care analysis commissioned by AbbVie Deutschland GmbH & Co KG.

Authors' Contributions

FM and SM had full access to all the data in the study and take responsibility for the integrity of the data and accuracy of the data analysis. FM, SM, MW, PA, and JH conceptualized and designed the study. FM, SM, MW, PA, JH, PSK, HK, JK, and PK were involved in the acquisition of data. FM, SM, RD, JK, KB, MH, NV, and PP were involved in the analysis and interpretation of data. All the authors were involved in drafting the manuscript and critically revising it for important intellectual content, and they approved the final version of the manuscript to be submitted for publication.

Conflicts of Interest

MW reports support and grants for projects, talks, and discussions by Abbvie, BMS, Boehringer, Galapagos, Gilead, GSK, Hexal, Janssen, Novartis, Pfizer, and UCB. MW is also a member of RHADAR (RheumaDatrenRhePort).

Multimedia Appendix 1
Original questionnaire (German).
[PDF File (Adobe PDF File), 367 KB - jmir_v27i1e52601_app1.pdf]

Multimedia Appendix 2
Questionnaire - English translation.
[PDF File (Adobe PDF File), 365 KB - jmir_v27i1e52601_app2.pdf]

References

1. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. J Bus Res 2021;123:557-567 [FREE Full text] [doi: 10.1016/j.jbusres.2020.10.030]
2. Mühlensiepen F, Kurkowski S, Krusche M, Mucke J, Prill R, Heinze M, et al. Digital health transition in rheumatology: a qualitative study. Int J Environ Res Public Health 2021;18(5):2636 [FREE Full text] [doi: 10.3390/ijerph18052636] [Medline: 33807952]
3. Kernder A, Morf H, Klemm P, Vossen D, Haase I, Mucke J, et al. Digital rheumatology in the era of COVID-19: results of a national patient and physician survey. RMD Open 2021;7(1):e001548 [FREE Full text] [doi: 10.1136/rmdopen-2020-001548] [Medline: 33622673]
4. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: a systematic analysis of the World Health Organization Burden of Diseases database. Ann Rheum Dis 2019;78(6):844-848. [doi: 10.1136/annrheumdis-2019-215142] [Medline: 30987966]
5. Al Maini M, Adelowo F, Al Saleh J, Al Weshahi Y, Burmester GR, Cutolo M, et al. The global challenges and opportunities in the practice of rheumatology: white paper by the World Forum on Rheumatic and Musculoskeletal Diseases. Clin Rheumatol 2015;34(5):819-829 [FREE Full text] [doi: 10.1007/s10067-014-2841-6] [Medline: 25501633]
6. Ward IM, Schmidt TW, Lappan C, Battafarano DF. How critical is tele-medicine to the rheumatology workforce? Arthritis Care Res (Hoboken) 2016;68(10):1387-1389 [FREE Full text] [doi: 10.1002/acr.22853] [Medline: 26866514]
7. Miloslavsky EM, Bolster MB. Addressing the rheumatology workforce shortage: a multifaceted approach. Semin Arthritis Rheum 2020;50(4):791-796 [FREE Full text] [doi: 10.1016/j.semarthrit.2020.05.009] [Medline: 32540672]
8. de Thurah A, Bosch P, Marques A, Meissner Y, Mukhtyar CB, Knitza J, et al. 2022 EULAR points to consider for remote care in rheumatic and musculoskeletal diseases. Ann Rheum Dis 2022;81(8):1065-1071. [doi: 10.1136/annrheumdis-2022-222341] [Medline: 35470160]

9. Knitza J, Callhoff J, Chehab G, Hueber A, Kiltz U, Kleyer A, et al. [Position paper of the Commission on Digital Rheumatology of the German Society of Rheumatology: tasks, targets and perspectives for a modern rheumatology]. Z Rheumatol 2020;79(6):562-569. [doi: 10.1007/s00393-020-00834-y] [Medline: 32651681]
10. Kataria S, Ravindran V. Digital health: a new dimension in rheumatology patient care. Rheumatol Int 2018;38(11):1949-1957. [doi: 10.1007/s00296-018-4037-x] [Medline: 29713795]
11. Zarbl J, Eimer E, Gigg C, Bendzuck G, Korinth M, Elling-Audersch C, et al. Remote self-collection of capillary blood using upper arm devices for autoantibody analysis in patients with immune-mediated inflammatory rheumatic diseases. RMD Open 2022;8(2):e002641 [FREE Full text] [doi: 10.1136/rmdopen-2022-002641] [Medline: 36104118]
12. Muehlensiepen F, May S, Zarbl J, Vogt E, Boy K, Heinze M, et al. At-home blood self-sampling in rheumatology: a qualitative study with patients and health care professionals. BMC Health Serv Res 2022;22(1):1470 [FREE Full text] [doi: 10.1186/s12913-022-08787-5] [Medline: 36461025]
13. Mucke J, Knitza J, Muehlensiepen F, Grahammer M, Stenzel R, Simon D, et al. TELERA-asynchronous TELEmedicine for patients with rheumatoid arthritis: study protocol for a prospective, multi-center, randomized controlled trial. Front Med (Lausanne) 2021;8:791715 [FREE Full text] [doi: 10.3389/fmed.2021.791715] [Medline: 34966765]
14. Muehlensiepen F, Knitza J, Marquardt W, Engler J, Hueber A, Welcker M. Acceptance of telerheumatology by rheumatologists and general practitioners in Germany: nationwide cross-sectional survey study. J Med Internet Res 2021;23(3):e23742 [FREE Full text] [doi: 10.2196/23742] [Medline: 33690147]
15. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]
16. Vossen D, Knitza J, Klemm P, Haase I, Mucke J, Kernder A, et al. [Acceptance of video consultation among patients with inflammatory rheumatic diseases depends on gender and location: results of an online survey among patients and physicians]. Z Rheumatol 2023;82(2):108-113 [FREE Full text] [doi: 10.1007/s00393-021-01052-w] [Medline: 34448915]
17. Richter JG, Chehab G, Reiter J, Aries P, Muehlensiepen F, Welcker M, et al. Evaluation of the use of video consultation in German rheumatology care before and during the COVID-19 pandemic. Front Med (Lausanne) 2022;9:1052055 [FREE Full text] [doi: 10.3389/fmed.2022.1052055] [Medline: 36507506]
18. Muehlensiepen F, Knitza J, Marquardt W, May S, Krusche M, Hueber A, et al. Opportunities and barriers of telemedicine in rheumatology: a participatory, mixed-methods study. Int J Environ Res Public Health 2021;18(24):13127 [FREE Full text] [doi: 10.3390/ijerph182413127] [Medline: 34948737]
19. Becker C, Diener M, Hueber A, Henes J, Krusche M, Ignatyev Y, et al. Unmet information needs of patients with rheumatic diseases: results of a cross-sectional online survey study in Germany. Int J Environ Res Public Health 2022;19(12):7071 [FREE Full text] [doi: 10.3390/ijerph19127071] [Medline: 35742318]
20. Labinsky H, Gupta L, Raimondo MG, Schett G, Knitza J. Real-world usage of digital health applications (DiGA) in rheumatology: results from a German patient survey. Rheumatol Int 2023;43(4):713-719 [FREE Full text] [doi: 10.1007/s00296-022-05261-7] [Medline: 36543961]
21. Muehlensiepen F, Petit P, Knitza J, Welcker M, Vuillerme N. Factors associated with telemedicine use among patients with rheumatic and musculoskeletal disease: secondary analysis of data from a German nationwide survey. J Med Internet Res 2023;25:e40912 [FREE Full text] [doi: 10.2196/40912] [Medline: 36705950]
22. Krusche M, Klemm P, Grahammer M, Mucke J, Vossen D, Kleyer A, et al. Acceptance, usage, and barriers of electronic patient-reported outcomes among German rheumatologists: survey study. JMIR Mhealth Uhealth 2020;8(7):e18117 [FREE Full text] [doi: 10.2196/18117] [Medline: 32390592]

Abbreviations

DHT: digital health technology
DiGA: digital health application
ePRO: electronic patient-reported outcome
EULAR: European Alliance of Associations for Rheumatology
HCP: health care provider
mHealth: mobile health
RMD: rheumatic and musculoskeletal disease

Edited by N Cahill; submitted 09.09.23; peer-reviewed by M de Zwaan, S Góngora Alonso; comments to author 18.04.24; revised version received 28.05.24; accepted 09.10.24; published 06.01.25.

Please cite as:
May S, Darkow R, Knitza J, Boy K, Klemm P, Heinze M, Vuillerme N, Petit P, Steffens-Korbanka P, Kladny H, Hornig J, Aries P, Welcker M, Muehlensiepen F
Digital Transformation of Rheumatology Care in Germany: Cross-Sectional National Survey
J Med Internet Res 2025;27:e52601
URL: https://www.jmir.org/2025/1/e52601
doi:10.2196/52601
PMID:

Original Paper

Application of Large Language Models in Medical Training Evaluation—Using ChatGPT as a Standardized Patient: Multimetric Assessment

Chenxu Wang[1,2,3], BSc; Shuhan Li[1,2], BSc; Nuoxi Lin[1,2], BSc; Xinyu Zhang[3], BSc; Ying Han[3], NP, MSc; Xiandi Wang[3], MD; Di Liu[1,2,4], PhD; Xiaomei Tan[2], PhD; Dan Pu[3], MD; Kang Li[1,2,4*], PhD; Guangwu Qian[1,5*], PhD; Rong Yin[1,2,4*], PhD

[1]West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China
[2]Department of Industrial Engineering, Pittsburgh Institute, Sichuan University, Chengdu, China
[3]Department of Medical Simulation Center, West China Hospital, Sichuan University, Chengdu, China
[4]Med-X Center for Informatics, Sichuan University, Chengdu, China
[5]Department of Computer Science, Pittsburgh Institute, Sichuan University, Chengdu, China
[*]these authors contributed equally

Corresponding Author:
Kang Li, PhD
West China Biomedical Big Data Center
West China Hospital
Sichuan University
No.37, Guoxue Lane, Wuhou District
Chengdu, 610041
China
Phone: 86 02881739902
Email: likang@wchscu.cn

Abstract

Background: With the increasing interest in the application of large language models (LLMs) in the medical field, the feasibility of its potential use as a standardized patient in medical assessment is rarely evaluated. Specifically, we delved into the potential of using ChatGPT, a representative LLM, in transforming medical education by serving as a cost-effective alternative to standardized patients, specifically for history-taking tasks.

Objective: The study aims to explore ChatGPT's viability and performance as a standardized patient, using prompt engineering to refine its accuracy and use in medical assessments.

Methods: A 2-phase experiment was conducted. The first phase assessed feasibility by simulating conversations about inflammatory bowel disease (IBD) across 3 quality groups (good, medium, and bad). Responses were categorized based on their relevance and accuracy. Each group consisted of 30 runs, with responses scored to determine whether they were related to the inquiries. For the second phase, we evaluated ChatGPT's performance against specific criteria, focusing on its anthropomorphism, clinical accuracy, and adaptability. Adjustments were made to prompts based on ChatGPT's response shortcomings, with a comparative analysis of ChatGPT's performance between original and revised prompts. A total of 300 runs were conducted and compared against standard reference scores. Finally, the generalizability of the revised prompt was tested using other scripts for another 60 runs, together with the exploration of the impact of the used language on the performance of the chatbot.

Results: The feasibility test confirmed ChatGPT's ability to simulate a standardized patient effectively, differentiating among poor, medium, and good medical inquiries with varying degrees of accuracy. Score differences between the poor (74.7, SD 5.44) and medium (82.67, SD 5.30) inquiry groups ($P<.001$), between the poor and good (85, SD 3.27) inquiry groups ($P<.001$) were significant at a significance level (α) of .05, while the score differences between the medium and good inquiry groups were not statistically significant ($P=.16$). The revised prompt significantly improved ChatGPT's realism, clinical accuracy, and adaptability, leading to a marked reduction in scoring discrepancies. The score accuracy of ChatGPT improved 4.926 times compared to unrevised prompts. The score difference percentage drops from 29.83% to 6.06%, with a drop in SD from 0.55 to 0.068. The performance of the chatbot on a separate script is acceptable with an average score difference percentage of 3.21%. Moreover, the performance differences between test groups using various language combinations were found to be insignificant.

Conclusions: ChatGPT, as a representative LLM, is a viable tool for simulating standardized patients in medical assessments, with the potential to enhance medical training. By incorporating proper prompts, ChatGPT's scoring accuracy and response realism significantly improved, approaching the feasibility of actual clinical use. Also, the influence of the adopted language is nonsignificant on the outcome of the chatbot.

(J Med Internet Res 2025;27:e59435) doi:10.2196/59435

KEYWORDS

ChatGPT; artificial intelligence; standardized patient; health care; prompt engineering; accuracy; large language models; performance evaluation; medical training; inflammatory bowel disease

Introduction

Background

In recent decades, large language models (LLMs) have experienced significant advancements [1,2]. LLMs are artificial intelligence (AI) models designed to comprehend and process natural human language, as well as generate it [1,3]. Concurrent with the progression of AI, LLMs have exhibited substantial potential in executing tasks involving natural language processing [4]. LLM applications range from article synthesis to summarization to patient diagnosis, illustrating their flexibility in providing valuable assistance [5-9]. Among the various types of LLMs such as the generalist language model, Flamingo, and Minerva [10], ChatGPT stands out as one of the critical milestone models, which is the primary focus of this study.

ChatGPT, developed by OpenAI, a US-based company, represents a significant advancement in language models [11,12]. This model facilitates user interaction with follow-up questions and is fine-tuned for controlled output [13,14]. Additionally, it can be integrated into custom applications via an application programming interface, allowing developers to craft chatbots and virtual assistants with tailored behavior [13]. Built on GPT architecture, ChatGPT currently comprises 2 primary variants, GPT-3.5 and GPT-4. A key characteristic of ChatGPT is that its output is significantly influenced by the input prompt and can be fine-tuned by modifying this prompt. Consequently, prompt engineering is essential for optimizing ChatGPT's performance [15]. Despite ChatGPT and other LLMs, as well as the broader field of AI, being in developmental stages, they have demonstrated great potential in assisting human activities and may substitute human in certain tasks due to their low cost and high efficiency [16]. In the realm of medical education, traditional tests using standardized patients are no exception, involving substantial human labor and training costs [17-20].

The concept of a standardized patient, initially introduced by Barrows and Abrahamson [21], has become a widely recognized method in medical training and assessment [22,23]. In this context, "standardized" implies that the patient in a standardized patient scenario is trained to consistently portray a specific set of symptoms, medicine allergies, and medical history. Here, "patient" refers to an individual acting as if seeking medical care, unlike real patients, these are trained doctors or individuals simulating the diagnosis process to access medical professionals [23,24]. Consequently, a standardized patient is an individual trained to simulate a real patient, accurately portraying a set of symptoms or conditions. Medical students or doctors interact with standardized patients to practice and evaluate their clinical skills [22]. A critical aspect of this procedure is conversing with the standardized patient to obtain essential information for diagnosis, which is known as history-taking tasks. As the entire training process for an SP is both time-consuming and expensive, rural hospitals may even fail to provide such medical training due to poor access to high-quality health care services [25-27]. Our research aims to fill this research gap by applying LLMs, such as ChatGPT, in the assessment process involving standardized patients to tackle these problems. A critical factor influencing ChatGPT's response is the prompt input given to the model [28,29]. Nevertheless, ChatGPT's accuracy and precision require enhancement for medical applications [30,31]. Therefore, this study explored a novel approach to standardized patients in medical training evaluation by substituting human standardized patients with ChatGPT, as well as maintaining relatively high accuracy and precision. The study is conducted along 2 distinct dimensions. Initially, the capability and feasibility of ChatGPT functioning as a standardized patient are investigated. Subsequently, this study explores prompt engineering and revisions to enhance ChatGPT's performance. Finally, this study examines the differences in scores and wording between original and revised prompt (RP) results generated by ChatGPT.

Objectives

This study applied an exploratory approach to standardized patients in medical training evaluation by substituting human standardized patients with ChatGPT, as well as maintaining relatively reasonable accuracy and precision. The main objective can be divided into 2 stages of experiment. Initially, the capability and feasibility of ChatGPT functioning as a standardized patient are investigated. Subsequently, exploring prompt engineering and revisions to enhance ChatGPT's performance, and also examining the differences in scores and wording between the original prompt (OP) and RP results generated by ChatGPT.

Methods

Standardized Patient Resource

The West China Medical Simulation Center of West China Hospital of Sichuan University has released a comprehensive standardized patient training script, encompassing a scenario-based conversational analysis and a specific performance criterion for participants, detailed in Table 1. This criterion emphasizes a procedure-oriented approach, where interactions with the standardized patient are critically evaluated and graded. The criterion comprises 2 principal aspects for evaluation: inquiry skills and humanistic care. The inquiry skills evaluation encompasses 4 dimensions, conversation arrangement, type of question, verifications, and use of professional jargon. In the domain of humanistic care, 2 dimensions are considered: speech and amiable behavior. Each dimension has 5 tiers, from tier 1 to tier 5, with ascending levels of performance. Comprehensive judgment criteria are illustrated in Table 1. ChatGPT exhibits limitations concerning criteria based on subjective judgment, as it lacks emotional response and is not fully adaptable to subjective grading. Therefore, this study focuses solely on the objective grading criteria. For example, an evaluation point could be whether the user inquired about the medicine history or not. Such criteria can be objectively assessed based on the conversation.

Table 1. Standardized patient consultation skills grading and scoring criteria. This scoring sheet can be used in standardized patient assessments to quantitatively evaluate the performance of candidates, where 5= best performance.

Skills	Ranking tiers				
	5-tier	4-tier	3-tier	2-tier	1-tier
Inquiry skills					
Conversation arrangement	The beginning, middle, and end of the consultation are clear and precise, with questions asked in an orderly manner.	Between 5-point and 3-point	Most of the consultation is conducted in an orderly fashion, but the beginning and ending are not clearly defined.	Between 3-point and 1-point	The consultation lacks coherence and organization.
Question types	Reasonable use of open-ended or closed-ended questions.	Between 5-point and 3-point	No open-ended questions, directly asking with closed-ended questions.	Between 3-point and 1-point	Frequently uses sequential and leading questions.
Verifications	Conduct a comprehensive and thorough verification and reference.	Between 5-point and 3-point	The verification and reference are incomplete and not sufficient.	Between 3-point and 1-point	Did not conduct verification and reference.
Professional jargon	The explanation is clear and easy to understand, not using complicated medical terminology.	Between 5-point and 3-point	The explanation is understandable, with minimal use of complex medical terminology.	Between 3-point and 1-point	Frequently uses complicate medical terminology.
Humanistic care					
Speech	Appropriate speech speed and tone.	Between 5-point and 3-point	The speech speed and tone are mildly uncomfortable.	Between 3-point and 1-point	The speech speed and tone are noticeably uncomfortable.
Amiable behavior	Appropriate response and comfort.	Between 5-point and 3-point	Provides responses and comfort.	Between 3-point and 1-point	No response or comfort.

Study Design

To date, ChatGPT has introduced 2 versions of its language model: GPT-3.5 and GPT-4. OpenAI reports that GPT-4 exhibits improved word processing and memory retention capabilities. These advancements make GPT-4 particularly suitable for the aims of our research [32]. The preliminary step involves ascertaining the feasibility of ChatGPT functioning as a standardized patient in medical training evaluations. To gain preliminary insights, we conduct a preliminary, exploratory conversation with ChatGPT-4. We chose inflammatory bowel disease (IBD), a relatively rare condition with multiple symptoms that can be easily confused with other diseases [33,34]. In this experiment, the chatbot was given only the name of the disease, without detailed criteria, and was asked to perform as a standardized patient to assess users' performance as if taking a standardized patient examination.

We asked a series of questions regarding IBD's symptoms to determine if ChatGPT could respond appropriately like a real standardized patient. Subsequently, ChatGPT was requested to provide a score for our inquiry process, along with its criteria for assessment. The chatbot was evaluated using 3 different approaches: using poor, medium, and good inquiries. The criteria were as follows: in the poor inquiry approach, no questions pertinent to IBD were posed, and the language was unprofessional and overly casual. The medium inquiry approach exhibited a combination of relevant and irrelevant questions, with varying degrees of accuracy. Conversely, the good inquiry approach demonstrated both linguistic precision and a consistently accurate focus on relevant topics. Each level underwent 3 consecutive tests, and in each test, the chatbot generated the answer 10 times, from which an average score out of 30 (5 scores each × 6 groups) was calculated for each level.

For further exploration, a second phase of the experiment was designed. In this subsequent experiment, we used a detailed script accompanied by clear criteria provided by our clinical skill training center. An example from the script, featuring a perfect, full-score conversation between a standardized patient and a medical student, was used to test the chatbot. By modifying the dialog, including additions and deletions, the final score could vary. To initiate the conversation, the chatbot was provided with a patient's medical record, along with a prompt outlined in Table 2. During the subsequent conversation, the questions on the script list were asked respectively to assess if the chatbot can extract information from the medical record precisely. Each answer was checked and compared to identify the flaws in responses, which were considered as key points for future improvement. At last, a score based on the given criteria was assigned by ChatGPT, along with a detailed analysis of the score. Repeated experiments (OP-1 to OP-5) were conducted 5 times, for each experiment, ChatGPT was asked to generate the score 30 times to obtain an average error rate for this overall preliminary experiment. From OP-1 to OP-5, the questions asked differed; therefore, the total standard score also varied. The questions asked and the total score were different, but the criterion was the same. From OP-1 to OP-5, the total score was in increasing order: 20/100, 60/100, 75/100, 85/100, and 90/100, respectively. A score of 20/100 is classified as "poor inquiry," while scores of 60/100 and 75/100 are classified as "medium inquiry," and 85/100 and 90/100 are classified as "good inquiry." Given that the total score varied across each trial, we used a metric known as the score difference percentage (SDP) to quantify ChatGPT's performance. This metric represents the error rate in terms of the percentage difference between expected and actual scores. The formula used is as follows.

After the preliminary experiment was done, a full inspection of the entire conversation was given. ChatGPT's performance on the following key points is focused on: the degree of anthropomorphism, clinical accuracy, and adaptability. Considering the degree of anthropomorphism, ensuring ChatGPT mimics a real patient accurately is crucial for its validity as a standardized patient, the chatbot should emulate the tone, behavior, and emotional responses typical of a real patient. In an authentic standardized patient assessment, to simulate a real-life patient, standardized patients may pose questions driven by anxiety, for instance, "Doctor, why does my stomach ailment keep recurring? Will it eventually lead to cancer?" [35]. It is acknowledged that ChatGPT may not initiate such inquiries due to its inherent limitations; however, we will investigate whether integrating a predesigned prompt enables the chatbot to generate these types of questions [36]. Moreover, the chatbot needs to communicate the symptoms effectively, maintaining a balance between being too vague and excessively detailed. It should provide information that is similar to what a real patient may disclose. For clinical accuracy, it is essential that the chatbot delivers precise information aligned with the provided medical record, which is a critical aspect of using ChatGPT in standardized patient assessments. The response from the chatbot must align with the corresponding description in the medical record. ChatGPT is expected to recall previous responses and consistently adhere to the documented medical history. For adaptability, in actual standardized patient tests, medical students may request information from the standardized patient that is not provided in the script or medical history; therefore, ChatGPT must be capable of inferring unspecified information. Following each evaluation, ChatGPT's performance would be evaluated based on the earlier criteria, subsequently generating a list of issues. Based on the identified issues, the prompt was modified. Subsequently, a new series of experiments began, using RP. The final version of the prompt is determined when ChatGPT demonstrates satisfactory performance in terms of anthropomorphism, clinical accuracy, consistency in responses, and adaptability, among other factors. Additionally, the scores provided were closely inspected. We designed a comparative experiment involving the OP group. RP-1 to RP-5 were crafted correlating with OP-1 to OP-5. For each paired experiment, the questions posed and the total score were identical. The only difference between the OP and RP groups was the variation in the prompts provided. For each group, 30 runs were conducted to minimize the potential bias. Ultimately, the error rates will be compared with results from OPs to assess the efficacy of RP. A sample dialogue with ChatGPT is included in Multimedia Appendix 1.

Finally, the performance of the RPs was evaluated using a different standardized patient script to verify generalizability. In preliminary experiments, the dialogues with the chatbot were conducted in Chinese to minimize potential errors introduced by translation and the prompt provided was in English. However, according to a previous study on Chinese National Medical Licensing Examination questions, the translation of Chinese questions into English resulted in only a minor improvement (P=.16) [37]. To validate this research, the test group would be divided into 4 subgroups, representing all possible combinations of 2 languages used in prompts and dialogues, which were prompts in English (PE) + dialogue in English (DE); prompts in Chinese (PC) + dialogue in English (DE); prompts in English (PE) + dialogue in Chinese (DC); and prompts in Chinese + dialogue in Chinese (DC). All the tests in each group would be conducted 15 times to minimize the deviation caused by the inconsistency of the chatbot responses.

Table 2. Initial prompt given to ChatGPT[a].

Prompt	Consideration
I'm a first-year medical student. Please help me with an interactive test case. Are you familiar with the procedure of a standardized patient assessment? Now, I want you to play the role of a standardized patient, while I play the role of the student. Follow these steps.	Initialization
Your basic situation: [The detailed medical history is provided to ChatGPT based the script]	Set up specific scenario
You need to answer the questions I'm asking you based on the facts.	Restrict the free response of ChatGPT
At the end of my consultation, you need to rate my consultation process, based on following criteria: [The specific scoring rules are provided]	Grade with detailed criteria

[a] This is the initial, unrevised prompt used for interactions with ChatGPT in the phase 2 study design, prior to prompt engineering.

Data Analysis

The analysis software used in this research is R (version 4.3.3; R Foundation for Statistical Computing). For the first phase of the experiment, a 2-tailed *t* test is performed on the results to ascertain whether ChatGPT can differentiate between varying levels of problem relevance. In the second phase, the Shapiro-Wilk normality test was conducted to examine whether the data collected were normally distributed. For data not normally distributed, the Mann-Whitney *U* test was used to evaluate the significance of the differences in ChatGPT scores and standard scores, as well as between the OP and RP groups, thereby ascertaining the enhancements postprompt engineering. The data were collected during November 2023, December 2023, and August 2024.

Results

Feasibility Test

In the feasibility test, upon receiving the keyword "IBD," ChatGPT automatically generated a virtual medical record, in subsequent interactions, it adhered to the information within the record, simulating a real patient's behavior. For each inquiry quality group, 30 runs were conducted. Subsequently, ChatGPT provided an average score of 74.7/100 (SD 5.44) for a poor inquiry, an average score of 82.7 (SD 5.30) for a medium inquiry, and an average score of 85.0 (SD 3.27) for a good inquiry. These statistical data are presented in Table 3.

To demonstrate ChatGPT's ability to distinguish the results, 2 inquiry model groups were established. The score differences between the poor and medium inquiry groups, between the medium and good inquiry groups, and between the poor and good groups were respectively examined. Both the score differences between the poor and medium inquiry groups ($P<.001$), between the poor and good groups ($P<.001$) were significant at a significance level of $\alpha=.05$, while the score differences between the medium and good inquiry groups were not statistically significant ($P=.16$).

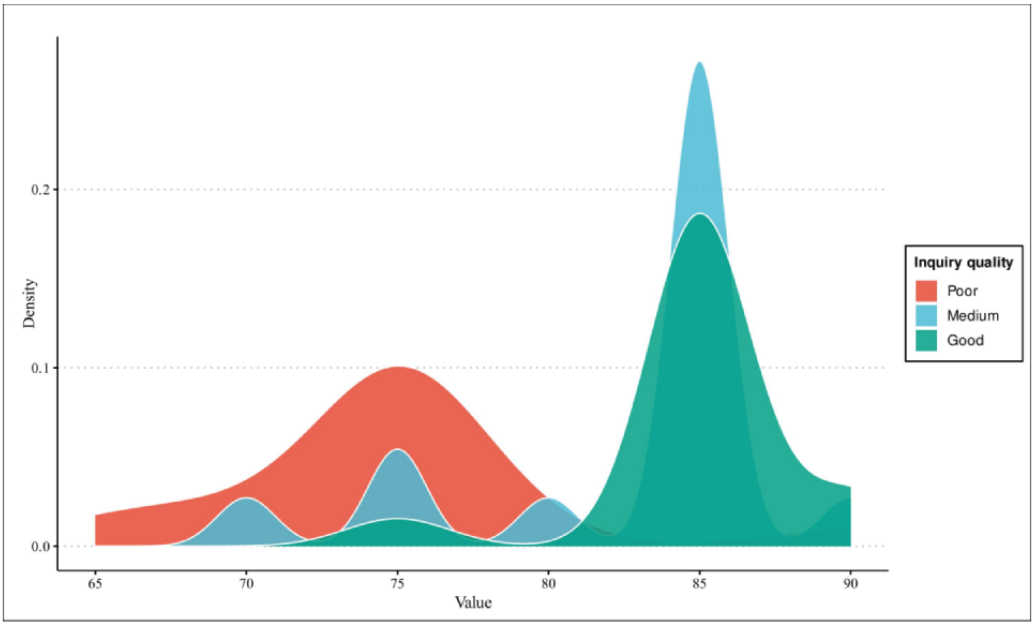
The density diagram of the scores is illustrated in Figure 1. Along with each score, the chatbot provided a detailed rating scale. Across various trials, ChatGPT's criteria varied slightly, yet the core remained consistent: it consisted of 5 components: relativeness, inquiry into medical history, clinical reasoning, professionalism, and problem-solving.

Table 3. Scores given by ChatGPT in feasibility test[a].

Inquiry quality	Mean (SD)
Poor	74.7 (5.44)
Medium	82.67 (5.30)
Good	85.00 (3.27)

[a] The table displays the mean (SD) scores provided by ChatGPT for different inquiry qualities related to inflammatory bowel disease symptom questions. The test involved three levels: "poor" (irrelevant questions and casual language), "medium" (mixed relevance), and "good" (precise and relevant). For each level, 30 interactions were conducted, and ChatGPT adhered to a virtual medical record generated for inflammatory bowel disease throughout the conversation, simulating a standardized patient.

Figure 1. Scores given by ChatGPT in feasibility test. This figure illustrates the density distribution of scores given by ChatGPT when evaluating different levels of inquiry quality (poor, medium, and good) regarding IBD symptoms. Each quality level reflects varying degrees of relevance in questioning, with ChatGPT adhering to an IBD standardized patient scenario. The distribution shows how ChatGPT assessed each inquiry type. IBD: inflammatory bowel disease.



Performance Enhancement

The results from the preliminary experiment of the second phase demonstrated that ChatGPT can accurately replicate the clinical symptoms recorded in medical records; however, the scores it provided showed significant divergences compared to the standard scores. The results are presented in Figure 2. The upper band of each bar indicates the mean value of the SDP for each dataset, while the vertical line denotes SD. The data collected from the OP experiments are compared with the RP group subsequently.

To enhance performance, we focused on the following key aspects: degree of anthropomorphism, clinical accuracy, and adaptability. Regarding the degree of anthropomorphism, some responses provided by ChatGPT were notably overly professional with medical jargon. For instance, when inquired "Where do you feel unwell?" ChatGPT responded with "Processus xiphoideus." This term is rarely used in everyday language, typically, one might say "My chest hurts." Consequently, the chatbot was instructed to "Please perform like an ordinary person that does not have much professional knowledge in the medical field and avoid using professional jargons" to adjust its behavior. Moreover, without prompts, ChatGPT cannot replicate real-life standardized patients who may ask certain questions due to anxiety. To address this, the OP was enhanced with "When I present a summary of symptoms and seek your confirmation, please first respond to my inquiry." Subsequently, simulate a patient tone characterized by anxiety, posing questions like "Can my illness be treated?" or "Is this a serious or minor illness?" Concerning clinical accuracy, several issues were identified. When queried, "Do you have any vomiting symptoms?" ChatGPT responded "Yes, basically undigested food. And my stools are normal, no black stools appeared" (this response originated from the provided medical history). The chatbot specifically inquired about the vomit habits; however, it responded with details on vomit habits and additionally mentioned the stool condition, which was documented in the medical record subsequent to the vomiting information. This indicates that ChatGPT may provide premature and overly comprehensive responses. A prompt can be introduced to address this issue: "Just answer the question each time I gave you, do not provide information that is not related to my question." Furthermore, ChatGPT provided responses not aligning with the medical records. This issue, although infrequent, emerged several times during our experiment, when asked "Where are you feeling discomfort," ChatGPT replied "head" instead of "stomach." Although the chatbot can follow the content in medical records strictly in most situations, clinical accuracy is greatly favored, the chatbot must not make mistakes in this part. Additional prompts like "Please follow strictly with the information in medical record I provided you, do not compile information already provided in the medical record" can be added. When grading, ChatGPT sometimes did not focus on the conversation about standardized patients. After providing the criteria, the chatbot would just assume the conversation has happened. This could result in a high-scoring error rate. However, when initially adding the prompt "When scoring, please concentrate solely on the dialogue that took place during the standardized patient simulation. For each criterion listed, review our conversation history to determine whether I asked the specified question. If I did not, please assign a zero for that criterion." Regarding adaptability, when ChatGPT was asked about diseases or symptoms not provided in the medical record, the chatbot was unable to answer these questions accurately, while the correct response should be "No." For instance, when being asked "Have you experienced a heart attack," ChatGPT responded with "Sorry, I cannot confirm this, as it is not indicated in the medical record." As a medical record serves as a checklist for the patients' disease history and symptoms experienced to date, it should be a complete reflection of the patients' medical situation. Therefore, when asked about the symptoms or disease that is not provided in the medical record, ChatGPT's answer should be "No, I do not have this symptom or disease." Additional prompt can be added to ascertain the adaptability: "The patient's all disease history and symptoms are as described in the medical record. The patient does not have any disease or symptoms not mentioned." Our results are shown in the circular diagram in Figure 3. Combining all the revisional prompts, new prompts are listed in Table 4.

The results of the RP group are shown in Figure 2. We first used the Shapiro-Wilk normality test to determine whether the data in each experiment were normally distributed. The results show that all the data we got, both in OP and RP groups, were not normally distributed (all *P*<.001). Then Mann-Whitney *U* test was used to determine the significance between the mean of each dataset and the standard score. For instance, considering the OP-1 group, the standard score is 20/100, we used the Mann-Whitney *U* test to test if the difference between the mean of real scores ChatGPT provided in OP-1 and number of 60 has significance. Then, Mann-Whitney *U* tests were conducted between each paired OP and RP groups, respectively, to determine if RP indeed improved the performance of ChatGPT. The detailed result is shown in Table 5.

Using another new patient script, a total of 60 experiments were conducted on 4 test groups. The results are shown in Figure 4. The PE+DC group exhibited the highest performance, achieving an average SDP value of 3.21%. The PC+DE group demonstrated the lowest performance, recording an average SDP value of 4.2%. The average SDP values achieved by the PC+DC and PE+DE groups were generally identical (SDP=3.46%). However, the PE+DE group showed superior performance over the PC+DC group, as evidenced by a smaller deviation in results. Meanwhile, the results from the Mann-Whitney *U* tests, conducted across all groups, indicated no significant differences (*P*=.65, between PC+DC and PE+DC; *P*=.46, between PC+DC and PC+DE; *P*=.98, between PC+DC and PE+DE; *P*=.29, between PE+DC and PC+DE; *P*=.51 between PE+DC and PE+DE; and *P*=.37, between PC+DE and PE+DE).

Figure 2. Error bar histogram of RP and OP experiments. This figure displays SDP across 5 experiments (OP-1 to OP-5 and RP-1 to RP-5) comparing OP and RP used for ChatGPT's evaluation as a standardized patient. Each experiment involved 30 interactions, measuring the deviation of ChatGPT's provided scores from the standard scores. The blue bars represent the mean SDP for each OP group, and the orange bars indicate the corresponding RP group. Error bars denote the SD of SDP, highlighting variations in accuracy across prompt types. OP: original prompt; RP: revised prompt; SDP: score difference percentage.

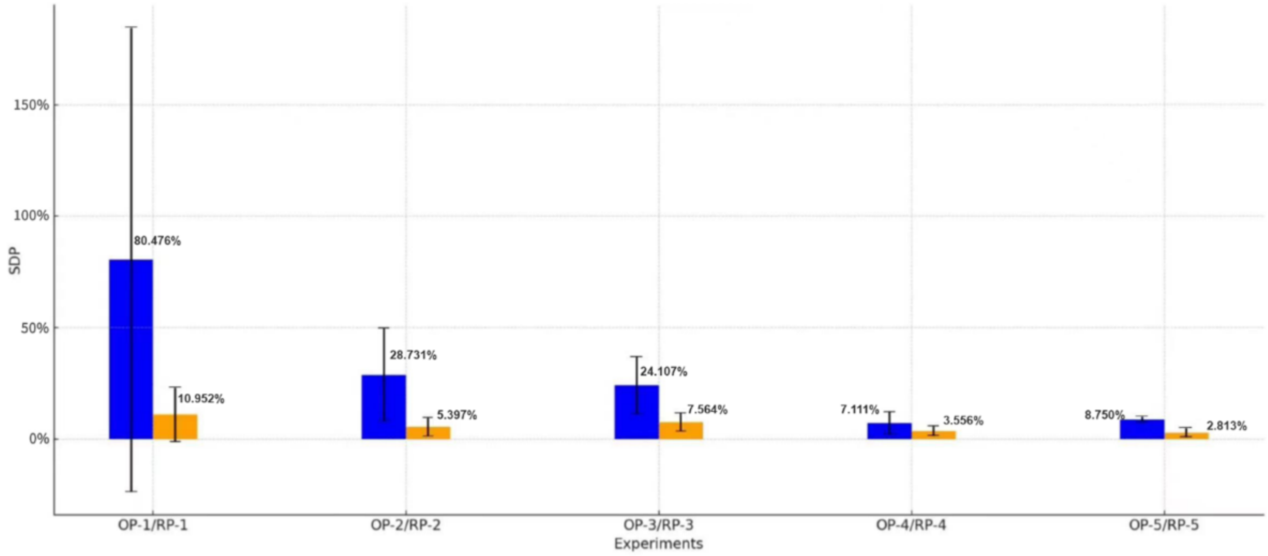


Figure 3. Prompt revision for ChatGPT in standardized patient assessment. This diagram summarizes key areas for prompt revision aimed at improving ChatGPT's performance as a standardized patient. Focus areas include clinical accuracy, degree of anthropomorphism, and adaptability. Each segment outlines specific issues identified during initial tests—such as premature answers, use of professional jargon, and deviation from medical records—and corresponding prompt modifications. These adjustments were designed to enhance ChatGPT's ability to simulate realistic patient behavior, adhere to medical records, and respond consistently during medical interactions.

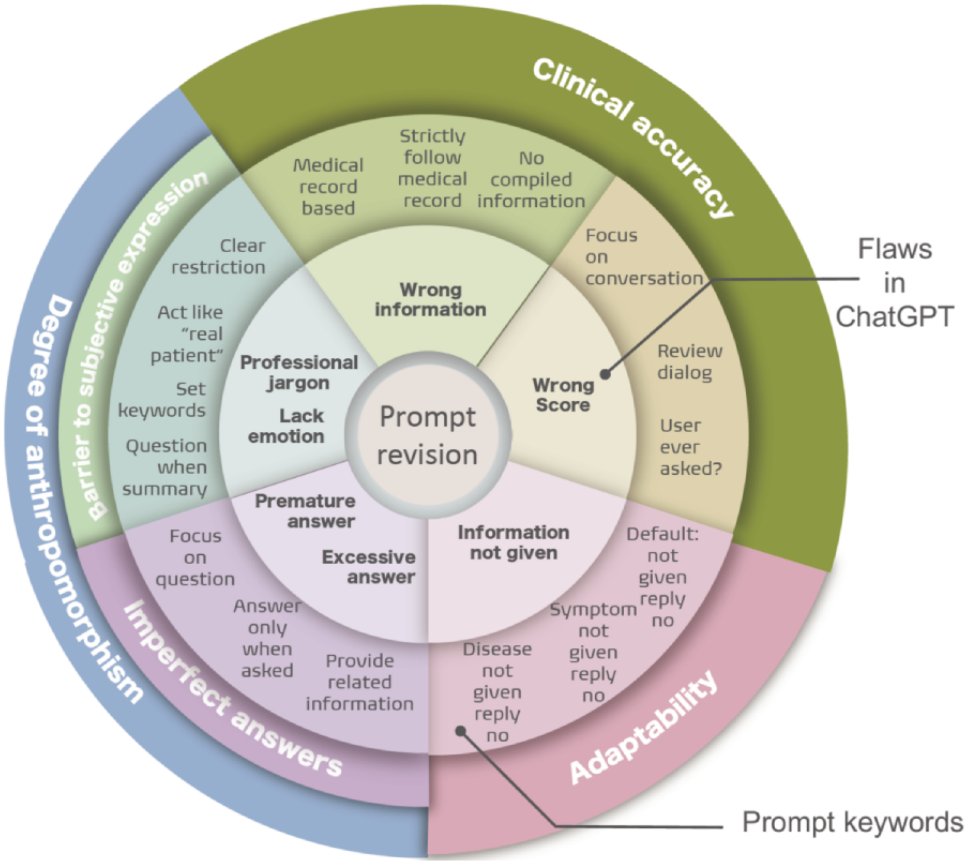


Table 4. The revised prompt given to ChatGPT[a].

Prompt	Consideration
I'm a first-year medical student. Please help me with an interactive test case. Are you familiar with the procedure of a standardized patient assessment? Now, I want you to play the role of a standardized patient, while I play the role of the student. Follow these steps:	Initialization
Your basic situation: [The detailed medical history is provided to ChatGPT based the script]	Set up specific scenario
You need to answer the questions I'm asking you based on the facts.	Restrict the free response of ChatGPT
Please perform like an ordinary person that does not have much professional knowledge in medical field. Avoid using professional jargons.	Avoid professional jargon
When I present a summary of symptoms and seek your confirmation, please first respond to my inquiry. Subsequently, simulate a patient tone characterized by anxiety, posing questions like 'Can my illness be cured?' or 'Is this a serious or minor illness?'	Set up question due to anxiety
Just answer the question each time I gave you, do not provide information that is not related to my questions.	Avoid premature and excessive answer
Please follow strictly with the information in medical record I provided you, do not compile information that is already provided in medical record.	Avoid wrong information
When giving score, please only focus on the conversation happened in standardized patient simulation, for each list in criteria, review our conversation history to check if I have ever asked this question, if not, then you give me a zero for this list.	Improve grading accuracy
The patient's all disease history and symptoms are as described in medical record. The patient does not have any disease or symptoms not mentioned.	Improve adaptability
At the end of my consultation, you need to rate my consultation process, the specific scoring rules are as follows: [The specific scoring rules are provided]	Grade with detailed criteria

[a] This is the revised prompt used for interactions with ChatGPT in the phase 2 study design, after prompt engineering.

Table 5. Statistical data for experiment, using SDP[a,b].

Experiment order	Real score			
	Mean (SD), %	Mean ChatGPT score	Accurate score	*P* value
OP[c]-1	80.476 (1.058)	12.633	7	<.001
RP[d]-1	10.952 (0.123)	7.033	7	.28
OP-2	28.731 (0.212)	26.567	21	<.001
RP-2	5.397 (0.041)	20.333	21	<.001
OP-3	24.103 (0.131)	32	26	<.001
RP-3	7.564 (0.039)	24.033	26	<.001
OP-4	7.111 (0.051)	32.134	30	<.001
RP-4	3.556 (0.021)	29.8	30	.81
OP-5	8.750 (0.013)	34.8	32	<.001
RP-5	2.813 (0.021)	32.833	32	<.001
OP-overall	29.834 (0.550)	—[e]	—	—
RP-overall	6.056 (0.068)	—	—	—

[a] SDP: score difference percentage.

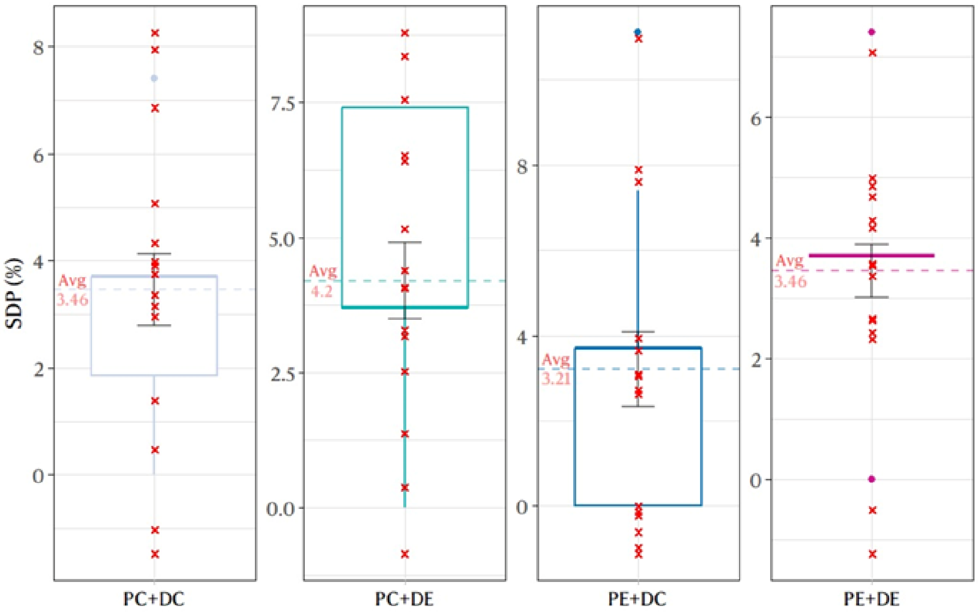
[b] The table presents the accuracy of ChatGPT's scoring during standardized patient assessments, comparing the OPs and RPs. The scoring measures how well ChatGPT's responses align with predefined standard scores for each scenario. Metrics include the mean and SD of the SDP, indicating the deviation between ChatGPT's scores and the expected standards. The table also provides accurate scores and *P* values from Mann-Whitney *U* tests, evaluating the significance of differences between ChatGPT's scoring and the expected results.

[c] OP: original prompt.

[d] RP: revised prompt.

[e] No data.

Figure 4. Experiment results using new script. This figure displays the performance of ChatGPT using revised prompts on another script across 4 language combinations: PC+DC, PC+DE, PE+DC, and PE+DE. Each group underwent 15 trials, and the results were analyzed based on the SDP metric, which measures the accuracy of ChatGPT's scoring. The Mann-Whitney U tests indicated no significant differences between any group pairs, suggesting similar performance across language combinations. PC: prompts in Chinese; PE: prompts in English, DC: dialogue in Chinese; DE: dialogue in English. SDP: score difference percentage.



Discussion

Principal Findings

Phase 1: Preliminary Experiment

In this study, we conducted an exploratory study to examine the potential of using ChatGPT as standardized patients in medical education. Also, the performance enhancement after RP was evaluated. The knowledge of ChatGPT is not updated every day, therefore the precision and accuracy of the chatbot may decrease when it comes to new diseases and medicine [30,38]. All the papers designed in this study for ChatGPT evaluation do not include diseases newly found, as all the analyses and questions provided to ChatGPT were conducted between November 2023 to December 2023. We found that with refined prompts, ChatGPT could simulate patient interactions effectively, demonstrating improvements in realism, clinical accuracy, and adaptability. The use of ChatGPT as a standardized patient offers a cost-effective alternative to traditional human standardized patients, potentially enhancing access to medical training, especially in resource-limited settings.

Our results suggest that, in the preliminary experiment, poor inquiry is most dense in a relatively high score of around 75, this score seems to be higher than expected as all 4 questions inputted to ChatGPT are basically irrelevant to the standardized patient test. However, ChatGPT is an LLM that is finetuned by humans, it is finetuned to minimize offense to humans [39,40] It is understandable that ChatGPT tends to give the user a higher score even if all the questions are wrong. As to the chatbot, it may seem to be offensive if providing a low score to human users while not having a detailed criterion. This problem disappeared in the later experiments where detailed criterion was provided. As for the medium inquiry way, ChatGPT provided an average score higher than that of a poor inquiry model. Our results suggest that the score of the medium inquiry group is the most spread, as well as the largest data range among the 3 models. This reflects that in the field of standardized patient assessment, ChatGPT can distinguish between a relevant question and an irrelevant one, but its performance is not stable and easily affected without proper definition and restriction.

Even though the average score difference between medium and good inquiries is not significant, good inquiry has the smallest variance with the highest data density. As our results suggest, the difference between poor and medium groups is significant, while the difference between medium and good is not statistically significant. Considering other trends, however, the increase in the average score, along with other data features, with respect to the increase and difference in inquiry question relevance, indicate that ChatGPT is aware of which question does not belong to a standardized patient assessment process and which question belongs to it. Although it cannot distinguish clearly without a concrete prompt, the chatbot is familiar with the process of standardized patient assessment. However, its criterion is different from the criterion provided by experts, and the scores it provides are erroneous. This is sufficient to provide evidence for the feasibility of using ChatGPT to perform standardized patient training.

Phase 2: Prompt Engineering

The preliminary experiment of the performance enhancement set up a blueprint of the whole experiment. As illustrated in Figure 2, in the first few experiments, the improvement of RP is quite significant. For instance, in the OP-1 experiment, the mean score has the SDP around 0.8 which means that the score provided by ChatGPT is 1.8 times higher than the accurate score. With an overall average SDP of 29.834%, this shows that ChatGPT's behavior is highly unpredictable and unreasonable, and thus not feasible for application based on the current method. One noticeable trend from Figure 2 is that, when the experiment continues, the SD and average score become more accurate. This is not because of the fact that the performance of ChatGPT improves as the experiment is conducted, rather, literature suggests ChatGPT falls short when handling similar tasks. Without prompt limitation, the chatbot tends to include the content of the whole conversation into grading, beyond the user's query. For example, if the user asked 1 question that addressed 1 key point in the criteria, but ChatGPT's reply went beyond the scope of this question, its answer may cover two or more key points in the criteria. Therefore, even though the user did not explicitly inquire about certain key points, ChatGPT's expansive response might still count these unasked points during grading. Consequently, this feature will let ChatGPT tend to provide a higher overall score.

To ensure a stable, accurate performance in standardized patient assessment, prompt engineering plays a critical role. ChatGPT's output can be generally modified and controlled by changing prompts, thus a good prompt with many restrictions can constrain ChatGPT to guide the chatbot to our ideal output. The enhancement and revision of prompts become pivotal, which is the most significant and pivotal contribution of this study. The reasons why choosing the degree of anthropomorphism, clinical accuracy, and adaptability as our evaluation key point is based on ChatGPT's prompt optimizing key features [41,42]. In the context of standardized patient training, the chatbot is required to perform like a real person, moreover, it must adhere more closely to the behaviors of a real-life patient, which is the essence of the test. In achieving so, the fundamental requirement for the chatbot is to have a high level of anthropomorphism which consists of 2 major problems discovered in this research: barrier to subjective expression and imperfect answers. For the former aspect, ChatGPT tends to use professional medical jargon to illustrate symptoms in medical reports which do not accurately reflect its role as a real patient. Also, in only 1 of 5 experiments, ChatGPT responded to a question as if it were experiencing anxiety. In OP-3, when asked "Have you ever been to the hospital before for this disease?" ChatGPT first answered the question objectively, then subjectively added, "I fear that this disease has lasted for too long, so I came here to consult you." Although the case that ChatGPT has an anxiety tone is rare, this demonstrates that the chatbot can imitate this tone, requiring only some constraints in the prompt. Future studies may explore the possible prompts engineering to improve its performance in simulating human emotions. In essence, the performance of ChatGPT in the preliminary experiment is unstable and poor; however, it possesses the capability for preliminary testing yet lacks guidance, thereby necessitating optimization via prompt engineering [43], underscoring the significance and necessity of the second-phase experiment.

Ultimate Performance Enhancement

After applying RP, ChatGPT used descriptive words to answer questions like "Where is your discomfort," instead of detailed medical jargon, and the chatbot asked questions like "Why my symptoms persist?" "Can my disease be cured?" After the user has summarized the conversation. Though the chatbot's tone may seem "mechanical," it indicates that RP indeed works. For the latter aspect, the imperfect answers here refer to the chatbot's premature and excessive answers, which significantly influence grading. This part has a great influence on grading. For instance, when asked "Have you ever had other diseases in stomach?" the chatbot first answered yes, and explained when diagnosed and with what diseases. After providing sufficient information for the question, ChatGPT added "I have been to the hospital for the first time for this 2 months ago." The answer may seem reasonable. Yet, during grading, ChatGPT automatically awarded full marks for "Asking about information on previous hospital visits," justifying that the user had inquired about it. However, no question related to this was asked in that conversation. When asked for more detailed information, ChatGPT responded that since the content of the conversation mentioned the patient had gone to the hospital 2 months ago, it should count; even though this information was provided by the chatbot subjectively, not in response to the user's question. Therefore, premature and excessive questions not only influence degrees of anthropomorphism but also contribute dramatically to grading errors. Subsequently, clinical accuracy is also a concern [44]. When restrictive prompts such as "Focus on my question" and "Answer only what is asked" are used, ChatGPT began to use plain, short answers to respond, from one extreme to another, but the answer contains the essence without premature questions, making these responses preferable. During the preliminary experiments, ChatGPT can strictly follow the medical record, only once mistaking a stomachache for a headache. Additionally, in 1 conversation, the chatbot provided compiled information on the medical history. Therefore, to assure accuracy, a restrictive prompt asking ChatGPT to base answers only on the record is crucial. Also, the chatbot must be accurate when giving the score, as this is the only quantitative reflection of ChatGPT's performance in correctly evaluating medical students' performance who are taking the standardized patient examination.

With RP, the chatbot now focuses on the content that appears in the standardized patient training conversation, before that, it sometimes mistook the content of the criteria as part of the judgment, resulting in full mark. Adaptability is also important, RP made a ChatGPT response with "No" for the symptoms asked but not in medical history in the following experiment. After applying RP, ChatGPT's performance improves dramatically. One of the most serious problems in using ChatGPT is that the answer it provides has a high level of randomness and this is hard to avoid. The average SDP improved to 20 times higher than OP results. Considering the high randomness in answers provided by the chatbot, this result is quite acceptable. Also, the score deviation for each experiment is quite small, the average SD decreased to 0.068 from 0.55 in the original group results. By adding these prompts, we reduce SD to nearly one-fifteenth of its original value. Massive differences in average value and SD indicate that by giving our extra, RP, the performance of ChatGPT as a standardized patient improves significantly. The significant difference between the OP and RP results indicates that RP improved ChatGPT's performance significantly. RP-2, RP-3, and RP-5 generally have lower SD with a lower mean SDP, which represents a higher accuracy and precision than their corresponding OP groups. To fully use ChatGPT in a standardized patient training process, good accuracy and consistency in responses are required. After applying our RP, ChatGPT has an overall SDP of 6.056%. This score is closer to 0% and improves a lot compared to the un-RP group (29.834%), even though there still remains space to improve. The results also indicate that ChatGPT’s performance can be fine-tuned through the revision of prompts provided to it. Future studies may conduct a more comprehensive examination to reduce the SDP further.

The efficacy of RP was confirmed through its universal applicability when tested on another script, achieving a similar average SDP to our preliminary script. Furthermore, validation tests conducted on the new script also indicated that the impact of language on chatbot performance is insignificant ($P>.05$, for all groups). The observed performance deviation is within acceptable limits. However, the highest accuracy rate was observed in the test group that used English prompts and Chinese dialogues.

Limitations

There are several limitations in this study. First, to minimize errors associated with word translation, we conducted the simulated standardized patient test dialog with ChatGPT in standard simplified Chinese [45]. Nevertheless, the prompt provided to ChatGPT was in English to ensure accuracy. Although a separate experiment was conducted to confirm the minor impact of language used on the performance of ChatGPT, future studies may test the performance of ChatGPT on standardized patients in other languages and corresponding prompt engineering to improve its performance. Secondly, the criteria provided to ChatGPT involved only objective key points. In real standardized patient tests, some subjective judgments may also be made accordingly. Finally, the prompt revision we provided represents the optimal approach. Further research is necessary to use ChatGPT more effectively in standardized patient training.

Conclusions

ChatGPT, as a representative LLM, has much potential application in medical assessment. Our study suggests that it is feasible to use ChatGPT as a simulated patient for standardized patient evaluation and scoring. The output performance concerns introduced by the randomness of ChatGPT could be improved by adding detailed, restrictive prompts. By using prompts revised in this study, the accuracy of ChatGPT’s scoring significantly improved to an acceptable level that could be used in actual standardized patients involved in medical training. Additionally, ChatGPT’s responses during the conversation became more accurate and lifelike. It is worth noting that, in this application scenario, the influence of the adopted language on the chatbot’s outcome is nonsignificant. Overall, ChatGPT’s accuracy and performance in standardized patient assessment are acceptable, but they highlight the need for continuing improvement before it can be used as a fully trustworthy clinical assessment method.

Acknowledgments

The authors used the generative artificial intelligence tool ChatGPT by OpenAI to examine the potential use of LLMs in replacing the role of standardized patients, mainly focusing on their feasibility and accuracy. A sample dialogue with ChatGPT is made available in Multimedia Appendix 1. This study was funded by the Aier Eye Hospital-Sichuan University Research Grant (23JZH043). This study was supported by Training Programs for Innovation and Entrepreneurship of Sichuan University (project 20241039L); the Interdisciplinary Crossing and Integration of Medicine and Engineering for Talent Training Fund, West China Hospital, Sichuan University; and the 1·3·5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYYC21004). The sponsor had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. The views expressed are those of the authors and not necessarily those of the West China Hospital and Sichuan University.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to institutional case privacy and a large number of interaction dialogs but are available from the corresponding author on reasonable request. Other than the sample dialogue provided in the Multimedia Appendix 1, additional detailed dialogues with ChatGPT may be available by contacting the corresponding author.

Authors' Contributions

RY conceptualized the study, curated the data, performed formal analysis, and managed the project. YH and XW contributed to the conceptualization of the study. GQ also conceptualized the study, acquired funding, and managed the project. CW developed the methodology, curated the data, performed formal analysis, conducted the investigation, prepared the original draft of the manuscript, and contributed to visualization. SL and NL developed the methodology, conducted the investigation, and prepared the original draft of the manuscript. DL, XT, and XZ contributed to the visualization and preliminary manuscript revision. KL acquired funding, supervised the project, and managed the project. DP supervised the project and managed the project. All authors reviewed and edited the manuscript and have read and agreed to the published version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A sample interaction dialogue with ChatGPT for RP-2 (revised prompts) group (standard score=21).
[DOCX File , 26 KB - jmir_v27i1e59435_app1.docx]

References

1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
2. Rane NL, Tawde A, Choudhary SP, Rane J. Contribution and performance of ChatGPT and other large language models (LLM) for scientific and research advancements: a double-edged sword. Int Res J Modernization Eng Technol Sci 2023;5(10):875-899. [doi: 10.56726/irjmets45213]
3. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023;307(2):e230163. [doi: 10.1148/radiol.230163] [Medline: 36700838]
4. Mattas PS. ChatGPT: a study of AI language processing and its implications. Int J Res Publication Rev 2023;04(02):435-440. [doi: 10.55248/gengpi.2023.4218]
5. Fui-Hoon Nah F, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. J Inf Technol Case Application Res 2023;25(3):277-304. [doi: 10.1080/15228053.2023.2233814]
6. Huang H, Zheng O, Wang D, Yin J, Wang Z, Ding S, et al. ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. Int J Oral Sci 2023;15(1):29 [FREE Full text] [doi: 10.1038/s41368-023-00239-y] [Medline: 37507396]
7. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
8. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
9. Al-Worafi YM, Goh KW, Hermansyah A, Tan CS, Ming LC. The use of ChatGPT for education modules on integrated pharmacotherapy of infectious disease: educators' perspectives. JMIR Med Educ 2024;10:e47339. [doi: 10.2196/47339] [Medline: 38214967]
10. Parikh PM, Shah DM, Parikh KP. Judge juan manuel padilla garcia, ChatGPT, and a controversial medicolegal milestone. IJMS 2023;75:3-8. [doi: 10.25259/ijms_31_2023]
11. ChatGPT. OpenAI. URL: https://help.openai.com/en/articles/6825453-chatgpt-release-notes [accessed 2023-12-18]
12. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ 2023;9:e46885 [FREE Full text] [doi: 10.2196/46885] [Medline: 36863937]
13. Wu T, He S, Liu J, Sun S, Liu K, Han QL, et al. A brief overview of ChatGPT: the history, status quo and potential future development. IEEE/CAA J Automatica Sin 2023;10(5):1122-1136. [doi: 10.1109/jas.2023.123618]
14. Roumeliotis KI, Tselikas ND. ChatGPT and open-AI models: a preliminary review. Future Internet 2023;15(6):192. [doi: 10.3390/fi15060192]
15. Short CE, Short JC. The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation. J Bus Venturing Insights 2023;19:e00388. [doi: 10.1016/j.jbvi.2023.e00388]
16. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individual Differ 2023;103:102274. [doi: 10.1016/j.lindif.2023.102274]
17. Keiser MM, Turkelson C. Using students as standardized patients: development, implementation, and evaluation of a standardized patient training program. Clin Simul Nurs 2017;13(7):321-330. [doi: 10.1016/j.ecns.2017.05.008]
18. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare 2023;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
19. Lee H. The rise of ChatGPT: exploring its potential in medical education. Anat Sci Educ 2024;17(5):926-931. [doi: 10.1002/ase.2270] [Medline: 36916887]
20. Wong RSY, Ming LC, Raja Ali RA. The intersection of ChatGPT, clinical medicine, and medical education. JMIR Med Educ 2023;9:e47274 [FREE Full text] [doi: 10.2196/47274] [Medline: 37988149]
21. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. J Med Educ 1964;39:802-805 [FREE Full text] [Medline: 14180699]
22. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. Acad Med 1993;68(6):443-451. [doi: 10.1097/00001888-199306000-00002] [Medline: 8507309]

23. Ainsworth MA, Rogers LP, Markus JF, Dorsey NK, Blackwell TA, Petrusa ER. Standardized patient encounters. a method for teaching and evaluation. JAMA 1991;266(10):1390-1396. [doi: 10.1001/jama.266.10.1390] [Medline: 1880869]
24. Plaksin J, Nicholson J, Kundrod S, Zabar S, Kalet A, Altshuler L. The benefits and risks of being a standardized patient: a narrative review of the literature. Patient 2016;9(1):15-25. [doi: 10.1007/s40271-015-0127-y] [Medline: 26002043]
25. King AM, Perkowski‐Rogers LC, Pohl HS. Planning standardized patient programs: case development, patient training, and costs. Teach Learn Med 1994;6(1):6-14. [doi: 10.1080/10401339409539636]
26. Petrusa ER. Taking standardized patient-based examinations to the next level. Teach Learn Med 2004;16(1):98-110. [doi: 10.1207/s15328015tlm1601_19] [Medline: 14987183]
27. Tong Y, Tan CH, Sia C, Shi Y, Teo HH. Rural-urban healthcare access inequality challenge: transformative roles of information technology. MIS Q 2022;46(4):1937-1982. [doi: 10.25300/misq/2022/14789]
28. Kocoń J, Cichecki I, Kaszyca O, Kochanek M, Szydło D, Baran J, et al. ChatGPT: jack of all trades, master of none. Inf Fusion 2023;99:101861. [doi: 10.1016/j.inffus.2023.101861]
29. Lee U, Jung H, Jeon Y, Sohn Y, Hwang W, Moon J, et al. Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in English education. Educ Inf Technol 2023;29(9):11483-11515. [doi: 10.1007/s10639-023-12249-8]
30. Haze T, Kawano R, Takase H, Suzuki S, Hirawa N, Tamura K. Influence on the accuracy in ChatGPT: differences in the amount of information per medical field. Int J Med Inform 2023;180:105283. [doi: 10.1016/j.ijmedinf.2023.105283] [Medline: 37931432]
31. Muftić F, Kadunić M, Mušinbegović A, Abd Almisreb A. Exploring medical breakthroughs: a systematic review of ChatGPT applications in healthcare. Southeast Eur J Soft Comput 2023;12(1):41 [FREE Full text]
32. GPT-4. Chatbot App. URL: https://tinyurl.com/bdhpmh7p [accessed 2023-12-18]
33. Pithadia AB, Jain S. Treatment of inflammatory bowel disease (IBD). Pharmacol Rep 2011;63(3):629-642 [FREE Full text] [doi: 10.1016/s1734-1140(11)70575-8] [Medline: 21857074]
34. Yin R, Neyens DM. Online health resource use by individuals with inflammatory bowel disease: analysis using the National Health Interview Survey. J Med Internet Res 2020;22(9):e15352 [FREE Full text] [doi: 10.2196/15352] [Medline: 32969831]
35. King JJC, Das J, Kwan A, Daniels B, Powell-Jackson T, Makungu C, et al. How to do (or not to do) using the standardized patient method to measure clinical quality of care in LMIC health facilities. Health Policy Plan 2019;34(8):625-634 [FREE Full text] [doi: 10.1093/heapol/czz078] [Medline: 31424494]
36. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023;25:e48568 [FREE Full text] [doi: 10.2196/48568] [Medline: 37379067]
37. Fang C, Wu Y, Fu W, Ling J, Wang Y, Liu X, et al. How does ChatGPT-4 preform on non-English national medical licensing examination? An evaluation in Chinese language. PLOS Digit Health 2023;2(12):e0000397 [FREE Full text] [doi: 10.1371/journal.pdig.0000397] [Medline: 38039286]
38. Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
39. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber Phys Syst 2023;3:121-154. [doi: 10.1016/j.iotcps.2023.04.003]
40. Liu B, Xiao B, Jiang X, Cen S, He X, Dou W. Adversarial attacks on large language model-based system and mitigating strategies: a case study on ChatGPT. Secur Commun Netw 2023;2023:1-10. [doi: 10.1155/2023/8691095]
41. Henrickson L, Meroño-Peñuela A. Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. AI Soc 2023;39(5). [doi: 10.1007/s00146-023-01752-8]
42. Giray L. Prompt engineering with ChatGPT: a guide for academic writers. Ann Biomed Eng 2023;51(12):2629-2633. [doi: 10.1007/s10439-023-03272-4] [Medline: 37284994]
43. Lubiana T, Lopes R, Medeiros P, Silva JC, Goncalves ANA, Maracaja-Coutinho V, et al. Ten quick tips for harnessing the power of ChatGPT in computational biology. PLoS Comput Biol 2023;19(8):e1011319 [FREE Full text] [doi: 10.1371/journal.pcbi.1011319] [Medline: 37561669]
44. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. Res Sq 2023. [doi: 10.21203/rs.3.rs-2566942/v1] [Medline: 36909565]
45. Li L, Zhang H, Li C, You H, Cui W. Evaluation on ChatGPT for Chinese language understanding. Data Intell 2023;5(4):903. [doi: 10.1162/dint_a_00232]

Abbreviations

AI: artificial intelligence
DC: dialogue in Chinese
DE: dialogue in English
IBD: inflammatory bowel disease

LLM: large language model
OP: original prompt
PC: prompts in Chinese
PE: prompts in English
RP: revised prompt
SDP: score difference percentage

Edited by A Coristine; submitted 11.04.24; peer-reviewed by A Ciftci Sisman, Y Wang, H Huang; comments to author 09.08.24; revised version received 24.08.24; accepted 11.11.24; published 01.01.25.

Please cite as:
Wang C, Li S, Lin N, Zhang X, Han Y, Wang X, Liu D, Tan X, Pu D, Li K, Qian G, Yin R
Application of Large Language Models in Medical Training Evaluation—Using ChatGPT as a Standardized Patient: Multimetric Assessment
J Med Internet Res 2025;27:e59435
URL: https://www.jmir.org/2025/1/e59435
doi:10.2196/59435
PMID:

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org