Contents

Viewpoint

SOMAScience: A Novel Platform for Multidimensional, Longitudinal Pain Assessment (e47177)
Chloe Gunsilius, Joseph Heffner, Sienna Bruinsma, Madison Corinha, Maria Cortinez, Hadley Dalton, Ellen Duong, Joshua Lu, Aisulu Omar, Lucy Owen, Bradford Roarr, Kevin Tang, Frederike Petzschner

40

Original Papers

Application of eHealth Tools in Anticoagulation Management After Cardiac Valve Replacement: Scoping Review Coupled With Bibliometric Analysis (e48716)
Ying Wu, Xiaohui Wang, Mengyao Zhou, Zhuoer Huang, Lijuan Liu, Li Cong

70

Investigating Citizens’ Acceptance of Contact Tracing Apps: Quantitative Study of the Role of Trust and Privacy (e48700)
Grace Fox, Lisa van der Werff, Pierangelo Rosati, Theo Lynn

195

Documentation Completeness and Nurses’ Perceptions of a Novel Electronic App for Medical Resuscitation in the Emergency Room: Mixed Methods Approach (e46744)
Kin Cheung, Chak Yip

208

Evidence of How Physicians and Their Patients Adopt mHealth Apps in Germany: Exploratory Qualitative Study (e48345)
Tanja Schroeder, Maximilian Haug, Andrew Georgiou, Karla Seaman, Heiko Gewald

218

Engagement With a Remote Symptom-Tracking Platform Among Participants With Major Depressive Disorder: Randomized Controlled Trial (e44214)
Katie White, Ewan Carr, Daniel Leightley, Faith Matcham, Pauline Conde, Yatharth Ranjan, Sara Simblet, Erin Dawe-Lane, Laura Williams, Claire Henderson, Matthew Hotopf

233

Effectiveness of a Nurse-Led Mobile-Based Health Coaching Program for Patients With Prostate Cancer at High Risk of Metabolic Syndrome: Randomized Waitlist Controlled Trial (e47102)
Kyoungjin Lee, Jeongok Park, Eui Oh, Juhhee Lee, Chang Park, Young Choi

253

Use and Engagement With Low-Intensity Cognitive Behavioral Therapy Techniques Used Within an App to Support Worry Management: Content Analysis of Log Data (e47321)
Paul Farrand, Patrick Raue, Earline Ward, Dean Repper, Patricia Areán

288
A Behaviorally Informed Mobile App to Improve the Nutritional Quality of Grocery Shopping (SwapSHOP): Feasibility Randomized Controlled Trial (e45854)
Carmen Piernas, Charlotte Lee, Alice Hobson, Georgina Harmer, Sarah Payne Riches, Michaela Noreik, Susan Jebb. 301

Effectiveness of Metaverse Space–Based Exercise Video Distribution in Young Adults: Randomized Controlled Trial (e46397)
Rami Mizuta, Noriaki Maeda, Tsubasa Tashiro, Yuta Suzuki, Sayo Kuroda, Ayano Ishida, Sakura Oda, Tomoya Watanabe, Yuki Tamura, Makoto Komiya, Yukio Urabe. 315

Evaluation of Chinese HIV Mobile Apps by Researchers and Patients With HIV: Quality Evaluation Study (e52573)
Peng Liu, Lingmeng Wang, Fuzhi Wang. 364

mHealth Apps for the Self-Management of Low Back Pain: Systematic Search in App Stores and Content Analysis (e53262)
Tianyu Zhou, David Salman, Alison McGregor. 382

Short-Term Effect of a Health Promotion Intervention Based on the Electronic 12-Hour Dietary Recall (e-12HR) Smartphone App on Adherence to the Mediterranean Diet Among Spanish Primary Care Professionals: Randomized Controlled Clinical Trial (e49302)
Luis Béjar, Pedro Mesa-Rodriguez, María García-Perea. 394

The Association of eHealth Literacy Skills and mHealth Application Use Among US Adults With Obesity: Analysis of Health Information National Trends Survey Data (e46656)
George Shaw Jr, Bianca Castro, Laura Gunn, Keith Norris, Roland Thorpe Jr. 409

Barriers and Implications of 5G Technology Adoption for Hospitals in Western China: Integrated Interpretive Structural Modeling and Decision-Making Trial and Evaluation Laboratory Analysis (e48842)
Linyun Zhou, Minghuan Jiang, Ran Duan, Feng Zuo, Zongfang Li, Songhua Xu. 426

Reviews

Dissemination Strategies for mHealth Apps: Systematic Review (e50293)
Henri Moungui, Hugues Nana-Djoungu, Che Anyiang, Mireia Cano, Jose Ruiz Postigo, Carme Carrion. 85

Effects of mHealth-Based Lifestyle Interventions on Gestational Diabetes Mellitus in Pregnant Women With Overweight and Obesity: Systematic Review and Meta-Analysis (e49373)
Yirong He, Chuanyu Huang, Qiuyang He, Shujuan Liao, Biru Luo. 105

Effectiveness of Telecare Interventions on Depression Symptoms Among Older Adults: Systematic Review and Meta-Analysis (e50787)
Man Wu, Chaoyang Li, Ting Hu, Xuexiang Zhao, Guiyuan Qiao, Xiaolian Gao, Xinhong Zhu, Fen Yang. 124

Effects of Digital Physical Health Exercises on Musculoskeletal Diseases: Systematic Review With Best-Evidence Synthesis (e50616)
Johanna Nagel, Florian Wegener, Casper Grim, Matthias Hoppe. 140

Mobile and Computer-Based Applications for Rehabilitation Monitoring and Self-Management After Knee Arthroplasty: Scoping Review (e47843)
Sabhya Pritwani, Purnima Shrivastava, Shrutri Pandey, Aijt Kumar, Rajesh Malhotra, Ralph Maddison, Niveditha Devasenapathy. 158
Functionality and Quality of Asthma mHealth Apps and Their Consistency With International Guidelines: Structured Search and Evaluation (e47295)
Billy Robinson, Eleni Proimos, Daniel Zou, Enying Gong, Brian Oldenburg, Katharine See. ................................................................. 328

Developing a Comprehensive List of Criteria to Evaluate the Characteristics and Quality of eHealth Smartphone Apps: Systematic Review (e48625)
Janette Ribaut, Annette DeVito Dabbs, Fabienne Dobbels, Alexandra Teynor, Elisabeth Mess, Theresa Hoffmann, Sabina De Geest. .......... 347
An Introduction to Smart Home Ward–Based Hospital-at-Home Care in China

Weibin Cheng, MD; Xiaowen Cao, MSc; Wanmin Lian, MSc; Junzhang Tian, MD

Abstract

Hospital-at-home has been gaining increased attention as a potential remedy for the current shortcomings of our health care system, allowing for essential health services to be provided to patients in the comfort of their own homes. The advent of digital technology has revolutionized the way we provide medical and health care, leading to the emergence of a “hospital without walls.” The rapid adoption of novel digital health care technologies is revolutionizing remote health care provision, effectively dismantling the conventional boundary separating hospitals from the comfort of patients’ homes. The Guangdong Second Provincial General Hospital has developed a 5G-powered Smart Home Ward (SHW) that extends medical care services to the home setting and is tailored to meet the needs and settings of each patient’s household. The SHW was initially tested for its suitability for treating 4 specialized diseases, including cardiovascular disease, stroke, Parkinson disease, and Alzheimer disease. Understanding and addressing the potential challenges and risks associated with SHWs is essential for the successful implementation and maintenance of safe and effective home hospitalization.

(JMIR Mhealth Uhealth 2024;12:e44422) doi:10.2196/44422

KEYWORDS
smart home ward; telemonitoring; telemedicine; home care; hospital at home; healthcare delivery; implementation; smart ward; medical monitoring; medical care; rehabilitation; health care

The Future of Older Adult Care in China: Innovations in Health Care Delivery

China has emerged as a rapidly aging society. Aging contributes significantly to the health care burden in China due to the increased prevalence of chronic diseases and disabilities among older adults. The health care burden for the population of older individuals in China is further compounded by inadequate health care infrastructure and limited access to health care services in rural areas. The Chinese government has implemented various policies and initiatives to address these challenges, including expanding health care coverage, promoting preventative care, and increasing investment in health care infrastructure. Currently, the top-level design of China’s older adult care model is “9073,” which means that 90% of older people receive older adult care services at home, 7% of them receive short-term care in the community, and 3% of them receive institutional care. This implies that the age-friendly and livable environment at home affects 90% of the population of older adults.

Hospital-at-home, which seeks to provide essential health services to patients in the comfort of their own homes, has been gaining increasing attention as a feasible solution for at-home older adult care and medical services [1-3]. The COVID-19 pandemic has underscored the risks of overreliance on physical medical institutions, emphasizing the continued need to develop a decentralized medical service ecosystem that revolves around patients’ families and communities. Hospital-at-home programs have been implemented in many high-income countries for years [4-7]; however, successful programs are limited [8]. Experience translating from high-income countries to low- and middle-income countries can be challenging, where medical resources are limited and public health literacy is low [2,3,9,10].

The Growth of Emerging Technologies for Home Hospitalization

New digital health care technologies are being rapidly adopted for remote health care provision, which is breaking down the traditional barrier between the hospital and home. The emergence of 5G technology posed the potential to further enhance home hospitalization by enabling remote patient monitoring and real-time communication between health care providers and patients. The introduction of sensor-based wearables and devices has changed the way that clinical data are collected and stored, leading to groundbreaking advancements in how care is provided [11,12]. Digital biomarkers—a set of objectives, quantifiable measures of physiological and behavioral characteristics that are acquired via wearables, implants, and other devices—are becoming increasingly essential in this process [13,14]. Home hospitalization can greatly benefit from the use of digital
biomarkers, providing care teams with a more comprehensive understanding of a patient’s health. These biomarkers enable the tracking of vital signs such as heart rate and respiration rate, as well as changes in sleep patterns, activity levels, and dietary habits [12,15,16].

**Smart Home Ward for Hospital-at-Home Care**

In 2021, Guangdong Second Provincial General Hospital developed the “Hospital Intelligent Twins,” a 5G-powered smart hospital that integrates the Internet of Things (IoT), artificial intelligence (AI), cloud computing, and 5G applications to create all-scenario intelligence for health care and hospital management [17,18]. Leveraging this 5G smart hospital infrastructure, Guangdong Second Provincial General Hospital further explored the Smart Home Ward (SHW), which seeks to break through the “wall” of centralized hospital-based health care services by extending care services to the home setting (Figure 1). The SHW is a seamless hospital unit managed by health professionals in the hospital and is designed to provide patients with equivalent hospital-level services at home, including medical monitoring, ward rounds, consultation, and medical care, and so on. SHW can support 2 hospital-at-home models, early-supported discharge and admission avoidance, offering integrated health monitoring, medical care, and rehabilitation services all within the comfort of their homes [19,20].

**Figure 1.** Demonstration of the Smart Home Ward in the hospital.
The management of patients in the SHW is handled by an interdisciplinary care team overseen by responsible specialists from relevant clinical departments. At present, the Departments of Cardiology and Neurology are pilot testing SHWs for patients with cardiovascular conditions and those with cerebrovascular conditions, respectively. Each specialty establishes a dedicated home ward team comprising physicians from that department, along with nurses, pharmacists, physical therapists, occupational therapists, and social workers. For example, the post–coronary heart disease treatment team includes cardiologists, cardiac nurses, and rehabilitation therapists. Patients undergo an initial clinical assessment by their specialist physician to determine their suitability for remote home care. The inclusion and exclusion criteria for specific conditions such as coronary heart disease and stroke are provided in Multimedia Appendix 1. If deemed feasible based on the assessment and if the patient provides consent, the interprofessional SHW staff conduct a home environment evaluation. Necessary modifications are made and monitoring equipment or devices installed to safely support care at home with remote specialist oversight.

In the SHW, series of digital biomarkers are monitored, including blood pressure, heart rate, respiratory rate, blood oxygen saturation, body temperature, and electrocardiographic signals, with patients’ daily activities, treatment compliance, and risk occurrences monitored via contact or contactless devices. These data can be leveraged to tailor treatments or notify care providers of any deviations from expected parameters.

Data generated from the SHW are transmitted securely to the hospital’s electronic medical record (EMR) system in real time. The integration follows privacy and security guidelines set forth by our hospital’s Health Insurance Portability and Accountability Act (HIPAA)–compliant policies. Only deidentified data points that are relevant for clinical care, such as vital signs, activity levels, and medication adherence data, are integrated into the EMR. Data transmission from the SHW to the EMR is one-way to ensure the security of sensitive hospital information. Authorized care team members, including attending physicians, nurses, pharmacists, physical therapists, and case managers, have access to consolidated patient data reports within the EMR system. This allows them to monitor trends, recognize any deviations from normal ranges, and act accordingly without needing to use separate systems.

The primary care team, led by the patient’s attending physician, is responsible for reviewing the daily reports and contacting patients if follow-up care is needed, based on the remote patient monitoring data. Family members and caregivers participate in care conferences to stay updated on the patient’s progress but do not have access to the EMR. We aim to expand access to aggregated reports to allow for greater caregiver involvement while maintaining privacy and security.

A pretraining ward has been established in the hospital to help patients transit from hospital to home living. Patients are trained on the use of smart home devices and receive guidance on remote rehabilitation training. In addition, the pretraining ward includes medical assessments to ensure that the patient is eligible for admission to SHW care and will benefit from the treatments provided at home. Specialized staff consisting of nurses, physical therapists, and occupational therapists are dedicated to supporting the SHW program. These clinicians provide personalized patient education, demonstrations, and skills training. They ensure that patients and caregivers are comfortable with the remote monitoring system and therapy program prior to discharge. The goal is to maximize treatment adherence and outcomes through empowering patients and families with knowledge. A dedicated tech support team is also available to assist families in successfully setting up the in-home system and addressing any technical issues that may arise post discharge. Readily available support from clinical and technical experts further enables safe and independent living at home with remote care and monitoring.

**Characteristics of the SHW**

The SHW constructed in this project not only integrates advanced medical IoT technology and products, but also involves the renovation of the environment to make it suitable for older adults and patients, extending homogeneous medical services provided in the hospital wards to patients’ homes. Its characteristics are mainly reflected in the following 4 aspects.

First, the SHW has a multitude of functions including ambient sensing, medical monitoring, rehabilitation training, exercise and diet guidance, psychological counseling, and sleep management. To achieve these capabilities, a variety of cutting-edge technologies such as 5G or Wi-Fi 6, Internet of Medical Things (IoMT), smart wearable devices, smart home appliances, and health monitoring equipment have been deployed in the home setting. For instance, an intelligent closetool can detect urine levels and monitor heart rate, body fat, and the length of time spent on the toilet [21]. Additionally, an intelligent mattress and pillow monitor heart rate, respiration rate, and body movement frequency during sleep [22, 23]. Applications such as those for fall detection in the bathroom and the amount of stillness in daily activities can trigger an automatic alarm for an emergency, which connects to the hospital [24].

Second, to facilitate hospital-at-home management, a platform was developed with the integration of cross-system and cross-ecological IoMT devices (Figure 2; see the screenshot of the SHW Medical Management Information System). Thus, IoMT devices and ambient sensors are interconnected, and data are generated, gathered, managed, and processed by various terminals, including data from a hospital and a SHW (including data for the home environment, daily activities, treatment, rehabilitation, etc.). This enables interoperability for home ward management, allowing patients, health professionals, and caregivers to use mobile terminals such as tablets to view the patient’s physiological data in real time and manage the home ward setting with ease. As medical teams have the capability to monitor a patient’s health from a distance, fewer in-person visits are needed, thus decreasing the cost of care [25].
Third, deployment of an SHW is tailored to cater the needs and settings of a patient’s household. To promote patient autonomy in the home setting, the SHW undergoes various levels of renovations suitable for aging individuals, creating a barrier-free, safe, and secure living space. Furthermore, ambient sensors enable automatic configuration to ensure the comfort of the environment [26]. These sensors continuously detect the air, temperature, and humidity in the ward. Whenever the indoor concentrations of particulate matter of diameter ≤ 2.5 μm, formaldehyde, and carbon dioxide exceed standard levels, the SHW automatically adjusts its equipment to improve indoor air quality.

Fourth, the extension of hospital care services to the home setting is achieved through the integration of various technologies and resources. This challenge is addressed by the SHW through the deployment of advanced IoMT technology and products, which allow for the passive monitoring of patients’ health and the implementation of timely medical interventions. Additionally, a comprehensive EMR system has been implemented, enabling seamless sharing of patients’ medical records between the hospital and home health care systems. Furthermore, a robust 5G-based telemedicine infrastructure has been established, enabling real-time communication between patients and doctors, including digital consultations, remote monitoring, and telemedicine-enabled home visits, as well as providing on-demand home care services [27].
**SHW Care Plans for Specialized Diseases**

The SHW’s feasibility of treating 4 specialized diseases, namely Parkinson disease, cardiovascular disease, stroke, and Alzheimer disease, is being tested. For each disease, corresponding care plans have been developed, and necessary monitoring and treatment facilities have been suggested.

In the case of Parkinson disease, exercise rehabilitation treatment is mainly used as the primary treatment modality in the home ward. Patients are guided by doctors remotely to implement various forms of internet-based rehabilitation exercise training, such as relaxation training, joint range of motion training, muscle training, breathing training, gait training, balance training, and cognitive training. During exercise, the patient’s heart rate is monitored using a smart watch to prevent excessive exercise intensity. A smart lunch box is provided for daily monitoring of the frequency and amplitude of the patient’s hand tremors while eating to evaluate the effectiveness of rehabilitation treatment [28].

For patients with cardiovascular disease, the home ward caters to those with ST-segment elevation myocardial infarction, non–ST-segment elevation acute coronary syndrome, stable angina pectoris, ischemic cardiomyopathy, chronic systolic heart failure, and sudden cardiac death syndrome. Cardiac rehabilitation exercise therapy is guided by health professionals from the hospital, which includes aerobic exercise, resistance training, and neuromuscular training. A treadmill with heart and lung function monitoring is set up as a rehabilitation tool at home, which records and monitors the patient’s exercise status and heart rate, blood oxygen saturation, and calorie consumption. This allows doctors to adjust the program promptly [29].

For stroke, given that individuals who have experienced a stroke often face difficulties with mobility, the adaptation of living spaces to suit their needs is crucial to promote independent living and facilitate rehabilitation. To this end, we have implemented appropriate accessibility retrofits within the home setting to establish a barrier-free environment for patients. Such efforts can contribute to not only increased mobility but also heightened patient confidence. To evaluate the efficacy of functional rehabilitation, regular assessments are conducted through both home visits and hospital evaluations.

In the case of Alzheimer disease, a comprehensive treatment approach is adopted, including cognitive training, task training, and music therapy. These treatments are supported by the augmented reality and virtual reality applications that allow doctors to simulate complex scenarios to stimulate patients’ brain activities in a safe and controlled environment. Several physiological signals, such as heart rate variability, eye movements, and sleep patterns, are recorded and are used to train models for treatment evaluation. Furthermore, an AI-powered camera is installed in the home environment to monitor the patient’s activity status, such as sitting and lying time. A smart watch with GPS positioning and communication functions is also used to prevent patients from getting lost [30].

**Challenges for the Scale-Up of the SHW Care Program in China**

The widespread adoption of SHWs has been faced with several challenges. First, authoritative guidelines for the implementation of medical service standards for home ward care and those for setting up technology-enabled digital wards in China are lacking. Second, a sustainable fee-based model for home ward care has yet to be established. Third, the effective operation of home ward care requires close collaboration among hospitals, community health service organizations, and family members. However, there are currently no unified regulations to delineate the responsibilities of various stakeholders involved in providing home ward care. Fourth, the popularity and reliability of the technology needs improvement. SHWs require specialized equipment and technologies, a secure and trust-based environment, and staffing competencies to ensure patient safety and privacy. A unique standard needs to be established to ensure the interconnectedness and interoperability of various devices. Meanwhile, potential problems associated with the cost of equipment, access to necessary data, and data privacy also exist.

**Implications of Hospital-at-Home Care Using SHWs for Practitioners, Researchers, and Policy Makers**

The implementation of SHWs has the potential to extend the quantity and quality of hospital care services in response to the increasing demand for medical care in China’s aging society. For health care practitioners, the integration of smart home digital health technologies can facilitate remote patient monitoring and management, enabling timely and effective care delivery in patients’ homes. This approach can reduce the burden on hospitals and clinics, while improving patient outcomes and satisfaction. To enable the successful implementation of SHWs, new team structures with corresponding workflows must be created in clinical settings to optimize health care systems and patients’ usage of this technology. Clinical physicians must work closely with technicians in developing workflows and integrated AI tools and in the process of care provision, such as remote monitoring.

For researchers, the implementation of SHWs presents a unique opportunity to study the impact of digital health technology on health care delivery, patient outcomes, and cost-effectiveness. However, ethical considerations must be navigated when accessing the vast amounts of data gathered at SHWs to ensure patient privacy and security [31]. All stakeholders must understand and adapt to these implications to fully harness their potential benefits while mitigating associated risks.

Policy makers can benefit from the use of SHW technologies, as they show promising potential in reducing health care costs and improving access to care, particularly for patients in rural or underserved areas. However, policy makers must establish regulations to safeguard patient privacy and security, ensure equitable access to technology, and enhance digital infrastructure for widespread adoption. With careful consideration of these factors, the implementation of SHW can revolutionize
hospital-at-home care, improving the quality of care, reducing health care costs, and ultimately benefiting patients and the health care system as a whole.

**In Summary**

The advent of digital technology has revolutionized the way we provide medical and health care, leading to the emergence of “hospitals without walls.” This SHW is an innovative care model that promises to bring a host of improvements to health care by providing convenient access to “boundless ambulatory care,” as well as “boundless inpatient care.” Despite the potential benefits, these digital projects are still faced with some challenges, such as public acceptance and adoption of the technology, willingness to pay for services, and encouraging medical insurance uptake. Therefore, it is essential for governments to promote public awareness of the advantages of digital technology and introduce incentives that motivate people to take up medical insurance in order for these services to be widely available.

**Acknowledgments**

This study was funded by the National Key R&D (research and development) Program of China (number 2021YFC2009400).

**Authors' Contributions**

WC, XC, and WL conceived the study and drafted the manuscript. JT provided professional support and made critical revisions to the manuscript. All authors read and approved the final manuscript.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Inclusion/exclusion criteria for specific conditions.

[DOCX File, 18 KB - mhealth_v12i1e44422_app1.docx ]

**References**


**Abbreviations**

- AI: artificial intelligence
- EMR: electronic medical record
- HIPAA: Health Insurance Portability and Accountability Act
- IoMT: Internet of Medical Things
- IoT: Internet of Things
- SHW: Smart Home Ward
© Weibin Cheng, Xiaowen Cao, Wanmin Lian, Junzhang Tian. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.1.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Advances and Opportunities of Mobile Health in the Postpandemic Era: Smartphonization of Wearable Devices and Wearable Deviceization of Smartphones

Wonki Hong, PhD
Department of Digital Healthcare, Daejeon University, Daejeon, Republic of Korea

Corresponding Author:
Wonki Hong, PhD

Abstract
Mobile health (mHealth) with continuous real-time monitoring is leading the era of digital medical convergence. Wearable devices and smartphones optimized as personalized health management platforms enable disease prediction, prevention, diagnosis, and even treatment. Ubiquitous and accessible medical services offered through mHealth strengthen universal health coverage to facilitate service use without discrimination. This viewpoint investigates the latest trends in mHealth technology, which are comprehensive in terms of form factors and detection targets according to body attachment location and type. Insights and breakthroughs from the perspective of mHealth sensing through a new form factor and sensor-integrated display overcome the problems of existing mHealth by proposing a solution of smartphonization of wearable devices and the wearable deviceization of smartphones. This approach maximizes the infinite potential of stagnant mHealth technology and will present a new milestone leading to the popularization of mHealth. In the postpandemic era, innovative mHealth solutions through the smartphonization of wearable devices and the wearable deviceization of smartphones could become the standard for a new paradigm in the field of digital medicine.

(JMIR Mhealth Uhealth 2024;12:e48803) doi:10.2196/48803

KEYWORDS
mobile health; mHealth; smartphonization; wearable deviceization; new form factor; sensor-integrated display

Background
In the postpandemic era, the significance of mobile health (mHealth) has been highlighted, and explosive growth in this area is expected to continue [1,2]. Cutting-edge technologies are converging with health care, and mHealth, based on hyperconnected intelligence, is leading the paradigm shift in medical care [3,4]. Many countries have already entered a superaged society, and the proportion of gross domestic product expenditures for medical care is increasing due to an upsurge in the number of people with chronic diseases. In addition, the excessive demand compared to the available supply, the lack of health care infrastructure, and the unbalanced distribution of medical staff are also problems. Therefore, prediction, prevention, and management through artificial intelligence (AI)–based medical big data analysis are required, and for this purpose, ubiquitous and accessible medical services using personalized devices must be provided [5]. mHealth is a strong candidate to make this possible, and the ultimate goal is to dramatically improve the standard and satisfaction of living by providing quality services at affordable prices [6,7].

Wearable electronics and smartphones are representative types of mobile systems optimized for personalized health care sensing. As shown in Figure 1 [8-19], wearable devices that cover the human body and smartphones, a necessity for modern people, enable comprehensive health management in real time.
However, the pace of the development and popularization of mHealth technology is progressing more slowly than expected. From the perspective of a paradigm shift from the smartphonization of wearable devices and the wearable deviceization of smartphones, this viewpoint aimed to propose ways to unleash the potentiality of mHealth in the postpandemic era. The smartphonization of wearable devices and the wearable deviceization of smartphones do not simply mean that current smartphones become wearable devices and that current wearable devices maintain the functions of current smartphones. The smartphonization of wearable devices is to completely replace the smartphone function with a wearable device, while upgrading health care performance by embedding the current smartphone’s computational power and sensor-integrated display, including large-area panels and user interaction, in the wearable system. In addition, the wearable deviceization of smartphones refers to a change in the form factor so that health care sensing can be performed by switching from the current rigid form to a form that can be attached to a curved skin surface. The new form factor, which features both wearable computer and smartphone functions, will improve detection performance through large-area sensing and increase the penetration rate.

This viewpoint investigated recent trends in health care sensing methods using wearable devices and smartphones, which are
the central axis of mHealth. In the case of wearable devices, the form factor for each detailed location on the body and the corresponding detection target technology was described. In the case of smartphones, it covered the detection target and principles of health care according to the application of internal and external sensors, materials, and software. This viewpoint also analyzed the prospects of and current challenges in existing mHealth systems and considered new health care solutions using flexible displays for the convergence form factor of smartphones and wearables. The differentiating point was to consider the direction of mHealth from the perspective of a sensor-integrated and new form factor display. Ultimately, from a display perspective, solutions for the smartphonization of wearable devices and the wearable deviceization of smartphones will provide insight into the health care paradigm shift.

**Recent Progress in the Development of Wearable Electronics for Health Care**

The primary classification of wearable electronics based on the attachment position can be divided into the face, upper body, limbs, and whole body. Wearable clothes all over the body can also be classified separately.

**Face**

**Head**

The face, which is closest to the brain, is significant from a sensory point of view because it is where the 5 senses are concentrated. Face-wearable devices with various form factors, such as bands, caps, headsets, lenses, glasses, tattoos, mouthguards, and masks, may be distributed at each part of the head, eyes, nose, mouth, and ears to sense critical biosignals. In the case of the head, a wearable system that can analyze brain waves and psychological states can be applied [8,20,21]. Figure 2A [8] shows a wireless wearable electroencephalogram (EEG) measurement device based on a tattoo. AI can enhance decision-making by deep learning classification of received EEG data. Namely, it advances the decision performance of AI by feedback through brain waves. Additionally, it would be possible to grasp the degree of brain activation and mental condition of the frontal lobe and temporal lobe through the measurement of biosignals, such as brain waves.
**Figure 2.** Wearable devices attached to the face for mHealth. (A) Wearable EEG analysis platform with tattoo electrodes for EEG measurement and earbuds for wireless interaction. The images were reprinted from Shin et al [8,19]. (B) Stretchable corneal lenses for ocular electrodagnosis. The images were reprinted from Kim et al [19,22]. (C) Intraoral electronics for sodium intake analysis through wireless remote control. The images were reprinted from Lee et al [23]. (D) Sensing platform for gaseous CO2 real-time determination inside filtering face piece 2 (FFP2) facemasks. The images were reprinted from Escobedo et al [10,19]. EEG: electroencephalogram; mHealth: mobile health.

**Eyes and Nose**

System form factors worn on the eye may be divided mainly into lenses and glasses. In the case of lenses, eye health factors, such as glucose, intraocular pressure, and electroretinographic measurements, can be determined using noninvasive methods [9,22,24-26]. For example, a corneal sensor embedded in a disposable soft contact lens can be deployed for electroretinography based on electrochemical anchoring, as shown in Figure 2B [22]. These corneal lenses are functional sensors tailored for ophthalmic electroretinographic testing in human eyes via a user-friendly interface and a design that can be deployed noninvasively. Glasses for health care are prescribed by doctors as an auxiliary tool for surgery and can also analyze the electrolyte and metabolite content of sweat flowing from the head [27-29]. In addition, a wearable system placed on the nose in the form of a nose pad on the glasses can sense the pulse wave, respiratory rate, and electrooculographic measurements [30,31].
Mouth and Ears

Wearable electronics related to the mouth take the form of mouth guards, tooth sensors, and masts and can analyze saliva and nutrients and monitor air quality [10,23,32-35]. For example, a small stretchable circuit and sensor that can be inserted into the human oral cavity may be integrated into a breathable, flexible microporous membrane for a tissue-friendly design, as shown in Figure 2C [23]. Such a device may be used in research to study the prevention of hypertension by facilitating continuous quantification analysis of sodium intake. Figure 2D [10] shows a sensing platform for detecting gaseous CO$_2$ inside a face mask via stable inorganic phosphors whose luminescence is controlled by a pH indicator. A mask combining a battery-free printed near-field communication (NFC) tag and a photochemical sensor for noninvasive CO$_2$ measurement was used to achieve detection performance with a resolution of 103 ppm. Practicality in physical activity has been increased through the compensation of the temperature noise and characterized analytical specifications of measurement systems. Moreover, health care wearable systems attached to the ears use earbuds to perform heart rate and sleep monitoring functions [36,37].

Upper Body

In addition to the face, wearable systems can be applied to the neck, chest, abdomen, internal organs, back, and waist to extract significant health values.

Neck

In the case of the neck, wearable devices with a necklace and patch form factor can record an electrocardiogram (ECG) and voice pressure and monitor the diet through an electroglottogram (EGG) using a neckband [38-40]. For example, a neck-attached wearable device incorporating a cross-linked polymer film and hole-patterned diaphragm structure detects and quantifies voice with an excellent sensitivity of 5.5 V Pa$^{-1}$ over the voice frequency range, as shown in Figure 3A [39]. This device can be used for voice health management and security authentication by eliminating vibration distortions on the curved skin surface through excellent skin compatibility via using ultrathin profiles of $\geq 5$ µm.
Figure 3. Wearable electronics mounted onto the upper body. (A) Vibration-responsive patch for sensing voice pressure. The images were reprinted from Lee et al [19,39]. (B) Epidermal cardiopulmonary patch based on laser fabrication. The images were reprinted from Rachim et al [41]. (C) Air-silicon composite transducer (ASiT) for breathing pattern monitoring. The images were reprinted from Cotur et al [19,42]. (D) Spine tracker sensor system. The image was reprinted from Stollenwerk et al [19,43]. (E) A belt for waistline measurement. The images were reprinted from Nakamura et al [12,19]. EPE: electrophysiological electrode; MES: mechano-acoustic sensor; PCB: printed circuit board.

**Thorax**

Thorax-related wearable electronics, such as patches, chest belts, and brassieres, enable ECG recording, temperature measurement, sleep monitoring, posture analysis, and galvanic skin response (GSR) assessment [11,41,44-48]. Figure 3B [41] shows a sensor designed for continuous monitoring of the cardiopulmonary biosignal via a CO$_2$ laser-based manufacturing process. The epidermal patch consisting of a mechano-acoustic sensor and electrophysiological electrodes provides advanced functionality through a gas-permeable and biocompatible layer.

**Abdomen**

Abdomen-attached mHealth systems can sense glucose and breathing patterns through patches and straps [42,49]. For instance, an air-silicon composite transducer monitors respiratory activity by continuously measuring the force applied to the air channel embedded in the silicon-based elastomer, as shown in Figure 3C [42]. The system, which uses a pressure sensor and mixed-signal radio electronics, follows the principle of sensing the air pressure change inside the channel when breathing force is applied to the transducer surface. In particular, tactile sensing, including pressure sensing, is critical in health.
care. This is because tactile sensors attached to the skin detect physical stimuli, such as breathing patterns, heart rate, pulse, muscle activity, and body temperature, linked to biological signals. Skin, the most widely distributed organ among the five sense organs in the human body, is a tactile sensor with receptors that detect pressure, delicate movements, and temperature and is also an actuating organ that emits the same physical stimulation. Flexibility is a crucial element for the tactile sensor to be conformally attached to the skin to detect minute physical changes in detail and increase user convenience [50-52].

Furthermore, digestible pills check medication compliance. Management of medication adherence can prevent patients with severe mental illness from experiencing relapses and hospitalizations [53]. In addition, capsule endoscopy can monitor the colon health or bladder pressure state [54,55].

**Back**

A wearable system attached to the back can be used to analyze changes in the spine’s shape during training. A spine tracker device shown in Figure 3D consists of 5 sensors, with each sensor attached to the lumbar spine, and can correct posture by providing real-time feedback [43].

**Waist**

In addition, a waist belt can be useful for obesity management [12,56]. The belt automatically measures waist circumference with high accuracy, with an $F_1$-score of 0.95, and monitors the daily lifestyle using a magnetometer, an accelerometer, and a gyroscope, as shown in Figure 3E [12].

**Limbs**

In the case of the limbs, the main categories include the hands, arms, legs, and feet by attachment location.

**Hands**

The measurable health factors in a hand-related wearable device, such as a patch, ring, or glove, include rehabilitation evaluation analysis, ECG characteristics, oxygen saturation, dietary monitoring, pulse wave, and temperature [13,57-61]. For instance, a multisensory electronic skin integrated into a polyimide network simultaneously detects physical properties, such as temperature, strain, humidity, light, magnetic field, pressure, and proximity, in real time, as shown in Figure 4A [13]. It can also be used for rehabilitation evaluation using personalized intelligent prostheses.
Figure 4. Wearable devices attached to the hands and arms. (A) Stretchable and conformable electronic skin for multifunctional sensing. The images were reprinted from Hua et al [13,19]. (B) Power generation textile for wearable health care. The images were reprinted from Zhao et al [15,19]. (C) Stand-alone patch for health monitoring based on a stretchable organic optoelectronic system. The images were reprinted from Lee et al [62]. (D) Thermal patch for self-care treatment through temperature distribution sensing and thermotherapy based on wireless graphene. The images were reprinted from Kang et al [63]. (E) Sensor conformably attached to skin decoding epicentral human motions. The images were reprinted from Kim et al [19,64]. (F) A single wearable biosensor platform that simultaneously monitors sweat and interstitial fluid (ISF). The images were reprinted from Kim et al [16,19]. MEG: magnetoelastic generator; OLED: organic light-emitting diode; PDMS: polydimethylsiloxane; PI: polyimide; PPG: photoplethysmogram; PVA: polyvinyl alcohol; Temp.: temperature.

Arms

mHealth systems of various form factors related to the arm can also be useful for health management. Among them, wristwatches, bands, and bracelet devices can detect health factors, such as the heart rate, oxygen saturation, number of steps, blood pressure, ECG characteristics, glucose, blood sugar, and sweat metabolites [14,15,65-75]. Figure 4B [15] shows a magnetoelastic generator that provides the power to drive a wearable biosensor system. This generator can help measure cardiovascular parameters underwater without encapsulation for telemedicine and has excellent water vapor transmission characteristics.

A patch sensor attached to the arm can measure the pH, sweat rate, lactate, heart rate, temperature, electromyogram (EMG) and ECG characteristics, blood pressure, and water content and
can also be applied for wound treatment and rehabilitation evaluation [62-64,76-85]. For instance, a stand-alone organic skin patch for health care with an organic light-emitting display with sufficient pixels reports the heart rate via a stretchable photoplethysmogram (PPG) sensor, as shown in Figure 4C [62]. An ultrathin patch of 15 μm is configured on a soft elastomer substrate and can operate stably at 30% strain using a combination of a stress relief layer and deformable microcracks. Figure 4D [63] shows a wireless graphene patch that simultaneously provides thermal sensing and thermotherapy capabilities. This thermal patch consists of a graphene-based capacitive sensor, a graphene thermal pad, and a flexible wireless communication module to continuously monitor temperature changes with high resolution and sensitivity and perform thermal treatment through a graphene-based heater. Beyond the existing complex multisensor structure, skin patches alone may decode movements of 5-finger gestures by detecting microdeformation using the laser-induced crack structure, as shown in Figure 4E [64]. Based on the same principle, it can be attached to various body parts to track physical movements. Furthermore, ECG, EMG, temperature, sweat, and interstitial fluid analyses can be performed following health care monitoring through arm tattoos [16,86]. For instance, a noninvasive epidermal biosensing system includes physically separated electrochemical biosensors for the extraction of interstitial fluid at the cathode and sweat stimulus extraction at the anode, as shown in Figure 4F [16]. Namely, this biomarker monitoring system is a single wearable epidermal platform that simultaneously samples and analyses different biofluids.

**Legs and Feet**

Figure 5 describes a wearable health care device that may be applied to the legs, feet, or whole body. The mobile form factors applicable to the legs include patches, wearable robots, and straps, which perform moisture analysis at the wound area, gait analysis, ECG measurement, and rehabilitation evaluation [17,87-92]. For instance, appropriate dressing changes for exudative wounds are essential. Using a moisture sensor mounted on the bandage, as shown in Figure 5A [89], the change in the amount of dressing on the wound can be detected and the replacement time determined, increasing patient convenience. A motion capture device can accurately measure the movement of limbs during daily activities, strenuous exercise, and long-term exercise, as shown in Figure 5B [17]. Existing drift and instability problems are solved by integrating microtriaxis inertial and microtriaxis flow sensors. Additionally, it is possible to evaluate gait performance on irregular and uneven surfaces using a wearable sensor in the form of a strap with 6 inertial measurement units (IMUs) and an analysis algorithm, as shown in Figure 5C [92]. It is possible to implement edema measurement, gait analysis, and ulcer detection via plantar pressure analysis using wearable sensors attached to the shoes, socks, or soles of the feet [93-97].
Figure 5. mHealth apps for the legs and the whole body. (A) Moisture sensor for exudative wounds. The image was reprinted from Henricson et al [19,89]. (B) A motion capture device capable of detecting limb movements with high accuracy. The images were reprinted from Liu et al [17,19]. (C) Wearable strap sensor for gait analysis. The image was reprinted from Luo et al [19,92]. (D) An electronic textile conformable suit for distributed sensing wirelessly. The images were reprinted from Wicaksono et al [19,98]. mHealth: mobile health.

Whole Body
Furthermore, clothes worn on the whole body are also a type of wearable device. Figure 5D [98] shows a personalized and conformable suit of an electronics-based textile for multimodal health care sensing. The platform’s elasticity ensures intimate contact between the electronic device and the skin, and it can detect the skin temperature, heart rate, and respiration with high accuracy and precision. The suit with electronic textiles can measure the body temperature, respiratory rate, heart rate, oxygen saturation, and EMG and ECG characteristics and can also perform phototherapy [98-103]. As described before, form factor and detection targets by body part on wearable devices are summarized in Tables 1-4.
<table>
<thead>
<tr>
<th>Body position and form factor</th>
<th>Target(s) of detection</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Head</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tattoo</td>
<td>EEG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>[8]</td>
</tr>
<tr>
<td>Band, cap, headset</td>
<td>Mental stress through EEG</td>
<td>[20]</td>
</tr>
<tr>
<td>Band, cap, headset</td>
<td>EEG</td>
<td>[21]</td>
</tr>
<tr>
<td><strong>Eyes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lenses</td>
<td>Glucose</td>
<td>[24,25]</td>
</tr>
<tr>
<td>Lenses</td>
<td>Intraocular pressure</td>
<td>[9,26]</td>
</tr>
<tr>
<td>Lenses</td>
<td>Electroretinogram</td>
<td>[22]</td>
</tr>
<tr>
<td>Glasses</td>
<td>Auxiliary surgical tool</td>
<td>[27,28]</td>
</tr>
<tr>
<td>Glasses</td>
<td>Sweat electrolytes, metabolites</td>
<td>[29]</td>
</tr>
<tr>
<td><strong>Nose</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nose pad</td>
<td>Pulse wave, respiration rate</td>
<td>[30]</td>
</tr>
<tr>
<td>Nose pad</td>
<td>Electrooculogram</td>
<td>[31]</td>
</tr>
<tr>
<td><strong>Mouth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouthguard</td>
<td>Saliva monitoring</td>
<td>[32-34]</td>
</tr>
<tr>
<td>Mouthguard</td>
<td>Nutrition analysis</td>
<td>[23]</td>
</tr>
<tr>
<td>Tooth sensor</td>
<td>Nutrition analysis</td>
<td>[35]</td>
</tr>
<tr>
<td>Mask</td>
<td>Air quality monitoring</td>
<td>[10]</td>
</tr>
<tr>
<td><strong>Ears</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earbuds</td>
<td>Heart rate</td>
<td>[36]</td>
</tr>
<tr>
<td>Earbuds</td>
<td>Sleep monitoring using EEG</td>
<td>[37]</td>
</tr>
</tbody>
</table>

<sup>a</sup>EEG: electroencephalogram.
Summary of form factor and detection targets on wearable devices for the upper body.

<table>
<thead>
<tr>
<th>Body position and form factor</th>
<th>Target(s) of detection</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neck</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Necklace</td>
<td>ECG^[a]</td>
<td>[38]</td>
</tr>
<tr>
<td>Patch</td>
<td>Voice pressure</td>
<td>[39]</td>
</tr>
<tr>
<td>Band</td>
<td>EGG^[b]</td>
<td>[40]</td>
</tr>
<tr>
<td><strong>Thorax</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patch</td>
<td>ECG</td>
<td>[11,41,44]</td>
</tr>
<tr>
<td>Patch</td>
<td>ECG, temperature</td>
<td>[45]</td>
</tr>
<tr>
<td>Patch</td>
<td>Sleep monitoring</td>
<td>[46]</td>
</tr>
<tr>
<td>Chest belt</td>
<td>Trunk posture</td>
<td>[47]</td>
</tr>
<tr>
<td>Brassiere</td>
<td>Galvanic skin response</td>
<td>[48]</td>
</tr>
<tr>
<td><strong>Abdomen</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patch</td>
<td>Glucose</td>
<td>[49]</td>
</tr>
<tr>
<td>Strap</td>
<td>Respiratory patterns</td>
<td>[42,50-52]</td>
</tr>
<tr>
<td><strong>Internal organs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingestible pill/capsule</td>
<td>Medication compliance</td>
<td>[53]</td>
</tr>
<tr>
<td>Ingestible pill/capsule</td>
<td>Intravesical pressure and colon monitoring</td>
<td>[54,55]</td>
</tr>
<tr>
<td><strong>Back</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strap</td>
<td>Spine monitoring</td>
<td>[43]</td>
</tr>
<tr>
<td><strong>Waist</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belt</td>
<td>Obesity management</td>
<td>[12,56]</td>
</tr>
</tbody>
</table>

^[a]ECG: electrocardiogram.
^[b]EGG: electroglottogram.
<table>
<thead>
<tr>
<th>Body position and form factor</th>
<th>Target(s) of detection</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patch</td>
<td>Rehabilitation</td>
<td>[13]</td>
</tr>
<tr>
<td>Ring</td>
<td>ECG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>[57]</td>
</tr>
<tr>
<td>Ring</td>
<td>SpO&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>[58]</td>
</tr>
<tr>
<td>Ring</td>
<td>Dietary management</td>
<td>[59]</td>
</tr>
<tr>
<td>Ring</td>
<td>Pulse wave, temperature</td>
<td>[60]</td>
</tr>
<tr>
<td>Glove</td>
<td>Rehabilitation</td>
<td>[61]</td>
</tr>
<tr>
<td><strong>Wrist</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>Heart rate, step number</td>
<td>[65]</td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>SpO&lt;sub&gt;2&lt;/sub&gt;</td>
<td>[66]</td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>SpO&lt;sub&gt;2&lt;/sub&gt;, heart rate, energy expenditure</td>
<td>[67,68]</td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>Blood pressure</td>
<td>[14,69]</td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>Pulse management</td>
<td>[15]</td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>ECG</td>
<td>[70,71]</td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>Diagnosis of Parkinson disease</td>
<td>[72]</td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>Glucose</td>
<td>[73,74]</td>
</tr>
<tr>
<td>Watch/band/bracelet</td>
<td>Sweat metabolites (glucose, lactate)</td>
<td>[75]</td>
</tr>
<tr>
<td>Patch</td>
<td>Sweat rate, pH, lactate, glucose, chloride</td>
<td>[76]</td>
</tr>
<tr>
<td>Patch</td>
<td>Heart rate</td>
<td>[62]</td>
</tr>
<tr>
<td>Patch</td>
<td>Wound management</td>
<td>[77,78]</td>
</tr>
<tr>
<td>Patch</td>
<td>Temperature, thermotherapy</td>
<td>[63]</td>
</tr>
<tr>
<td>Patch</td>
<td>ECG, EMG&lt;sup&gt;c&lt;/sup&gt;</td>
<td>[79-81]</td>
</tr>
<tr>
<td>Patch</td>
<td>EMG</td>
<td>[82,83]</td>
</tr>
<tr>
<td>Patch</td>
<td>Blood pressure, skin hydration, temperature</td>
<td>[84]</td>
</tr>
<tr>
<td>Patch</td>
<td>Biometrics</td>
<td>[85]</td>
</tr>
<tr>
<td>Patch</td>
<td>Rehabilitation</td>
<td>[64]</td>
</tr>
<tr>
<td>Tattoo</td>
<td>ECG, EMG, temperature</td>
<td>[86]</td>
</tr>
<tr>
<td>Tattoo</td>
<td>Sweat and Interstitial fluid analysis</td>
<td>[16]</td>
</tr>
<tr>
<td><strong>Legs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patch</td>
<td>ECG</td>
<td>[87]</td>
</tr>
<tr>
<td>Patch</td>
<td>Moisture analysis at the wound area</td>
<td>[88,89]</td>
</tr>
<tr>
<td>Wearable robot</td>
<td>Rehabilitation</td>
<td>[17,90,91]</td>
</tr>
<tr>
<td>Strap</td>
<td>Gait analysis</td>
<td>[92]</td>
</tr>
<tr>
<td><strong>Feet</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patch</td>
<td>Edema</td>
<td>[93]</td>
</tr>
<tr>
<td>Shoes</td>
<td>Gait analysis</td>
<td>[94-96]</td>
</tr>
<tr>
<td>Socks</td>
<td>Foot pressure ulcer</td>
<td>[97]</td>
</tr>
</tbody>
</table>

<sup>a</sup>ECG: electrocardiogram.

<sup>b</sup>SpO<sub>2</sub>: oxygen saturation.

<sup>c</sup>EMG: electromyogram.
Table. Summary of form factor and detection targets on wearable devices for the whole body (clothes using electronic textiles).

<table>
<thead>
<tr>
<th>Target(s) of detection</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature, respiration, heart rate</td>
<td>[98]</td>
</tr>
<tr>
<td>SpO$_2^a$, heart rate, temperature</td>
<td>[99]</td>
</tr>
<tr>
<td>Phototherapy, temperature, heart rate</td>
<td>[100,101]</td>
</tr>
<tr>
<td>EMG$^b$</td>
<td>[102]</td>
</tr>
<tr>
<td>ECG$^c$</td>
<td>[103]</td>
</tr>
</tbody>
</table>

$^a$SpO$_2$: oxygen saturation.

$^b$EMG: electromyogram.

$^c$ECG: electrocardiogram.

Recent Progress in the Development of Smartphone-Based Health Care Apps

In addition to wearable devices, health care delivery is also possible using smartphones through built-in sensors, smartphone-interlocked gadgets, display-related materials, and apps.

CMOS Only

Smartphones have built-in 20-30 sensors; in particular, complementary metal-oxide-semiconductor (CMOS) image sensors may be used to monitor heart, eye, and skin-related diseases [18,104-109]. As shown in Figure 6A [18], the atrial fibrillation screening ability using PPG pulse analysis based on a smartphone camera and a commercialized app showed a similar performance level to that of patches used for single-lead ECG monitoring. It has been proven that prodromal stroke symptoms can be detected using only a smartphone in a primary care setting. In addition, the fingertip motion signal and color intensity signal, both heterogeneous signals, are acquired and analyzed using a camera to remove finger movement and optical noise, as shown in Figure 6B [104]. In this way, a clean heart rhythm signal with high accuracy can be extracted via smartphone monitoring, while minimizing noise artifacts.
Figure 6. Health care apps using built-in smartphone sensors. (A) Smartphone built-in camera and app-based atrial fibrillation diagnosis. The images were reprinted from Chan et al [18,19]. (B) Heart rhythm analysis using CMOS image sensor. The images were reprinted from Tabei et al [19,104]. (C) Smartphone-based blood pressure measurement through the oscillometric finger-pressing method. The images were reprinted from Chandrasekhar et al [19,110]. (D) Set and acquisition graph of smartphone and 3D-printed mouthpiece adapter for spirometry. The images were reprinted from Thap et al [19,111]. CMOS: complementary metal-oxide-semiconductor; PPG: photoplethysmogram.

Hybrid Including CMOS

New functions, such as blood pressure measurement and temperature and dietary monitoring, can be established by combining pressure sensors, temperature sensors, and the phone microphone instead of CMOS alone [110,112-115]. For instance, as shown in Figure 6C [110], absolute blood pressure is measured via a blood flow oscillometric signal through finger pressure using a strain gauge on the front of the smartphone, in addition to CMOS. A light-emitting display may also be added to this, so it is possible to measure blood pressure ultimately with pure smartphone components.

IMU/Microphone/Ultrasonic Sensor

In addition, sleep position monitoring and treatment can be performed by detecting body movements through an IMU of the smartphone, and the gait of patients with Parkinson disease can also be analyzed [116-118]. The smartphone’s built-in microphone sensor can also assess lung capacity and breathing sounds and monitor sleep [111,119-121]. Figure 6D [111] reports lung capacity and function parameter measurements.
following smartphone microphone–based, high-resolution time-frequency spectral analysis. A moisture-resistant ultrasonic sensor using polyvinylidene fluoride can be used for biometric authentication through fingerprinting [122].

**Touch Sensor/Digitizer**

Moreover, general user interfaces, such as a touch sensor and digitizer, can also be used for health care purposes. For example, the heart rate can be checked by assessing capacitance changes according to the heartbeat with a capacitive touch sensor. The touch sensor is also helpful in diagnosing Parkinson disease through touch accuracy analysis [123,124]. In addition, a digitizer for writing can be applied to biometric authentication through handwritten signature recognition [125,126].

**Interlocked Gadgets**

There is a case of combining various mHealth sensing techniques, such as pesticide analysis, otitis media diagnosis, malaria infection detection, and ECG measurement, by adding a separate gadget rather than using just the smartphone itself [127-133]. The platform shown in Figure 7A [128] performs a visual, quantitative analysis of pesticides using an optical system that combines a dark cavity and an ultraviolet lamp with a smartphone. In other words, integrating a smartphone and a gadget-based paper strip enables real-time and on-site food evaluation. Additionally, it was confirmed that the diagnosis of acute otitis media is possible with the same level of accuracy as that attained with existing otoscopes through the combination of a commercialized optical system and a camera in a smartphone, as shown in Figure 7B [130]. Figure 7C [132] shows a smartphone-based immunodiagnostic platform that performs a chemiluminescence-based enzyme-linked immunosorbent assay using a lyophilized chemiluminescence reagent. This hand-held point-of-care-testing analyzer can detect active malaria infections with a sensitivity of 8 ng/mL.
**Figure 7.** Health care apps using gadgets mounted on smartphones. (A) Smartphone platform for pesticide evaluation of food, integrated with an ultraviolet lamp and a dark cavity by 3D printing. The images were reprinted from Chu et al [128]. (B) Smartphone otoscope for diagnosis of acute otitis media. The images were reprinted from Mousseau et al [130]. (C) Smartphone-based immunodiagnosis using microfluidic assays. The images were reprinted from Ghosh et al [19,132]. (D) Antibacterial touchscreen for preventing contamination. The images were reprinted from Ippili et al [134]. (E) Digital biomarkers that reflect users’ moods, behaviors, and cognitions using text logs, browser history, human-computer interactions, and various sensors. The images were reprinted from Chen et al [19,135].

**Display Materials**

Health care delivery can also be achieved through materials used in manufacturing smartphones, such as window coatings for antireflection and display processes. An ecofriendly antibacterial coating with Zn-doped silicon oxide thin films can prevent infectious diseases caused by microbial contamination of touch events, as shown in Figure 7D [134]. In addition, it is possible to reduce the deformation of retinal cells by decreasing the blue light of the display through the material development of organic light-emitting or color filters [136].

**Apps**

Furthermore, health care sensing is possible through apps incorporating digital phenotypes and digital therapeutics [135,137-140]. A digital phenotype refers to a disease or health condition that is unintentionally reflected in patterns of use of digital devices. Mobile apps can collect human-smartphone interaction data to monitor smartphone usage and construct...
long-term patterns and trend changes. As shown in Figure 7E [135], analyzing a digital biomarker that reflects human effects, moods, behaviors, and cognition can predict psychiatric conditions, such as depression and smartphone addiction. In addition, digital therapeutics delivered through games, education, coaching, and counselling are based on cognitive behavioral therapy and can treat insomnia, alcohol addiction, drug addiction, panic disorder, and attention deficit hyperactivity disorder. Additionally, it effectively improves physical diseases, such as obesity and high blood glucose. Table 5 summarizes the sensing methods and targets using smartphones.

Table 5. Summary of sensing methods and targets using smartphones.

<table>
<thead>
<tr>
<th>Type and sensing methods</th>
<th>Target(s) of detection</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Built-in sensors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMOSa</td>
<td>Atrial fibrillation</td>
<td>[18]</td>
</tr>
<tr>
<td>CMOS</td>
<td>Heart rate</td>
<td>[104,105]</td>
</tr>
<tr>
<td>CMOS</td>
<td>Diabetic retinopathy</td>
<td>[106]</td>
</tr>
<tr>
<td>CMOS</td>
<td>Skin cancer</td>
<td>[107-109]</td>
</tr>
<tr>
<td>CMOS + microphone</td>
<td>Heart rate, SpO₂b, blood pressure</td>
<td>[112]</td>
</tr>
<tr>
<td>CMOS + microphone + speaker</td>
<td>Diet management</td>
<td>[113]</td>
</tr>
<tr>
<td>CMOS + strain gauge + display</td>
<td>Blood pressure</td>
<td>[110,114]</td>
</tr>
<tr>
<td>CMOS + temperature sensor</td>
<td>Temperature, heart rate</td>
<td>[115]</td>
</tr>
<tr>
<td>IMUc</td>
<td>Sleep monitoring</td>
<td>[116,117]</td>
</tr>
<tr>
<td>IMU</td>
<td>Gait analysis</td>
<td>[118]</td>
</tr>
<tr>
<td>Microphone</td>
<td>Spirometry</td>
<td>[111,119]</td>
</tr>
<tr>
<td>Microphone</td>
<td>Breathing sound analysis</td>
<td>[120]</td>
</tr>
<tr>
<td>Microphone</td>
<td>Sleep monitoring</td>
<td>[121]</td>
</tr>
<tr>
<td>Ultrasonic sensor</td>
<td>Biometric using fingerprint</td>
<td>[122]</td>
</tr>
<tr>
<td>Touch sensor</td>
<td>Heart rate</td>
<td>[123]</td>
</tr>
<tr>
<td>Touch sensor</td>
<td>Parkinson disease</td>
<td>[124]</td>
</tr>
<tr>
<td>Digitizer</td>
<td>Biometrics using signature</td>
<td>[125,126]</td>
</tr>
<tr>
<td><strong>Gadgets interlocked with smartphones</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optical platform</td>
<td>Pesticide evaluation in food</td>
<td>[127-129]</td>
</tr>
<tr>
<td>Smartphone CMOS + lens</td>
<td>Otoscopy</td>
<td>[130,131]</td>
</tr>
<tr>
<td>Microfluidic platform</td>
<td>Malaria infection</td>
<td>[132]</td>
</tr>
<tr>
<td>Patch electrode</td>
<td>ECGd</td>
<td>[133]</td>
</tr>
<tr>
<td><strong>Materials</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Window coating</td>
<td>Antibacterial</td>
<td>[134]</td>
</tr>
<tr>
<td>Light emitting</td>
<td>Blocking of blue light</td>
<td>[136]</td>
</tr>
<tr>
<td><strong>Apps</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital phenotyping</td>
<td>Addiction, attention deficit hyperactivity disorder</td>
<td>[135,137,138]</td>
</tr>
<tr>
<td>Digital therapeutics</td>
<td>Mental health</td>
<td>[139,140]</td>
</tr>
</tbody>
</table>

aCMOS: complementary metal-oxide-semiconductor.
bSpO₂: oxygen saturation.
cIMU: inertial measurement unit.
dECG: electrocardiogram.

**Prospects for mHealth**

The industry of mHealth is expected to grow explosively in the future. In particular, the third generation of medicine and therapies that rely on novel solutions are emerging beyond the existing state of mHealth. Among them, bioelectronic medicine is a nonpharmacological treatment category that stimulates nerve functions with energy, such as electricity, light, and
ultrasonic waves. This approach uses an electronic device that controls metabolic function to maintain homeostasis by regulating hormones [141]. To date, electroceuticals have been used for obesity, asthma, sleep apnea, brain tumors, epilepsy, and Parkinson disease and have shown substantial and significant therapeutic effects [142-144]. It is also one of the most innovative fields in medicine because it has significant advantages when considering the development time and cost of existing drugs.

Using digital therapeutics, also referred to as “software as a medical device,” it is possible to manage and treat not only physical diseases but also psychiatric conditions, such as posttraumatic stress disorder and schizophrenia [145,146]. It is of great significance in terms of patient convenience that personal and sensitive mental health conditions can be diagnosed in real life, not in hospitals, through digital phenotypic analysis, such as smartphone usage patterns and uploaded social networking service (SNS) content.

From the point of view of the wearable form factor, since much of health care sensing is possible on the wrist, the smartwatch is currently playing a pivotal role in health care. The finger (as well as the wrist) is a body part to focus on as it can be used to assess health factors, such as the heart rate, oxygen saturation, ECG characteristics, blood pressure, blood sugar, biometric authentication, body temperature, and dietary monitoring. Therefore, it is expected that in the future, the ring type of device for health care will pair with the smartwatch as the 2 main pillars.

**Challenges for mHealth**

This viewpoint investigated comprehensive health care sensing technology using wearable electronics and smartphones. However, mHealth is less widely used than expected, unfortunately. Wearable devices are relatively more optimized for continuous and real-time health care sensing compared to smartphones [147,148]. However, the penetration rate compared to smartphones worldwide is sluggish [149-151]. A smartwatch, a representative wearable device, needs to be connected to a smartphone to operate, so users do not recognize the wearable device as an independent entity. Independent use is required to be fully positioned as a separate device. These devices lack effectiveness due to reduced user convenience because of their small screens, poor battery performance, low usage rate, clunky design, and high price. Wearable devices are recognized as a kind of subdevice rather than an essential and leading product because they do not have as much impact as smartphones. Therefore, in the case of wearable devices, innovative solutions are required to make them universal necessities for human beings, such as smartphones.

However, in the case of smartphones, the penetration rate is high worldwide, including low- and middle-income countries [152]. In the case of current smartphones, the fundamental value in terms of user experience as well as utility is high. However, it is not such a great solution from the perspective of health care. It is challenging to conduct biosignal sensing using a smartphone while being in close contact with human skin all day long, so it is challenging to implement continuous real-time big data-based predictive and preventive medical care using smartphones from the health care perspective. Smartphones desperately require a breakthrough that can allow them to monitor health in real time continuously, 24 hours a day, through a form more closely adherent to the skin, while maintaining the current phone function.

**Breakthroughs for mHealth**

The display is a crucial component of a health care system. In other words, smartphones and wearable devices, as central axes of the mHealth system, are inseparable from their displays. In addition, displays and sensors in mobile devices are closely related. To improve the convenience of user interaction, the proportion of the active area of mobile displays is increasing. However, the increase in the active area has a limitation that reduces the sensing performance, including sensitivity. To overcome this, the upper part of the sensor covers the display by lowering the resolution of the display to prevent the deterioration of the sensing transmittance. A typical example is under-panel camera (UPC) technology that covers the camera with the display by reducing the display resolution on the top of the CMOS image sensor to increase light transmittance.

**Sensor-Integrated Display Solution**

However, the ultimate and ideal method is a sensor-integrated display solution. A sensor-integrated display has many advantages from a health care sensing point of view. This is because (1) many mHealth sensors use an optical approach, (2) it is relatively easy to manufacture large-area sensors, and (3) the application of a new form factor display can lead to an increase in the body contact area.

First, the majority of mHealth sensing approaches are optical methods. Various health care parameters, such as the heart rate, oxygen saturation, blood pressure, blood sugar, body temperature, environmental monitoring, and ECH characteristics, can be measured optically. A display is an optical system that already has the means to transmit light. Therefore, a sensor-integrated display could be an optimized health care solution. To implement health care devices using optical systems, in addition to optical transmitters, receiver systems must also be equipped. For advanced performance, the light-emitting wavelength band needs to be expanded and supplemented, including infrared as well as visible light, through the development of materials for the light-emitting layer.

Second, since the sensing area and detection performance are proportional, health care ability can be improved through a sensor embedded in a wide display area. It enables health care sensing in a large area over the entire display area when the built-in optical system is applied, considering design rules. In addition, it is more advantageous for wearability because of a reduction in volume due to the implementation of microlayer thickness because of the sensor-integrated display. Additionally, compared to the number of photomasks needed to manufacture a conventional display, the number of additional photomasks required to implement a display health care system with built-in sensors is far less. It can contribute to popularization due to the low manufacturing price according to the integral type.
Ultrathin, low-cost health care devices with relatively simple processes have significant benefits over conventional, bulky, and expensive wearable computers.

Finally, the new form factor device, such as a stretchable sensor-integrated display, increases the area of contact with the body and improves detection capability through health care sensing in close contact with the skin. Flexible panels with user convenience could be applied to the human skin, considering ergonomic factors [153-159]. The flexibility of not only the active matrix backplane and core of the panel but also the touch sensor, fingerprint sensor, and pressure sensor must be ensured, as shown in Figure 8A [153]. In a complete sensor-integrated display, the flexibility of the backplane allows the sensor part to gain flexibility naturally.

**Figure 8.** New form factor display and principle. (A) Wearable display with flexible and ultrathin active matrix backplane, touch screen panel, and fingerprint sensor components. The image was reprinted from Park et al [19,153]. (B) A flexible active matrix organic light-emitting diode (AMOLED) with large-area MoS2-based backplane for human skin display. The images were reprinted from Choi et al [154]. (C) A graph of compressive and tensile strength as the thickness increases in a single-layer structure (solid blue line) and laminated structure (dotted red line).

Furthermore, Figure 8B [154] shows a wearable full-color organic light-emitting diode (OLED) display using a 2D material–based backplane transistor suitable for complex skin shapes. The 18×18 thin-film transistor array was fabricated on ultrathin MoS2 film and then transferred to Al2O3 (30 nm)/polyethene terephthalate (6 μm), providing mechanical flexibility beyond conventional OLED technology.
New Form Factor Display

The left picture of Figure 8C simulates a multilayered display, and when this display is bent, tensile strength is applied at the top and compressive strength is applied at the bottom. Assuming that it is formed with only a single layer of the same thickness rather than a laminated structure, extreme tensile and compressive forces occur on the upper and lower surfaces, resulting in cracks in the display, as shown by the solid blue line in the right graph. However, in the stacked structure, a pressure-sensitive adhesive (PSA) between the display layers continues to create new neutral planes, as shown by the dotted red line. In response, the magnitude of the tension and compression force at the top and bottom surfaces does not increase, even if the thickness of the display increases. In other words, using the PSA, it is possible to implement a flexible display without cracks.

No part of the human body is flat. When the health care system and the skin conformally adhere, sensing performance improves. Display technology based on PSA with the harmony of creep and recovery characteristics induces form changes in wearable devices and smartphones. A new form factor with flexibility based on PSA technology that creates a new neutral plane will facilitate a critical conversion of the mHealth system.

Conclusion

Personalized platforms, such as wearable devices and smartphones, can be applied to AI-based disease prediction, prevention, and treatment. This viewpoint researched the latest technology trends in mHealth regarding form factors and detection targets according to body attachment location and type. In particular, the sensor convergence technology of the new form factor display provides a framework to analyze health factors in real time by conformally adhering a large-area system to the skin. Innovation in form factors in sensor-integrated displays and convergence health care solutions enable the smartphonization of wearable devices and the wearable deviceization of smartphones. In addition, the strategy for the smartphonization of wearable devices and the wearable deviceization of smartphones can accelerate the development of mHealth, realizing the democratization of medical care so that anyone can use essential services of high quality. Furthermore, it is expected to create a new milestone for the medical paradigm shift in the post-pandemic era.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT RS-2022-00165631). This research was also supported by Daejeon University Research Grants (2022).

Authors' Contributions

WH conceived the idea and concept of the viewpoint, created the figure sets, summarized the tables, and wrote the manuscript.

Conflicts of Interest

None declared.

References


19. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-01-11]


72. Phillips C, Liaqat D, Gabel M, de Lara E. Wristo2: reliable peripheral oxygen saturation readings from wrist-worn pulse oximeters. Presented at: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops); Mar 22-26, 2021; Kassel, Germany. [doi: 10.1109/PerComWorkshops51409.2021.9430986]


Abbreviations

AI: artificial intelligence
CMOS: complementary metal-oxide-semiconductor
ECG: electrocardiogram
EEG: electroencephalogram
EGG: electroglottogram
EMG: electromyogram
IMU: inertial measurement unit
mHealth: mobile health
OLED: organic light-emitting diode
PPG: photoplethysmogram
PSA: pressure-sensitive adhesive

© Wonki Hong. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.1.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
SOMAScience: A Novel Platform for Multidimensional, Longitudinal Pain Assessment

Chloe Zimmerman Gunsilius1,2,3*, BA; Joseph Heffner4*, PhD; Sienna Bruinsma1,5, BA; Madison Corinha1, BSc; Maria Cortinez2, BSc; Hadley Dalton6, BA; Ellen Duong6, MSc; Joshua Lu6, BSc; Aisulu Omar6, MSc; Lucy Long Whittington Owen1, PhD; Bradford Nazario Roarr6, BA; Kevin Tang7, BA; Frederike H Petzschner1,8,9, PhD

1Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, RI, United States
2Neuroscience Graduate Program, Department of Neuroscience, Brown University, Providence, RI, United States
3Warren Alpert Medical School, Brown University, Providence, RI, United States
4Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, United States
5Department of Neuroscience, Brown University, Providence, RI, United States
6Center for Computation and Visualization, Brown University, Providence, RI, United States
7Industrial Design, Rhode Island School of Design, Providence, RI, United States
8Department of Psychiatry and Human Behavior, Brown University, Providence, RI, United States
9Center for Digital Health, Brown University, Lifespan, Providence, RI, United States

*these authors contributed equally

Corresponding Author:
Chloe Zimmerman Gunsilius, BA
Robert J. and Nancy D. Carney Institute for Brain Science
Brown University
Carney Institute for Brain Science, 4th floor
164 Angell Street
Providence, RI, 02912
United States
Phone: 1 401 863 6272
Email: chloe_zimmerman@brown.edu

Abstract

Chronic pain is one of the most significant health issues in the United States, affecting more than 20% of the population. Despite its contribution to the increasing health crisis, reliable predictors of disease development, progression, or treatment outcomes are lacking. Self-report remains the most effective way to assess pain, but measures are often acquired in sparse settings over short time windows, limiting their predictive ability. In this paper, we present a new mobile health platform called SOMAScience. SOMAScience serves as an easy-to-use research tool for scientists and clinicians, enabling the collection of large-scale pain datasets in single- and multicenter studies by facilitating the acquisition, transfer, and analysis of longitudinal, multidimensional, self-report pain data. Data acquisition for SOMAScience is done through a user-friendly smartphone app, SOMA, that uses experience sampling methodology to capture momentary and daily assessments of pain intensity, unpleasantness, interference, location, mood, activities, and predictions about the next day that provide personal insights into daily pain dynamics. The visualization of data and its trends over time is meant to empower individual users’ self-management of their pain. This paper outlines the scientific, clinical, technological, and user considerations involved in the development of SOMAScience and how it can be used in clinical studies or for pain self-management purposes. Our goal is for SOMAScience to provide a much-needed platform for individual users to gain insight into the multidimensional features of their pain while lowering the barrier for researchers and clinicians to obtain the type of pain data that will ultimately lead to improved prevention, diagnosis, and treatment of chronic pain.

(JMIR Mhealth Uhealth 2024;12:e47177) doi:10.2196/47177
KEYWORDS
acute pain; acute-chronic pain transition; chronic pain; clinical outcome measurement; digital health; ecological momentary assessment; EMA; ESM; experience sampling methodology; mHealth; mobile health; pain management; pain self-management; patient reported outcomes; smartphone app

Introduction
More of us are in chronic pain than you might think—20% of adults in the United States reported pain on most or several days in 2019 [1]. This presents a substantial burden on society, costing up to US $635 billion annually [2]. In 2016, chronic back and neck pain alone accounted for the highest amount of US health care spending across 154 conditions, including diabetes and heart disease [3]. Moreover, pain is the leading cause of health care use across all illnesses [4-6]. Against this backdrop, it has never been more important to develop accurate pain symptom assessment and prediction methods to help patients, caregivers, and other stakeholders make informed decisions about treatment and care.

Accurately measuring pain is crucial for predicting an individual’s pain trajectory [7]. Methods to identify objective biomarkers of pain intensity [8-10] are still in their infancy and have yet to be proven effective in predicting future self-reported pain [11]. To date, the most common way to assess if someone is in pain is to simply ask them [12]. Typically, this is done using an 11-point pain intensity scale, where individuals are asked to rate their pain from 0 (no pain) to 10 (worst pain imaginable) [12-14]. Since the 1980s, this simple pain intensity scale has played a significant role in the clinical assessment of pain by enabling defined targets for pain management and the dosing of pain-relieving medications [15,16]. If a person consistently reports pain intensity as more than 3 out of 10 for more than 3 months, the scale becomes part of the diagnostic criteria for chronic pain [17]. For patients, self-reporting their pain on the scale can validate and quantify their pain experience, leading to improved shared decision-making and enhanced communication with health care providers [18].

Despite its ease of use, the pain intensity scale has not led to significant advancements in pain management or patient satisfaction [18-23]. When measured in medical settings, people tend to over- or underreport their pain intensity depending on difficulties with recall, expected treatments, care standards, or other subjective factors such as mood [24,25]. Additionally, regulatory approval for pharmaceutical companies to promote “titration to effect” practices, whereby physicians were encouraged to increase opioid doses to achieve continued reductions on the pain intensity scale, contributed to opioid overprescribing that fueled the opioid epidemic [23,25]. These limitations highlight the need for more nuanced self-report measures of pain.

In recent years, clinical approaches to pain have sought to better assess the multidimensional experience of pain from a biopsychosocial perspective [12,26]. Multidisciplinary and individualized assessment and treatment of pain with both pharmacologic and nonpharmacologic interventions is now considered the ideal way to treat both acute and chronic pain [27]. To overcome the limitations of the unidimensional pain intensity scale, expert panelists at the Food and Drug Administration (FDA) and National Institutes of Health (NIH) have established “core outcome sets” of multidimensional questionnaires for use in research studies [13,28,29]. Dimensions assessed include pain, unpleasantness, interference, and impacts on mood and activity [25,30]. However, completing multiple questionnaires is time-consuming, and therefore not ideal for daily longitudinal studies. As a result, a large barrier remains in the ability of researchers to collect comprehensive, multidisciplinary pain data sets [31]. It is therefore imperative in pain research to implement a reliable method for multidimensional pain measurements that accurately captures the most important dimensions of pain symptoms and treatments as they evolve over time in the context of people’s daily lives.

A comprehensive approach to pain assessment necessitates the acquisition of both deep and wide pain data. Deep data involve in-depth evaluations of the multidimensional aspects of pain within individuals over extended periods, while wide data refer to data sets that encompass a large number of individuals across different demographic factors like age, geographic location, race, ethnicity, and socioeconomic status, as well as across different pain diagnoses (eg, arthritis, fibromyalgia, and postsurgical pain). The acquisition of deep data enables a deeper understanding of the mechanisms that trigger and sustain pain in individuals, while wide data provide the foundation for generalizing findings and developing biomarkers for pain persistence or recovery. Smartphone apps can provide large-scale platforms for data collection while also helping users track their daily symptom experience [32]. Such digital tools provide a promising solution for acquiring deep and wide data sets that enable new behavioral and scientific insights into the dynamics and evolution of pain.

Here, we introduce a novel mobile health (mHealth) platform for longitudinal pain assessment, called SOMA [Science]. This platform has been conceptualized to comprehensively capture multiple facets of pain through its smartphone app, SOMA. Unlike conventional pain assessment tools that focus primarily on daily pain intensity, SOMA offers an enriched multidimensional pain assessment. This includes factors like daily pain intensity, unpleasantness, interference, mood, expectations, and activities, based on experience sampling methodologies (ESMs) [33]. Our choice of measures is in accordance with the current Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations for pain assessment [28,34]. In addition, the design and validation of the SOMA app have been executed following the guidelines set forth by the American Psychiatric Association (APA) [35,36].

SOMA was designed with users in mind, offering a free and user-friendly tracking feature that facilitates real-time tracking of pain, medications, and treatment regimens. A “Trends” section distills the multidimensional data to visualize pain trends.
This feature aids users in self-managing their pain, recognizing patterns, and discerning between effective pain management and areas needing improvement. With these insights, users can communicate more efficiently with their health care providers, positioning SOMA as a valuable “companion” tool to be used alongside standard medical advice.

SOMA Science has been engineered to provide researchers with the ability to carry out independent single- and multicenter studies. Its pain assessments can be used as end points in research studies and within clinical practice. To support expansive research initiatives, the SOMA Science platform ensures seamless transfer of app-sourced data to researchers and affiliated institutions and offers open-source code to streamline data preprocessing.

Contemporary pain research standards underscore the significance of multidimensional pain assessments through established batteries of pain questionnaires [13,29]. However, we identified a scarcity of available tools that effectively capture such assessments in a format meaningful for researchers aiming to develop better measures of pain outcomes or symptom burden over time. While there are smartphone apps, like Manage My Pain and Pain Scale-Pain Tracker App, they primarily cater to users as self-management tools [37,38]. The foundational architecture of such apps and the nature of the data they procure do not typically align with the specifications researchers need for conducting or analyzing comprehensive, large-scale studies that meaningfully contribute to pain research. A further limitation is the proprietary nature of many of these apps, which can impede data quality, accessibility, and transparency.

To our knowledge, no other tools currently exist that are specifically designed to capture multidimensional and longitudinal pain metrics in compliance with the recommended standards and are simultaneously tailored for extensive open-source academic research. This distinctiveness sets SOMA Science apart as an unparalleled mHealth platform.

In this paper, we discuss the choices and technological considerations for the development of SOMA Science as well as the scientific rationale behind the selection of measures. Our aim is to outline how SOMA Science can be used by scientists and institutions to acquire large-scale, longitudinal, multidimensional pain data in single and multicenter studies in order to gain new insights into pain that will benefit patients.

Section 1: SOMA Science Platform

Overview

SOMA Science aims to address the current shortage of platforms for acquiring both deep (multidimensional and longitudinal) and wide (cross-spectrum and large-scale) pain data to create novel insights into the dynamics and evolution of acute and chronic pain. The term SOMA stems from the Greek word σώμα (meaning body or entire person), signifying that it takes a holistic approach to pain.

SOMA Science represents the combination of 3 branches: the SOMA smartphone app used for data acquisition, an application programming interface for data transfer, and open-source analysis code distributed through GitHub [39] (Figure 1).

Figure 1. Schematic overview of the SOMA Science platform. Data will be acquired using the SOMA app, a user-friendly smartphone app available on Google Play and the Apple App Store. Encrypted data are sent to our application programming interface (currently located at Brown University) and shared in a study-specific manner with individual researchers and institutions. To facilitate data analysis, we created a GitHub repository where researchers can download, modify, or even create new versions of our template scripts for data preprocessing and certain analysis techniques through GitHub. ESM: experience sampling methodology; IRB: institutional review board.

To request to run a study through the platform, researchers need to submit a research inquiry detailing the study purpose on the SOMA website [40]. Data for SOMA Science are acquired through the SOMA app, which is freely available on Google Play and the Apple App Store.
Play [41] and the Apple App Store [42] and can be found by searching for “SOMA Pain Manager.” Anyone is able to download and use the app, regardless of whether they are participating in a research study. For associating people’s app data with a specific study, the researcher will be assigned a unique study ID (1 per study) and a list of individual app IDs (1 per expected participant), which need to be sent to the study participants. Participants can then install and register on the SOMA app and enroll in a specific study using the study ID and unique app ID (instruction videos on [43]: Figure 2). This use of study and app IDs allows the assignment of individual participants’ data to single and multicenter studies.

**Figure 2.** Schematic overview of study enrollment on the SOMA app. Study participants can download and register on the SOMA app directly using the links to Google Play or the Apple App Store or by searching for “SOMA Pain Manager.” Inside the main menu of the app, they can sign into a specific research study using a pre-sent study ID and App ID. They are shown a copy of their institutional review board–approved study-specific informed consent form for their records. Encrypted research data on SOMA will be sent to our database and transferred to the researchers of each individual study.

Upon enrollment, participants will be sent regular reminders through the app to fill out short pain surveys (details about the ESM and data content are in the following sections). At the time of publication, assessments are restricted to the features listed below. Future releases may offer the option to request additional features and questions. After each assessment, encrypted data are transferred to our application programming interface, stored on an actively managed secure database (see “Maximizing Privacy and Security” in Multimedia Appendix 1), and then shared with researchers from individual studies (Figures 1 and 2).

**Implementation of APA Guidelines in SOMA**

The APA’s app evaluation model stands as a notable benchmark for evaluating the suitability of health-related smartphone apps intended for patient populations [35]. It delineates 5 pivotal criteria to assess apps: accessibility, privacy and security, clinical foundation, engagement style, and therapeutic goal [44]. To keep pace with the rapidly advancing field of health apps, the system is regularly updated and refined [45]. Stemming from this APA model is a comprehensive database [46], which facilitates app evaluations across the 5 core domains [47]. Such initiatives are crucial in establishing public-facing, user-friendly standards for health apps and ensuring the development of safe and efficacious apps that benefit users [48-50].

In the development of SOMAScience, we have deeply integrated the cardinal principles of the APA model. Recognizing the emphasis that this model (and other akin evaluation frameworks) places on robust privacy, security, usability, and clinical foundations, we meticulously factored in specific technological elements during the app’s creation [50]. For readers interested in the technological nuances and our dedicated approaches to privacy and security aligned with the APA guidelines, we direct you to Multimedia Appendix 1 [51-60].

**Development of SOMAScience**

A multidisciplinary team at the Psychiatry, Embodiment, and Computation Lab at Brown University, comprising academic neuroscientists, psychologists, pain physicians, software engineers, and designers, collaboratively conceptualized and initialized the early design and features of the SOMA app. The SOMAScience platform is supported by the Brainstorm program at the Carney Institute for Brain Science at Brown University, a new program to accelerate the translation of computational brain science into real-world applications that benefit patients, the scientific community, and society as a whole.

**Incorporation of Patient and User Feedback**

Any app meant for long-term use must provide a simple, user-friendly interface tailored to its target audience. With this in mind, we actively sought initial feedback from individuals experiencing chronic pain to shape our app’s delivery structure. Through comprehensive one-on-one Zoom (Zoom Video Communications, Inc) discussions with a select patient group (n=4), representing a diverse pain spectrum including migraines, postsurgical musculoskeletal pain, cancer pain, and inflammatory bowel disease, we gleaned insights into our early app prototypes. These discussions involved a detailed
walkthrough of each interface, where patients aired their thoughts in a guided interview.

Drawing from this feedback, we refined the “Check-In” interface’s design and flow. A common sentiment among initial users was the importance of daily tracking of pain and mood fluctuations in tandem with daily activities. To accommodate this, we introduced the “Trends” screen, a visual tool designed not only to foster self-management and a deeper understanding of pain dynamics but also to facilitate effective communication with health care providers. To further enhance the app’s utility, we incorporated screens to monitor various treatment modalities and transitioned the mood and emotion tracking interfaces to use intuitive visual analog scales (VASs) in lieu of a 2D rating system.

Responding to the patient’s desire for a more personalized experience, we introduced an interactive chatbot during the onboarding phase. This chatbot briefly engages users, gathering foundational demographic details and an introductory snapshot of their pain experiences. Existing studies vouch for the efficacy of chatbots in extending support to people with chronic pain [61], making this an evolving component with forthcoming features in SOMA that focus on pain interventions.

Our iterative refinement strategy incorporated a beta-testing phase. Initially, 30 internal testers actively engaged with the app daily across a gamut of devices and operating platforms, enabling us to identify and correct technical glitches and enhance the user experience, especially regarding the “Trends” data visualization. To expand our feedback, SOMA was then shared with a larger patient interest group (over 250 people with chronic pain), leading to critical refinements and the inclusion of user-suggested enhancements. This ongoing feedback mechanism ensures the continuous improvement and evolution of SOMA.

Section 2: Pain Data Acquisition Through SOMAScience

Deep Data Acquisition Using ESMs

Overview

The SOMA app uses ESMs to gather multidimensional and longitudinal pain data for SOMAScience. ESMs, also known as ecological momentary assessments, provide real-time, self-report data about individuals’ thoughts, feelings, and experiences (eg, “How do you feel right now?”) in the context of people’s daily lives [33]. Previous ESM studies on pain have shown high completion rates (>85%) and demonstrated the feasibility of using these daily self-reports for pain [62], in line with findings about the high completion rates of mobile-delivered ESM studies in general [63].

ESMs have several benefits over traditional self-report measures. First, they offer real-time data that are less prone to recall bias, allowing for the capture of critical experiences that might be missed by retrospective long-term self-report measures [64-67]. Second, ESMs can capture contextual information about an individual’s thoughts and experiences, such as knowing what activities a person engaged in when they experienced pain [68]. Thus, they provide insight into the longitudinal dynamics of multidimensional aspects of pain in people’s natural ecological environments across time and context [69,70]. This enables the generation of rich data sets that could be used to identify candidate behavioral “biomarkers” or “assays” that predict transitions in disease states based on self-report alone [68,71]. For example, there is preliminary evidence that longitudinal measures of pain can predict acute pain state transitions after surgery [72] and identify treatment response time courses in patients with chronic pain [73].

Momentary, Situational, Retrospective, and Prospective Assessments in SOMAScience

One limitation of existing ESM studies is that they typically solicit several short, momentary reports throughout the day [74]. While this approach reduces bias in pain reports resulting from memory recall or pain beliefs, it may still miss important short-term pain dynamics, such as flare-ups, and fail to assess the role of expectations in the development and treatment of pain [75]. To address this limitation, SOMAScience uses a multifaceted approach, which includes 4 daily assessment types on the SOMA app: momentary assessments (called random check-ins), voluntary self-initiated entries (called quick check-ins), and both retrospective and prospective assessments or coverage assessments [75] (which are both part of an evening routine at the end of the day; Figure 3).
Random and quick check-ins capture various aspects of mood, activities, pain, and pain location and can be completed in less than 30 seconds. Quick check-ins can be performed at any time, for example, during or shortly after a flare-up. Random check-ins reflect classical ESM assessments and only occur during randomly selected moments within a specific time window (eg, 3 checks per day between 8 AM and 6 PM). Users receive notifications on their phone when the random check-in is available and have the option to snooze the notification for a predefined time window (eg, a maximum of 60 minutes).

The evening routine assesses recall of pain, mood, activities, and any pain-related treatments over the past day (retrospective), as well as predictions of pain, mood, and activities for the next day (prospective). This routine is available during a prolonged, preset, but fixed time window at night (eg, 6 PM-11 PM) to promote habit formation that increases the likelihood of long-term app use. It takes less than 3 minutes to complete.

**Longitudinal Assessments with SOMA Science**

Pain is inherently dynamic, fluctuating not just daily but hourly and even on a minute-to-minute scale, even without changes in physiological markers [18,47,62]. Traditional methods, which measure pain intensity sporadically during clinical trials or medical visits, might not capture a patient’s holistic pain experience due to their limited assessment windows.

While some studies aim for more granular pain assessments, they often focus on brief periods. A systematic review found the median duration for ESM studies in pain to be 14 days and a mere 7 days for general mHealth ESM studies [40,61]. Such short durations can overlook pivotal phases, like the transition from acute to chronic pain over 3-6 months. Consequently, the role of self-reported pain dynamics in acute pain recovery or its evolution to chronic pain remains ambiguous.

The SOMA app is designed to bridge this gap. Its check-ins are concise, using straightforward design principles for ease of use. The chosen metrics cater to diverse pain experiences, facilitating collaboration and data sharing among researchers. Moreover, the app’s “Trends” feature empowers users to track their pain, treatments, activities, and mood over diverse timeframes (ie, weekly, monthly, and annually). This aids in providing users with a deeper understanding of their pain journey, ultimately supporting more effective self-management.

**Wide Data Acquisition Using Smartphones**

While large data sets on repeated multidimensional pain ratings beyond intensity alone are still few and far between, smartphones offer a unique opportunity to expand data acquisition beyond classical experimental settings [74]. Smartphone access has increased tremendously in the past decade (84% of US households reported owning at least 1 smartphone) [76]. Data acquired remotely through smartphone apps facilitate large-scale, real-world studies without the constraints of traditional laboratory studies. The results of such pragmatic studies are more generalizable than highly selective traditional randomized controlled trials [77-79]. SOMA Science was built to allow remote monitoring of longitudinal symptoms and treatments to maximize high-quality data in large-scale pragmatic studies. To further facilitate a much wider array of user input, the SOMA app is compatible with both Android and iOS devices, meaning anyone in the United States with a smartphone can use it.

Smartphone-based pain assessments offer a solution to the limited geographic, economic, and cultural diversity in traditional pain studies. Smartphones are pervasive, even in low-resource [80] or rural areas [72,81], where almost half of the world’s smartphone owners live [82]. They are also widely used by older adults [83], who are often left out of laboratory-based pain studies. Additionally, there is a need to consider how pain and its treatment vary across racial, ethnic, and cultural backgrounds for comprehensive care. Even within a specific culture, there are important differences in how pain is experienced and treated across different racial and ethnic backgrounds that need to be accounted for to deliver the best pain care [84-86]. Upcoming translations of the app into languages like Spanish and German, in collaboration with experts familiar with the culture, will further diversify data and
Section 3: Data Content

Overview

With the rising number of health-focused smartphone apps, there is also a growing need for transparency in the selection of measures for the app. Here we briefly detail the process of selection for each measure in SOMAScience as assessed through the SOMA app, its basis in the scientific and clinical literature, and what gaps it was chosen to address. The goal is to provide transparency in the design and selection process to facilitate the development of research studies using SOMAScience. Specific measures may be refined over time with user feedback and as scientific studies using SOMAScience identify areas for improvement.

Measuring Pain Intensity, Unpleasantness, and Interference

Multidimensional pain assessment is a crucial aspect of clinical practice as it helps determine the effectiveness of treatment and recovery. The SOMA app assesses the pain intensity scale in addition to pain interference and pain unpleasantness to provide a more holistic measure of daily pain symptom load [87]. Participants are asked to rate each pain question on a sliding VAS, providing a simple and comparable assessment of daily pain experience (Figure 4). This operationalizes a practice advocated in the International Statistical Classification of Diseases and Related Health Problems for providers to use a composite rating of pain intensity, distress, and interference of pain to determine overall burden [88]. The composite pain score provides a multidimensional solution that balances the limitations of the unidimensional pain scale while still being easily quantifiable. The ability to analyze the 3 measures individually and as a composite score helps identify meaningful individual variability, enabling specific predictions between individuals. For example, pain unpleasantness could be more predictive of future pain in individual A, while pain interference may be more important for individual B’s long-term outcomes.

Figure 4. Upper panel: example screens on the SOMA app for check-ins and trends. Lower panel: example screens on the SOMA app from Quick Check, Random Check, and Evening Routine. From left to right: assessing pain on a visual analog scale (VAS), indicating pain location, assessing mood on a VAS, emotion selection, and activity selection.

Importantly, SOMA’s 3 pain questions were chosen because they are directly comparable to results from established pain questionnaires, such as the Brief Pain Inventory [89] or the McGill Pain Inventory [90]. They also satisfy the standards set by major scientific and regulatory bodies, such as the IMMPACT recommendations, the NIH Helping to End Addiction Long-term initiative, and the FDA guidelines for assessing multidimensional components of pain [13,28,29,87]. In this way, SOMA’s multidimensional pain assessment of intensity, unpleasantness, and interference can provide important supplemental measures that are directly comparable to established clinical benchmarks and standards of care. This is...
critical for researchers looking to establish and validate novel pain biomarkers or end points.

**Measuring Pain Locations**

Pain localization is an important aspect of pain assessment. Conventional methods of measuring pain location in medical appointments and research studies involve having individuals indicate it on a body map, such as the Brief Pain Inventory [89], the McGill Pain Questionnaire [91], or the Michigan Body Map [92]. This approach can pinpoint differences in peripheral and central pain pathology based on the localization and stability of pain representation over time. For instance, nociceptive or inflammatory pain is usually precisely localized somatically and does not change much over time, while neuropathic or chronic primary pain is often experienced in multiple bodily locations, radiates, or changes over time [93]. More recent methods of digital quantification, like the ones used on the SOMA app, have established the reliability and validity of body maps for pain assessments [94,95]. Interactive body maps delivered through digital or tablet apps are more effective than traditional paper or laptop assessments [92,96]. Yet a review of smartphone apps that use the body map for tracking pain found that few actually quantified the location ratings or provided any summary feedback [97]. The SOMA app’s interactive body map offers 46 different discrete location options on the front and back of the body that participants indicate in every daily check-in (Figure 5 and Multimedia Appendix 1). The use of discrete points ensures uniformity across devices and accounts for differences in participants’ finger size or dexterity. The “Trends” section of the SOMA app displays the body map with the percentage of times a location has been selected, enabling users to visualize the frequency of pain at a given location. For participants who experience nonspecific, difficult-to-localize, or widespread pain, such as fibromyalgia, there is an additional option to indicate “My pain is everywhere” on the body map.

**Figure 5.** Pain map included on the SOMA app covering 46 discrete pain locations.

**Measuring Interventions**

The treatment of pain has been incredibly difficult to get right. The newest clinical guidelines advocate the use of multimodal, multidisciplinary approaches [27,98,99]. Such approaches emphasize a combination of pain treatments that include medications, restorative therapies (eg, physical therapy), interventional procedures (eg, epidural injections), behavioral interventions (eg, cognitive behavioral interventions), and complementary and integrative medicine (eg, acupuncture). Combinations of these therapies have been associated with the best long-term pain outcomes [100,101] and satisfy a biopsychosocial approach to pain [102].

It can be challenging for individuals and providers to determine which treatments are most effective for them, as the effects of many treatments for pain may not become apparent for weeks or even months (eg, cognitive and physical interventions, certain medications, and surgery) [103,104]. In determining how to measure treatments through the SOMA app, we followed the recommendations of the 2019 Department of Health and Human Services Pain Strategy [27].

In the case of medications, many different pathophysiologic mechanisms are targeted with different classes of medications. The use of different medications often changes over time, so we designed the medication screen on the SOMA app to be able to capture such changes. We, therefore, included 20 options across the main classes of pain medications for both acute and chronic pain on the SOMA app, detailed further in Multimedia Appendix 1.

A second treatment screen includes the recommended nonpharmacologic approaches to pain. Combinations of these treatments are often used by a single person over time to target different pain mechanisms [105]. SOMAScience currently provides the ability for people to track up to 20 different nonpharmacologic therapies across these 4 major classes, detailed in Multimedia Appendix 1. SOMAScience’s broad treatment tracking capabilities therefore facilitate the type of
wide data needed to understand differences in treatment use across users.

**Measuring Emotions**

Emotion and pain are fundamentally related. Definitions of pain acknowledge that pain is partially an emotional experience [106,107], yet few researchers would reduce pain down to just another emotional state such as sadness or happiness. The complexity of the pain-emotion relationship is highlighted by reviews of the neural circuits of each construct, showing both shared and functionally dissociable brain regions [108,109]. Unfortunately, the theoretical and empirical understanding of how emotion and pain are connected is limited, as much previous research only focuses on cross-sectional correlations between the 2 constructs [110,111].

Despite the renaissance of emotion research since the 1960s [112], emotion researchers continue to disagree on what constitutes an appropriate emotion measure [113]. The discrete emotion perspectives suggest that specific emotions such as anger, fear, happiness, sadness, disgust, and surprise are special kinds of biologically distinct responses associated with unique behavioral, physiological, and experiential correlates [114,115]. The dimensional perspectives consider emotions to be organized along a set of common dimensions such as valence (unpleasantness) and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of unpleasantness and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states. 

**Measuring Activities**

People’s experiences of both pain and emotion are intrinsically tied to the activities they are engaging in each day. There have been many large-scale studies investigating the relationship between emotions and behavior, showing that physical exercise is meaningfully associated with reduced mental health difficulties [123] and that people’s choice of activities is motivated by minimizing negative affect and maximizing positive affect [124,125]. In particular, people seem to engage in mood-elevating activities (eg, socializing) when they are feeling down and mood-depressing activities (eg, work and chores) when they are feeling up. At the same time, being in pain affects both mood and daily activities.

Most people intuitively reduce their activities when they experience new-onset pain. Acute pain generally functions as an alarm bell in the brain to signal tissue damage, with the urge to rest considered a protective mechanism to prevent further injury and promote healing [126]. However, outside of the initial acute phase, a lack of activity can hinder long-term recovery and may signal underlying changes in affective and motivational brain circuits that have been causally linked to the transition from subacute to chronic low back pain [126,127]. The synergistic impact of activity engagement on mood, pain, and physical function is why activity engagement is promoted by pain self-management and rehabilitation programs alike [128]. While many people assume that patients with chronic pain move less than those not in pain, objective actigraphy data do not differ between patients with chronic pain and no-pain controls [129,130]. What has been less studied is the types and range of daily activities and how they change between acute and chronic pain stages. While people may recognize changes in their daily activities as a result of pain (eg, no longer walking as much), they may not have insight into the relationship between pain, mood, emotion, and activities (ie, which activities increase or decrease pain in the short vs long term). It remains unclear whether certain patterns of activity engagement at different stages of pain experience are important for long-term outcomes.

For this reason, we added an activities-tracking feature alongside mood and pain tracking. The activities screen on the SOMA app contains 20 activities that are known to have a dynamic relationship with mood based on large-scale, longitudinal data sets [125] or to be common among patients with pain (eg, medical visits). Specific activities selected are detailed in Multimedia Appendix 1. In addition to having the person report either momentary or daily activities, we also have participants reflect on how each activity they completed made them feel that day and how much it affected their pain. Taken together, these measurements provide a full picture of participants’ daily activities and help understand the potential bidirectional relationships between emotion, pain, and behavior. Users are also able to visualize their ratings of how much a given activity affected their pain and mood over time in the dedicated trends screen, which may help people develop insight into how certain activities help or hinder their recovery.

**Measuring Predictions**

Expectations play an undeniably large role in pain perception. In artificial laboratory settings where healthy participants receive painful stimulation, a wealth of findings show that expectations about pain can increase the pain experience (nocebo effect) or conversely decrease it (placebo effect) [131,132]. This has important implications for the experience of pathological pain.

https://mhealth.jmir.org/2024/1/e47177
[133,134], where expectations about pain treatment are reliable predictors of treatment response [135]. These studies suggest that an ongoing cognitive modulation of pain is an important determinant of ongoing pain perception [136]. Expectations for pain relief most likely shift over time, the longer someone has experienced pain the more difficult the pain is to control [137,138]. Subtle longitudinal changes in expectations are believed to occur as pain becomes chronic, but capturing these changes in research studies is challenging. For this reason, we included an assessment of daily predictions about pain, mood, and activities on the SOMA app to capture how predictions vary over time within the same person or between different types of users (eg, patients with acute vs chronic pain).

We decided to assess expectations on the SOMA app by asking users to predict their expected levels of pain intensity, unpleasantness, interference, mood, and activities for the next day using the same scales used to capture their actual rating for that day. This allows the assessment of the bilateral influence of pain and mood expectations on actual experiences of pain, mood, and activities that are entered the following day.

**Discussion**

There is a great need for easy-to-use tools that help those in pain, their medical providers, and the larger health care system identify risk factors and predict the onset of chronic pain. Pain management is a rapidly evolving field that increasingly relies on assessments and treatments that are multimodal and multidisciplinary. Traditional, unidimensional assessments of self-reported pain fail to capture the nuances of pain experience and multimodal pain management. Therefore, there is an urgent need for research tools that have been specifically designed to capture this complexity.

To address this gap, we built the SOMA Science platform. Briefly, the platform uses the smartphone app SOMA to collect longitudinal, multidimensional, ESM-based pain data that capture daily pain intensity, unpleasantness, interference, mood, activities, and predictions. The SOMA app provides a simple and pleasing user interface that can promote pain self-management through visualization of pain trends over time, helping encourage individual insight into factors that exacerbate or alleviate pain. The visualizations provided can also be used to improve communication of multidimensional pain burdens to health care providers. At the same time, the larger SOMA Science platform enables user data to be included in registered single and multicenter studies.

In this paper, we have detailed the clinical and technological considerations taken into account in developing SOMA Science and the scientific rationale behind its measurements. We believe this platform is capable of meeting the requirement for tools to acquire deep and wide-ranging pain data over time, which has been largely absent from existing pain data sets. As such, SOMA Science can be used to answer a broad range of research questions, such as the correlation between initial pain dynamics and the eventual development of chronic pain (ie, predicting the transition from acute to chronic pain), evaluating both short-term and long-term effects of various treatments on pain experiences, or identifying distinct symptom clusters (ie, pain phenotypes). Moreover, the data available are sufficient to calculate more detailed multidimensional and longitudinal clinical trial or observational study end points.

The primary focus of this paper was to introduce SOMA Science as a platform for scientific studies. In the long term, we also plan to build platforms with a more clinical focus that facilitate integration with medical care (SOMA Clinic) and the support of treatments (SOMA Therapeutics). This will involve connecting the SOMA app to electronic medical records and including interfaces to health trackers (eg, daily actigraphy, heart rate, or sleep data from health kits or wearables). The intention is to have a significant positive impact, both in terms of advancing research on pain and improving the lives of people with pain.

**Acknowledgments**

This work was supported by the Brainstorm program at the Robert J & Nancy D Carney Institute for Brain Science (FHP), the COBRE Center for Nervous System Function (grant NIGMS 5P20GM103645; FHP), and conducted with the help of research staff at the Center for Computation and Visualization at Brown University and volunteers from the Rhode Island School of Design. CZG has been awarded an individual research grant that covered her effort on this paper from the National Center for Complementary and Integrative Health (grant F30AT012306). We want to thank everyone who has supported the development of SOMA Science and the SOMA app over the past years, including but not limited to Benjamin Andrew, Rashi Dhar, Louis Rakovich, and Isabel Restrepo.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

The technological and design details regarding the building of the SOMA app, as well as specific measures selected for SOMA app screens, that may be of interest to some readers.

[DOCX File, 22 KB - mhealth_v12i1e47177_app1.docx]

**References**


39. SOMA science: analysis repository for SOMA science. GitHub. URL: https://github/SOMAAnalysis/ [accessed 2023-12-08]

40. Research inquiry. SOMA. URL: https://somatheapp.com/researchinquiry/ [accessed 2023-12-08]


43. SOMA study registration tutorial. YouTube. URL: https://www.youtube.com/watch?v=z1hpomiYXFm [accessed 2023-12-08]


46. M-Health Index and Navigation Database. MindApps. URL: https://mindapps.org/ [accessed 2023-12-08]


52. mindLAMP digital app. Division of Digital Psychiatry at BIDMC. URL: https://www.digitalpsychiatry.org/lamp.html [accessed 2023-12-12]
53. Lamp Platform. URL: https://docs.lamp.digital/ [accessed 2023-12-12]
54. MongoDB. URL: https://www.mongodb.com/ [accessed 2023-12-12]
56. Lazard AJ, Brennen JSB, Belina SP. App designs and interactive features to increase mHealth adoption: user expectation survey and experiment. JMIR Mhealth Uhealth 2021;9(11):e29815 [FREE Full text] [doi: 10.2196/29815] [Medline: 34734829]
58. React Native. URL: https://reactnative.dev [accessed 2023-12-12]
60. SOMA app privacy policy. SOMA App. URL: https://somatheapp.com/pri [accessed 2023-12-12]
77. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365(9453):82-93. [doi: 10.1016/S0140-6736(04)17670-8] [Medline: 15639683]


Abbreviations

APA: American Psychiatric Association
ESM: experience sampling methodology
FDA: Food and Drug Administration
IMMPACT: Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
mHealth: mobile health
NIH: National Institutes of Health
VAS: visual analog scale
Clinical Efficacy of Mobile App–Based, Self-Directed Pulmonary Rehabilitation for Patients With Chronic Obstructive Pulmonary Disease: Systematic Review and Meta-Analysis

Chiwook Chung¹,², MD; Jong Won Lee³, MD, PhD; Sei Won Lee¹*, MD, PhD; Min-Woo Jo⁴*, MD, PhD

¹² these authors contributed equally

Corresponding Author:
Sei Won Lee, MD, PhD

Abstract

Background: Pulmonary rehabilitation is well known to improve clinical symptoms (including dyspnea), quality of life, and exercise capacity in patients with chronic obstructive pulmonary disease (COPD). However, researchers have reported difficulties in practicing center-based pulmonary rehabilitation. Recently, mobile app–based pulmonary rehabilitation has become available in clinical practice. We investigated the clinical outcomes of mobile app–based pulmonary rehabilitation in patients with COPD.

Objective: The objective of our study was to evaluate the clinical efficacy of mobile app–based pulmonary rehabilitation versus conventional center-based pulmonary rehabilitation for patients with COPD, using a systematic review and meta-analysis.

Methods: A systematic search of the literature published between January 2007 and June 2023 was performed, using the PubMed, Embase, Cochrane, and CINAHL databases to identify relevant randomized controlled trials involving patients with COPD. Pulmonary rehabilitation programs needed to provide an exercise program on a smartphone app. Study outcomes, including exercise capacity, symptom scores, quality of life, and hospitalization, were evaluated. The meta-analysis evaluated mean differences in 6-minute walk test distances (6MWDs), COPD Assessment Test (CAT) scores, modified Medical Research Council (mMRC) dyspnea scale scores, St. George Respiratory Questionnaire (SGRQ) scores, and risk ratios for hospitalization resulting from disease exacerbation.

Results: Of the 1173 screened studies, 10 were included in the systematic review and 9 were included in the meta-analysis. Further, 6 studies were multicenter studies. There were a total of 1050 participants, and most were aged ≥65 years. There were discrepancies in the baseline participant characteristics, smartphone apps, interventions, and study outcomes among the included studies. In the meta-analysis, 5 studies assessed 6MWDs (mean difference 9.52, 95% CI −3.05 to 22.08 m), 6 studies assessed CAT scores (mean difference −1.29, 95% CI −2.39 to −0.20), 3 studies assessed mMRC dyspnea scale scores (mean difference −0.08, 95% CI −0.29 to 0.13), 2 studies assessed SGRQ scores (mean difference −3.62, 95% CI −9.62 to 2.38), and 3 studies assessed hospitalization resulting from disease exacerbation (risk ratio 0.65, 95% CI 0.27-1.53). These clinical parameters generally favored mobile app–based pulmonary rehabilitation; however, a statistically significant difference was noted only for the CAT scores (P=0.02).

Conclusions: Despite some discrepancies in the baseline participant characteristics and interventions among studies, mobile app–based pulmonary rehabilitation resulted in favorable exercise capacity, symptom score, quality of life, and hospitalization outcomes when compared with conventional pulmonary rehabilitation. In the meta-analysis, the CAT scores of the mobile app–based pulmonary rehabilitation group were significantly lower than those of the control group (P=0.02). In real-world practice, mobile app–based pulmonary rehabilitation can be a useful treatment option when conventional center-based pulmonary rehabilitation is not feasible.

Trial Registration: PROSPERO CRD42023466965; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=466965

(JMIR Mhealth Uhealth 2024;12:e41753) doi:10.2196/41753

KEYWORDS
pulmonary rehabilitation; COPD; chronic obstructive pulmonary disease; mobile application; mobile app; 6MWD; 6-minute walk test distance; CAT; COPD Assessment Test; mMRC; modified Medical Research Council; SGRQ; St. George Respiratory Questionnaire; exacerbation; rehabilitation; mHealth; mobile health; clinical efficacy; PRISMA; mobile phone
Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by persistent respiratory symptoms and airflow limitation, which are usually caused by exposure to noxious gases or particles [1]. Recently, the prevalence of COPD has increased, making it a leading cause of morbidity and mortality worldwide [2,3]. Approximately 3,500,000 people experience COPD, and it is the third leading cause of disability-adjusted life years (1305 disability-adjusted life years per 100,000 population, 6.21% of total noncommunicable diseases disability-adjusted life years) in South Korea [4,5]. COPD has various extrapulmonary features and might be a systemic disease rather than a disease that only affects the airway [6]. Various clinical information is relevant to the mortality of patients with COPD, including information on physical activity, disability, lung function, long-term oxygen therapy, BMI, quality of life, depressive symptoms, marital status, comorbidity, and hospitalization [7-9]. Additionally, the BODE (BMI, Airflow Obstruction, Dyspnea, and Exercise Capacity) index, which includes BMI, airflow obstruction as assessed by the forced expiratory volume in 1 second (FEV₁), dyspnea as assessed by the modified Medical Research Council (mMRC) dyspnea scale, and exercise capacity as assessed by the 6-minute walk test distance (6MWD), is well known to predict mortality in patients with COPD [10,11].

Pulmonary rehabilitation is a comprehensive intervention for improving the physical and psychological conditions of people with chronic respiratory diseases through exercise training, education, and behavior modification [12]. Pulmonary rehabilitation has been shown to improve dyspnea, quality of life, and exercise capacity in patients with COPD [1,12-14]. Furthermore, patients with chronic respiratory diseases have decreased respiratory muscle mass and strength, which are accompanied by decreased respiratory function. In this population, pulmonary rehabilitation with exercise training is the only way to improve respiratory function [15]. The pulmonary rehabilitation programs used in previous landmark studies were composed of exercise training that was performed 30 to 45 minutes per day, 3 to 5 days per week, for at least 8 to 12 weeks [16,17]. However, researchers reported difficulties in practicing center-based pulmonary rehabilitation, including a lack of facilities; low health insurance coverage; a lack of awareness among physicians; and a lack of motivation, transport barriers, and low levels of social support among patients [4,18,19]. Thus, alternatives to center-based pulmonary rehabilitation are desperately needed [20]. Recently, the demand for telerehabilitation in pulmonary rehabilitation is increasing, owing to advances in telemedicine and challenges with face-to-face rehabilitation during the COVID-19 pandemic [20-22]. Among telerehabilitation modalities, mobile app–based pulmonary rehabilitation has been used in clinical trials; however, the clinical evidence for mobile app–based pulmonary rehabilitation from these studies has been inconclusive due to the heterogeneity in participants, study designs, and formats of apps [23-32]. Furthermore, previous systematic reviews focused on telerehabilitation [20], home telemonitoring [33], or patient support apps [34]. Therefore, we aimed to compare the clinical outcomes of mobile app–based, self-directed pulmonary rehabilitation programs (ie, those without telemonitoring but with exercise programs) in patients with COPD to those of conventional pulmonary rehabilitation because exercise programs are key components of pulmonary rehabilitation that improve chronic respiratory diseases and health-enhancing behaviors [12].

Methods

Data Sources and Literature Search

Literature searches were performed by using the PubMed, Embase, Cochrane, and CINAHL databases. The searches were conducted for literature published since 2007 because the iPhone (Apple Inc) and Android (Google LLC) smartphones were released in June 2007 and September 2008, respectively. The databases were searched for literature published up to June 30, 2023. Only full-text studies written in English were included. The search strategy was based on a PICOTS-SD (population, intervention, comparison, outcomes, time, setting, and study design) list (Multimedia Appendix 1). Briefly, the search algorithm focused on keywords related to “chronic pulmonary disease,” “mobile application,” and various clinical outcomes. If needed, authors were contacted for further information.

Eligibility Criteria and Study Selection

Each study was reviewed by 2 authors (CC and MWJ) independently according to the inclusion and exclusion criteria. The inclusion and exclusion criteria are presented in Table 1. The screening of titles and abstracts and the subsequent full-text review were performed by 2 authors (CC and MWJ) independently. Disagreements during the selection process were resolved through a discussion between 3 authors (CC, MWJ, and SWL).
<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article type</td>
<td>Full-text articles</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Study design</td>
<td>Randomized controlled trials</td>
</tr>
<tr>
<td>Participants' age</td>
<td>Adults</td>
</tr>
<tr>
<td>Disease</td>
<td>COPD^a</td>
</tr>
<tr>
<td>Smartphone app</td>
<td>Conventional or newly developed smartphone apps</td>
</tr>
<tr>
<td>Intervention</td>
<td>Pulmonary rehabilitation, including exercise programs, provided by a smartphone app</td>
</tr>
<tr>
<td>Control</td>
<td>Conventional pulmonary rehabilitation, including exercise programs (center-based rehabilitation or education)</td>
</tr>
<tr>
<td>Study outcome</td>
<td>At least 1 of the following outcomes: 6-minute walk test distance, COPD Assessment Test score, modified Medical Research Council dyspnea scale score, St. George Respiratory Questionnaire score, and hospitalization resulting from disease exacerbation</td>
</tr>
</tbody>
</table>

^aCOPD: chronic obstructive pulmonary disease.

^bN/A: not applicable.

**Data Collection and Risk of Bias Assessment**

Two authors (CC and MWJ) independently collected data regarding (1) general information about the study (authors, year, country, and study setting), (2) descriptions of study arms (number, sex, and age of participants), (3) characteristics of interventions, (4) inclusion and exclusion criteria, and (5) results for outcomes; they also double-checked these data. Two authors (CC and MWJ) independently assessed the risk of bias in the included studies. Discrepancies were resolved in discussions with the third author (SWL).

**Study Outcomes**

In the meta-analysis, study outcomes, including exercise capacity, symptom scores, quality of life, and hospitalization, were assessed. Exercise capacity was measured by using 6MWDs. The symptom scores were measured by using the COPD Assessment Test (CAT) and the mMRC dyspnea scale. Quality of life was measured by using the St. George Respiratory Questionnaire (SGRQ). Hospitalization was defined as hospitalizations resulting from disease exacerbation. The primary time points for the analysis were baseline and the end of the intervention.

**Statistical Analysis**

The continuous variables included the 6MWD, CAT score, and SGRQ score. The mMRC dyspnea scale score was a categorical variable, and it was calculated as a continuous value. Hospitalization resulting from disease exacerbation was a dichotomous variable. The variables at the time of follow-up were compared between groups. The mean differences and risk ratios between the intervention group and the control group were calculated, along with 95% CIs. The chi-square test and the $I^2$ statistic were used to assess statistical heterogeneity. If $I^2$ was <50%, the fixed effect model was used. Publication bias was visually assessed by using a funnel plot analysis because the limited number of studies with results for each outcome prevented us from performing the Egger test. The meta-analysis was performed by using Review Manager (RevMan) version 5.4 (The Cochrane Collaboration).

**Ethical Considerations**

This study complied with the Declaration of Helsinki, and all methods were performed in accordance with the relevant guidelines.

**Results**

**Study Selection**

An initial literature search identified a total of 1851 articles from the PubMed, Embase, Cochrane, and CINAHL databases; thereafter, 1173 articles remained after duplicates were removed. After evaluating titles and abstracts, 299 articles remained eligible for a full-text review. The full-text review was performed according to the criteria mentioned in the Eligibility Criteria and Study Selection section, and 10 articles were finally included in the systematic review [23-32]. Notably, 1 study was excluded from the meta-analysis because exercise capacity was evaluated by using the incremental shuttle walk test (ISWT) instead of 6MWDs [32]. Therefore, 9 studies were included in the meta-analysis [23-31] (Figure 1).
Characteristics of Included Studies

Characteristics of studies are described in Table 2. Studies were published after 2014, with almost half of them (4/10, 40%) published in 2020 [24,26,28,29]. Further, 6 studies were multicenter studies [24,25,27,29-31], and 3 studies enrolled fewer than 50 participants; the largest number of participants was 343 [25,28,29,32]. There were 1050 total participants, who were generally aged ≥65 years. More male participants were enrolled than female participants, and Wang et al [32] enrolled only male participants. In the study by North et al [28], participants were recruited after hospital admission with an acute exacerbation. In the studies by Vorrink et al [31] and Wang et al [32], participants were recruited after pulmonary rehabilitation. Kwon et al [27] recruited 2 groups of participants in the intervention arm, comprising the fixed regimen group and the fixed-interactive regimen group, according to exercise programs. Various formats of mobile apps were used for the studies; 2 studies in the United Kingdom used myCOPD, a digital health care app approved by the National Health Service [24,28], and 1 study in China used WeChat (Tencent Holdings Ltd), a popular mobile messenger app in China [26]. The follow-up duration ranged between 3 weeks and 12 months [23,31].
Table 3. Characteristics of included studies.

<table>
<thead>
<tr>
<th>Study (author, year)</th>
<th>Setting</th>
<th>Country</th>
<th>Sample size, n (%)</th>
<th>Age (y), mean (SD)</th>
<th>Mobile app</th>
<th>Follow-up duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barata et al [23], 2022</td>
<td>Single center</td>
<td>Romania</td>
<td>Control: Male: 16 (27.6); female: 6 (10.2); Intervention: Male: 54 (75); female: 18 (25)</td>
<td>64.3 (4.3)</td>
<td>Pneumocontrol app (newly developed)</td>
<td>21 d</td>
</tr>
<tr>
<td>Crooks et al [24], 2020</td>
<td>Multicenter</td>
<td>United Kingdom</td>
<td>Control: Male: 20 (64.5); female: 11 (35.5); Intervention: Male: 11 (37.9); female: 18 (62.1)</td>
<td>65.9 (7.3)</td>
<td>myCOPD</td>
<td>90 d</td>
</tr>
<tr>
<td>Demeyer et al [25], 2017</td>
<td>Multicenter</td>
<td>Belgium</td>
<td>Control: Male: 42 (72.4); female: 16 (27.6); Intervention: Male: 111 (64.9); female: 60 (35.1)</td>
<td>66 (8)</td>
<td>Fitbug app and a project-tailored coaching app</td>
<td>12 wk</td>
</tr>
<tr>
<td>Jiang et al [26], 2020</td>
<td>Single center</td>
<td>China</td>
<td>Control: Male: 43 (81.1); female: 10 (18.9); Intervention: Male: 44 (83); female: 9 (17)</td>
<td>70.9 (6.4)</td>
<td>WeChat official account based on social media</td>
<td>6 mo</td>
</tr>
<tr>
<td>Kwon et al [27], 2018</td>
<td>Multicenter</td>
<td>Republic of Korea</td>
<td>Control: Male: 11 (52.4); female: 10 (47.6); Intervention: Male: 23 (85.2); female: 4 (14.8);</td>
<td>65.1 (6.3)</td>
<td>efil breath (newly developed)</td>
<td>12 wk</td>
</tr>
<tr>
<td>North et al [28], 2020</td>
<td>Single center</td>
<td>United Kingdom</td>
<td>Control: Male: 13 (65); female: 7 (35); Intervention: Male: 11 (52.4); female: 10 (47.6)</td>
<td>68.1 (7.4)</td>
<td>myCOPD</td>
<td>90 d</td>
</tr>
<tr>
<td>Park et al [29], 2020</td>
<td>Multicenter</td>
<td>Republic of Korea</td>
<td>Control: Male: 42 (50); female: 42 (50); Intervention: Male: 19 (86.4); female: 3 (13.6)</td>
<td>70.5 (9.4)</td>
<td>COPD self-management program (newly developed)</td>
<td>6 mo</td>
</tr>
<tr>
<td>Spielmanns et al [30], 2023</td>
<td>Multicenter</td>
<td>Switzerland</td>
<td>Control: Male: 17 (50.5); female: 16 (48.5); Intervention: Male: 17 (50); female: 17 (50)</td>
<td>66.1 (6.8)</td>
<td>Kaia COPD app (newly developed)</td>
<td>6 mo</td>
</tr>
<tr>
<td>Vorrink et al [31], 2016</td>
<td>Multicenter</td>
<td>Netherlands</td>
<td>Control: Male: 36 (49.3); female: 37 (50.7); Intervention: Male: 42 (50); female: 42 (50)</td>
<td>62 (9)</td>
<td>Newly developed</td>
<td>12 mo</td>
</tr>
<tr>
<td>Wang et al [32], 2014</td>
<td>Single center</td>
<td>Taiwan</td>
<td>Control: Male: 14 (100); Intervention: Male: 12 (100)</td>
<td>71.4 (1.9)</td>
<td>Newly developed</td>
<td>6 mo</td>
</tr>
</tbody>
</table>

aThe fixed regimen group.
bThe fixed-interactive regimen group.
cParticipants were recruited after hospital admission with an acute exacerbation.
dCOPD: chronic obstructive pulmonary disease.
eParticipants were recruited after pulmonary rehabilitation.

The interventions in the studies are described in Table 3. Disease education and monitoring were provided in 5 studies [24,26,28-30,35], and the other 5 studies provided only exercise programs [23,25,27,31,32]. The level of exercise could be adjusted according to the participants’ exercise capacity in 5 studies [23,25,27,31,32]. In particular, Kwon et al [27] provided 2 kinds of exercise regimens, and walking distances were adjustable in both regimens. In cases of COPD exacerbation or...
poor compliance to pulmonary rehabilitation, participants could contact health care professionals in 7 studies [24-26,28-30,32,35]. Jiang et al [26] gave incentives to participants, that is, participants could obtain gifts at a mall by using acquired points.

**Table . Interventions of included studies.**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Exercise adjustment</th>
<th>Exercise monitoring</th>
<th>Disease education</th>
<th>Disease monitoring</th>
<th>Social support</th>
<th>Contact with health care professionals</th>
<th>Incentive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barata et al [23]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Crooks et al [24]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Demeyer et al [25]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Jiang et al [26]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Kwon et al [27]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>North et al [28]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Park et al [29]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Spielmanns et al [30]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Vorrink et al [31]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Wang et al [32]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Most studies (7/10, 70%) included adult participants with physician-diagnosed COPD; diagnoses were made according to the GOLD (Global Initiative for Chronic Obstructive Lung Disease) criteria [1]. Some studies did not include participants with severe COPD as defined by the GOLD criteria [29,31], and others did not set limitations for disease severity. Generally, participants with recent acute exacerbations, participants undergoing long-term home oxygen therapy, or participants with other medical conditions that did not allow for physical exercise were excluded. In the study by North et al [28], participants were included after hospitalization with an acute exacerbation (Table S1 in Multimedia Appendix 2).

Participants were evaluated on various dimensions of outcomes, including exercise capacity, disease severity, quality of life questionnaires, and acute exacerbation. Wang et al [32] reported favorable exercise capacity and serum inflammatory biomarker outcomes; however, this study was excluded from the meta-analysis because exercise capacity was reported based on the ISWT and limb muscle strength. Crooks et al [24] and North et al [28] reported that inhaler technique improved in the intervention group, which was beneficial to disease control. Demeyer et al [25] reported that lung function did not improve during pulmonary rehabilitation in the intervention and control groups, and musculoskeletal events occurred more often in the intervention group. Barata et al [23] reported that the maximal inspiratory and expiratory pressures improved in the intervention group (Table S2 in Multimedia Appendix 2).

**Risk of Bias in Studies**

The overall risk of bias in studies was considered low. However, the risk of performance bias was inevitably considered high in all studies because participant blinding was impossible, owing to the nature of the intervention (Figure 2). Funnel plots of comparisons showed fairly symmetrical distributions, which might mean less publication bias (Figure S1 in Multimedia Appendix 3).
Figure 2. Risk of bias in the included studies [23-31]. A: Risk of bias graph. B: Risk of bias summary; “Kwon H 2018 (1)” denotes the fixed regimen group, and “Kwon H 2018 (2)” denotes the fixed-interactive regimen group.

Meta-Analysis of Clinical Outcomes

Figure 3 shows the meta-analysis of study outcomes. In terms of statistical heterogeneity, the chi-square test and $I^2$ statistic for each meta-analysis showed no important heterogeneity. Exercise capacity was reported in various forms, including 6MWDs, ISWT results, the number of steps per day, and metabolic equivalents, in 8 studies [23-25,27,29-32]. Wang et al [32] reported on the ISWT only, and Crooks et al [24] and Spielmanns et al [30] reported the number of steps per day only. Thus, the 6MWD, which was used in 5 studies, was included in the meta-analysis [23,25,27,29,31]; there was no statistically significant difference between groups (mean difference 9.52, 95% CI −3.05 to 22.08 m; $P=.14$).
CAT scores were reported in 7 studies [23-28,30]; however, Demeyer et al [25] reported the CAT scores as medians and IQRs. Thus, the CAT scores from 6 studies were analyzed [23,24,26-28,30]. The CAT scores of the intervention group were significantly lower than those of the control group (mean difference $-1.29$, 95% CI $-2.39$ to $-0.20$; $P=.02$). Dyspnea was measured by using the mMRC dyspnea scale in 3 studies [26-28], and the scores did not significantly differ between groups (mean difference $-0.08$, 95% CI $-0.29$ to $0.13$; $P=.45$).

The quality of life was assessed in 6 studies, using various questionnaires [24-26,28,29,31], and SGRQ scores were reported in 2 trials [26,28]; there was no statistical difference in these scores between groups (mean difference $-3.62$, 95% CI $-9.62$ to $2.38$; $P=.24$).

The exacerbation of COPD was reported as outpatient clinic visits, emergency room visits, or hospitalizations in 4 studies [24,25,28,29]. Among them, hospitalizations were reported in 3 studies [24,28,29]. The frequency of hospitalization was not statistically different between groups (risk ratio 0.65, 95% CI 0.27 to 1.53; $P=.32$).

We also performed a subgroup analysis for the 6MWDs and CAT scores based on the baseline study results (6MWDs: $\geq 400$ m vs $<400$ m; CAT scores: $\geq 20$ vs $<20$) [23-31]. The subgroup
analysis did not show statistically significant differences (all \( P \) values were >.05). Furthermore, we performed a subgroup analysis for the CAT scores based on the rehabilitation programs (exercise program only vs exercise and self-management programs) [23,24,26-28,30]. Among studies offering both exercise and self-management programs, the CAT scores of the intervention group were significantly lower than those of the control group (mean difference \( -2.16, 95\% \text{ CI} -3.93 \text{ to } -0.39; P=.02 \); Figure S2 in Multimedia Appendix 3 [23-31]).

Discussion

Principal Results and Implications

We reviewed and described the clinical outcomes of mobile app–based pulmonary rehabilitation in patients with COPD. Participants and interventions were heterogeneous in their characteristics; however, participants who underwent mobile app–based pulmonary rehabilitation showed favorable exercise capacity, symptom score, quality of life, and hospitalization outcomes when compared to participants who underwent conventional pulmonary rehabilitation. In the meta-analysis, the 6MWDs, mMRC dyspnea scale scores, SGRQ scores, and exacerbations in the mobile app–based pulmonary rehabilitation group were not inferior to those in the control group, and the CAT scores were superior to those in the control group. Considering the difficulties in practicing conventional center-based pulmonary rehabilitation, mobile app–based pulmonary rehabilitation may be a useful treatment option when conventional pulmonary rehabilitation is not feasible.

Mobile App–Based Pulmonary Rehabilitation

Pulmonary rehabilitation has been traditionally delivered in outpatient, inpatient, and community settings, comprising \( \geq 2 \) sessions per week for at least 4 weeks [14]. In 2015, the American Thoracic Society/European Respiratory Society policy statement requested researchers to adopt alternative formats for pulmonary rehabilitation, demonstrate clinical outcomes that are at least comparable to those of traditional pulmonary rehabilitation programs, and evaluate cost-effectiveness and safety [36]. Since then, clinical trials have reported data on the clinical outcomes and safety of pulmonary rehabilitation program models, including home-based rehabilitation, telerehabilitation, web-based rehabilitation, community rehabilitation, primary care rehabilitation, rehabilitation requiring minimal resources, and combined heart failure/pulmonary rehabilitation models [22]. Mobile app–based pulmonary rehabilitation can be regarded as a type of telerehabilitation intervention [20] that provides health care at a distance through telecommunications or web-based technologies [37]. It may improve the accessibility of pulmonary rehabilitation for patients with chronic respiratory diseases by providing health care access and services for patients who are geographically or socially isolated, are engaged with full-time work, or are hard to transport due to the disease or comorbidities [20].

Further Development of Apps

Various types of apps were used in the studies. Some authors used newly developed apps, and others used myCOPD or the social messenger app WeChat [24,26,28]. Some apps, such as myCOPD, provided self-management programs for COPD, including education and symptom management programs [24,28]; however, other apps provided only exercise programs [25,27]. Although this study focused on clinical improvements in participants who underwent pulmonary rehabilitation, it should also be considered that overall self-management programs, such as disease education and symptom management programs, have affected clinical outcomes. However, pulmonary rehabilitation is defined as a comprehensive intervention that includes exercise training, education, and behavior change [12]. Recently, Holland et al [22] suggested that desirable components of pulmonary rehabilitation should include education, self-management training, smoking cessation, and an action plan for exacerbation, as well as a home exercise program. Therefore, apps that provide both exercise programs and self-management programs should be included in mobile app–based pulmonary rehabilitation.

Considering the challenges in center-based pulmonary rehabilitation and the shortage of health care resources, home-based pulmonary rehabilitation has been studied as an alternative to center-based pulmonary rehabilitation [38-43]. However, compliance to pulmonary rehabilitation is an important issue in home-based pulmonary rehabilitation, and a lack of motivation is an important reason for poor compliance [44]. In a study of home-based pulmonary rehabilitation without supervision, patients with good compliance showed significant improvements in CAT scores, BODE index scores, and FEV\(_1\) values when compared to patients with poor compliance [45]. Similarly, Crooks et al [24] described that there was an estimated \(-0.22 (95\% \text{ CI} -0.74 \text{ to } -0.31) \) decrease in the CAT score for every 7-day increase in app use (adjusted for baseline CAT score, COPD severity, and study site). However, North et al [28] reported that as time passed, the number of app users decreased in mobile app–based pulmonary rehabilitation. Therefore, patients are required to steadily run the app and perform pulmonary rehabilitation to achieve clinical improvement. Various methods were used in studies to enhance compliance, such as sending text messages with activity proposals to participants, contacting participants via telephone, providing incentives, and having participants communicate with other participants [25,26,29,31,32]. Additionally, activity level (step counts) was monitored by using a pedometer, and feedback was provided to participants [25,29,31,32]. In real-world practice, health care interventions and action plans should be considered in cases of poor compliance because poor compliance might reflect deconditioning or acute exacerbation among patients [22,26,44].

Further Development of Rehabilitation Programs

In clinical practice, exercise levels in pulmonary rehabilitation should be individualized according to patients’ exercise capacity [12,13]. Therefore, in mobile app–based pulmonary rehabilitation, maintaining appropriate exercise levels is a matter of concern. Some apps provided adjustable exercise regimens according to the changes in participants’ exercise capacity [25,27,31,32]. Kwon et al [27] designed exercise regimens in which the exercise levels were adjusted according to the maximum walking speed in the 6-minute walk test and the degree of breathing difficulty after exercise. Vorriink et al [31]...
designed physical activity goals that were set according to average steps per day. To maintain appropriate exercise levels, apps should provide adjustable and individualized exercise programs based on patients’ exercise capacity and activity level data that are collected via wearable devices or smartphone-mounted sensors.

Considering the study designs included in this review, it is important to develop strategies for improving compliance to rehabilitation and design individualized exercise programs to achieve significant improvements in clinical outcomes in future studies. Moreover, most studies (6/10, 60%) had rather small sample sizes (<100 individuals) for demonstrating the efficacy of pulmonary rehabilitation programs [24,27-30,32]. In addition, most studies (8/10, 80%) did not provide data regarding app usage, which could have been used in the subgroup analysis related to compliance [23,25-27,29-32]. Therefore, further studies with larger sample sizes and data on app usage are needed.

Nutrition support is also an important part of pulmonary rehabilitation [13,22]. In this review, some of the included apps provided disease education; however, a nutrition support program was not provided [24,26,28-30,35]. Nutrition support may help patients with COPD to maintain an adequate BMI and increase their muscle mass [13,22]. Exercise training that is accompanied by nutrition support might improve respiratory sarcopenia and enhance clinical benefits [15]; thus, further studies are needed in this area.

Clinical Outcomes and Prognosis

Exercise capacity and physical activity data can be used to predict the prognosis of patients with COPD. Exercise capacity inversely correlates with mortality in patients with COPD [46]. Physical activity also inversely correlates with exacerbation and mortality in patients with COPD [47]. Some of the reviewed studies reported physical activity as daily step counts, and these data were too widely distributed to be synthesized in the meta-analysis [24,30]. Moreover, the 6MWDs were not significantly different in the meta-analysis (P=14), and Wang et al [32] reported improvements in the ISWT and limb muscle mass in the intervention group. Thus, further studies are required to ascertain whether mobile app–based pulmonary rehabilitation can improve exercise capacity and physical activity in patients with COPD.

In some studies, we noticed that mobile app–based pulmonary rehabilitation improved quality of life, including the SGRQ, Clinical COPD Questionnaire, and Chronic Respiratory Disease Questionnaire scores [24,26,28,29,31]. Among these, the CAT scores significantly improved in the intervention group, as per the meta-analysis (P=02) [23,24,26-28,30]. The CAT scores correlated with the severity of airflow limitation and disease exacerbation in patients with COPD [48,49]. Taken together, mobile app–based pulmonary rehabilitation programs might improve clinical outcomes, such as acute exacerbation and mortality. Unfortunately, in the meta-analysis, there was no statistically significant difference in acute exacerbations between groups (P=32) because the study periods (range 3-6 mo) might have been too short to observe acute exacerbations [24,28,29]. Therefore, further studies with long-term follow-ups are required to evaluate the effect of mobile app–based pulmonary rehabilitation on acute exacerbations and mortality.

Limitations

First, discrepancies in the baseline status of participants were one of the main obstacles in synthesizing clinical outcomes. In the study by North et al [28], participants were evaluated after hospitalization with an acute exacerbation. In the studies by Vorring et al [31] and Wang et al [32], physical activity in participants with COPD was evaluated after pulmonary rehabilitation. Despite this heterogeneity, participants who underwent mobile app–based pulmonary rehabilitation showed consistently favorable results for clinical parameters. Second, discrepancies in the clinical parameters were also an obstacle in synthesizing clinical outcomes. Among the various parameters for exercise capacity, a meta-analysis could be performed on the 6MWD, as it was used in half of the reviewed studies (5/10, 50%) [23,25,27,29,31], and the 6MWD is a well-established surrogate marker in patients with COPD [1,50]. Questionnaires about quality of life, including the SGRQ, EQ-5D-5L, Clinical COPD Questionnaire, and Chronic Respiratory Disease Questionnaire, also showed generally favorable results in patients who underwent mobile app–based pulmonary rehabilitation [24-26,28,29,31]. Although clinical outcomes did not reflect statistically significant improvement in participants who underwent mobile app–based pulmonary rehabilitation and decisive evidence was hard to derive, this study showed that clinical outcomes generally favored mobile app–based pulmonary rehabilitation. Considering the difficulties with center-based pulmonary rehabilitation in real-world practice, mobile app–based pulmonary rehabilitation could be a reasonable alternative to conventional pulmonary rehabilitation.

Conclusion

In conclusion, this review shows that many mobile apps have been applied to pulmonary rehabilitation for patients with COPD. There were discrepancies in the baseline participant characteristics and interventions among studies. Nevertheless, in some studies, patients who participated in mobile app–based pulmonary rehabilitation showed favorable exercise capacity, symptom score, quality of life, and hospitalization outcomes when compared with those who underwent conventional pulmonary rehabilitation. In the meta-analysis, the 6MWDs, mMRC dyspnea scale scores, SGRQ scores, and exacerbations in the mobile app–based pulmonary rehabilitation group were not inferior to those in the control group, and the CAT scores were superior to those in the control group. Therefore, in real-world practice, mobile app–based pulmonary rehabilitation can be a useful treatment option when conventional center-based pulmonary rehabilitation is not feasible.
Acknowledgments
This research was supported by a grant of the Korea Health Promotion R&D Project, funded by the Ministry of Health & Welfare, Republic of Korea (grant HS21C0096), and a grant from the Korea Health Technology R&D Project, provided through the Korea Health Industry Development Institute (KHIDI) and funded by the Ministry of Health & Welfare, Republic of Korea (grant HI20C1058). We thank Asan Medical Library at University of Ulsan College of Medicine for helping with the literature search.

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
CC, JWL, SWL, and MWJ were responsible for the concept and design of the study. CC, SWL, and MWJ were responsible for data acquisition, data curation, and formal analysis. CC, JWL, and SWL drafted the original manuscript. All authors have revised and approved the final manuscript. All authors take responsibility for the accuracy of the content of the final manuscript. JWL and SWL obtained funding. Authors SWL and MWJ are co-corresponding authors. MWJ can be contacted at mdjominwoo@gmail.com. Generative artificial intelligence was not used in any portion of manuscript writing.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PICOTS-SD (population, intervention, comparison, outcomes, time, setting, and study design) search strategy for mobile apps for patients with chronic pulmonary disease.

Multimedia Appendix 2
Supplementary tables for the inclusion criteria, exclusion criteria, and clinical outcomes of the included studies.

Multimedia Appendix 3
Supplementary figures for the outcomes of included studies.

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

References


https://mhealth.jmir.org/2024/1/e41753 JMIR Mhealth Uhealth 2024 | vol. 12 | e41753 | p.68 (page number not for citation purposes)
Abbreviations

- 6MWD: 6-minute walk test distance
- BODE: BMI, Airflow Obstruction, Dyspnea, and Exercise Capacity
- CAT: COPD Assessment Test
- COPD: chronic obstructive pulmonary disease
- FEV₁: forced expiratory volume in 1 second
- GOLD: Global Initiative for Chronic Obstructive Lung Disease
- ISWT: incremental shuttle walk test
- mMRC: modified Medical Research Council
- PICOTS-SD: population, intervention, comparison, outcomes, time, setting, and study design
- SGRQ: St. George Respiratory Questionnaire
Application of eHealth Tools in Anticoagulation Management After Cardiac Valve Replacement: Scoping Review Coupled With Bibliometric Analysis

Ying Wu1,2*, MSc; Xiaohui Wang2*, MSc; Mengyao Zhou2, BSc; Zhuoer Huang2, MSc; Lijuan Liu3, MSc; Li Cong2, PhD

1Center for Moral Culture, Hunan Normal University, Changsha, China
2School of Medicine, Hunan Normal University, Changsha, China
3Teaching and Research Section of Clinical Nursing, Xiangya Hospital of Central South University, Changsha, China
* these authors contributed equally

Corresponding Author:
Li Cong, PhD
School of Medicine
Hunan Normal University
371 Tongzipo Road
Changsha, 410013
China
Phone: 86 0731 889124
Email: congli@hunnu.edu.cn

Abstract

Background: Anticoagulation management can effectively prevent complications in patients undergoing cardiac valve replacement (CVR). The emergence of eHealth tools provides new prospects for the management of long-term anticoagulants. However, there is no comprehensive summary of the application of eHealth tools in anticoagulation management after CVR.

Objective: Our objective is to clarify the current state, trends, benefits, and challenges of using eHealth tools in the anticoagulation management of patients after CVR and provide future directions and recommendations for development in this field.

Methods: This scoping review follows the 5-step framework developed by Arksey and O’Malley. We searched 5 databases such as PubMed, MEDLINE, Web of Science, CINAHL, and Embase using keywords such as “eHealth,” “anticoagulation,” and “valve replacement.” We included papers on the practical application of eHealth tools and excluded papers describing the underlying mechanisms for developing eHealth tools. The search time ranged from the database inception to March 1, 2023. The study findings were reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews). Additionally, VOSviewer (version 1.6.18) was used to construct visualization maps of countries, institutions, authors, and keywords to investigate the internal relations of included literature and to explore research hotspots and frontiers.

Results: This study included 25 studies that fulfilled the criteria. There were 27,050 participants in total, with the sample size of the included studies ranging from 49 to 13,219. The eHealth tools mainly include computer-based support systems, electronic health records, telemedicine platforms, and mobile apps. Compared to traditional anticoagulation management, eHealth tools can improve time in therapeutic range and life satisfaction. However, there is no significant impact observed in terms of economic benefits and anticoagulation-related complications. Bibliometric analysis suggests the potential for increased collaboration and opportunities among countries and academic institutions. Italy had the widest cooperative relationships. Machine learning and artificial intelligence are the popular research directions in anticoagulation management.

Conclusions: eHealth tools exhibit promise for clinical applications in anticoagulation management after CVR, with the potential to enhance postoperative rehabilitation. Further high-quality research is needed to explore the economic benefits of eHealth tools in long-term anticoagulant therapy and the potential to reduce the occurrence of adverse events.

(JMIR Mhealth Uhealth 2024;12:e48716) doi:10.2196/48716
KEYWORDS
eHealth tool; cardiac valve replacement; anticoagulation management; scoping review; bibliometrics analysis; rehabilitation

Introduction

Valvular heart disease involves damage to the cardiac valves caused by various factors such as valve degeneration and rheumatic heart disease [1]. Currently, an estimated 209 million people suffer from valvular heart disease worldwide. With the aging population, it is expected that the prevalence of valvular heart disease among those aged 65 years and older will increase by 50% globally by 2030 [2,3]. Furthermore, valvular heart disease caused approximately 300,000 deaths in 2019, accounting for nearly 2% of cardiovascular disease deaths worldwide. Every year, 8.7 million life years and 10.7 million disability-adjusted life years are lost to valvular heart disease [2]. The fundamental treatment for valvular heart disease is the repair or replacement of damaged heart valves through surgery or interventional therapy. Patients need long-term oral anticoagulants after cardiac valve replacement (CVR) surgery; however, improper use of anticoagulants can lead to bleeding, embolism, eventual valve failure, repeat surgeries, and even death [4,5]. Efficient anticoagulation management (AM) plays a critical role in promoting safe medication therapy after CVR, reducing adverse events, and improving the long-term prognosis of patients.

The traditional AM model has many obstacles and challenges. On the one hand, patients need to travel between their home and hospital regularly for follow-up and laboratory examinations and long-term anticoagulant medication, which costs them a lot of time, energy, and economy after CVR and also reduces patients’ confidence in rehabilitation and treatment compliance [6]. On the other hand, it is difficult for doctors to monitor and manage patients and adjust treatment plans promptly, which may affect the quality and effect of anticoagulant therapy, and increase the risk of adverse events [7]. Therefore, the exploration of personalized, intelligent, and efficient AM models is crucial in promoting cardiac recovery among patients undergoing CVR. Several guidelines highlight the potential of eHealth tools to facilitate AM and improve medical outcomes [8] and recommend the use of eHealth tools to assist physicians with anticoagulation decision-making support [9].

eHealth tools generally refer to digital devices based on information and communication technology, which can be used for disease prevention, diagnosis, treatment, monitoring, and management [10]. eHealth tools can provide patients with convenient, efficient, and accurate medical services through digital communication and remote disease management. Some studies have found that eHealth tools may become a potentially cost-effective and effective alternative to traditional “face-to-face” anticoagulation therapy. A meta-analysis of 12 randomized controlled trials (RCTs) showed that telemedicine combined with portable coagulometers significantly improved the time in the therapeutic range (TTR) and reduced the incidence of thrombotic events in patients with cardiovascular disease [11]. Inpatients undergoing anticoagulation therapy receiving the recommended dosage of anticoagulation software can effectively reduce readmissions and 30-day mortality as well as hospitalization costs [12]. A comprehensive understanding of the application status and effectiveness of eHealth tools is essential to establish a scientific, traceable, and integrated AM model.

Although several narrative reviews and meta-analyses have summarized the usage of mobile health technology in patients with cardiovascular diseases [13-15], uncertainties remain regarding the effectiveness and current status of using eHealth tools in AM of patients after CVR. A comprehensive summary of the development, application, and interrelationships among different research papers or groups is lacking. Scoping reviews entail extensive searches and a rigorous selection of research literature on a specific topic, which can include a comprehensive overview of the current research status and trends of the topic [16]. Scientometric analysis can quantitatively evaluate cooperative relationships by statistically analyzing publications and graphically presenting the social and intellectual connections of relevant literature [17]. Through the joint analysis of scoping review and scientometric analysis, the research status and trends of a certain field can be comprehensively evaluated from different perspectives, and the blind spots and unsolved problems can be determined to provide guidance and enlightenment for further research.

By conducting a scoping review and scientometric analysis, this study aimed to (1) summarize the application status of eHealth tools in AM of patients after CVR, (2) identify the hotspots and provide guidance for future research and practice, and (3) to provide a reference for promoting the wider application and sustainable development of eHealth tools in AM of patients after CVR.

Methods

Design

We used Arksey and O’Malley’s [18] 5-step framework for the scoping review. This review also followed the recommended items in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1). Five commonly used databases were searched, and the papers were screened based on the title, abstract, and full text. We exported the complete records of filtered papers as plain text files and imported them into VOSviewer software (version 1.6.18; Centre for Science and Technology Studies) to build visualization maps. We chose coauthorship analysis to construct network visualization maps of countries, institutions, and authors to understand the status of research collaboration in the field of eHealth tools. By cluster analysis and keyword overlay visualization, keywords are divided into different clusters and stacked over time, which can identify different themes and current research hotspots.

Scoping Review

In contrast to systematic and narrative reviews, scoping reviews focus on an initial appraisal of the current extent, scope, and
nature of the research literature and take the dissemination process further by summarizing the relevant existing research activities. It is also an appropriate way to map the key concepts and identify knowledge gaps [19]. We aimed to provide an overview of the use of eHealth tools to assist anticoagulation therapy in patients after CVR and highlight the current status, trends, and challenges in this field. Therefore, the scoping review was appropriate for this study.

Identifying the Research Question for Scoping Review
The first question that guided our scoping review was what are the range and effectiveness of eHealth tools services in AM of patients after CVR? The second one was what are the benefits and barriers of applying eHealth tools in AM?

Identifying Databases and Studies
PubMed, MEDLINE, Web of Science, CINAHL, and Embase were searched from inception to March 1, 2023. Searches were not limited to a specific geographic region, and any literature published in non-English languages was excluded. Searches included combinations of free text words and index terms using Boolean operators. Moreover, a manual retrospective search of the references was conducted as a supplement. Detailed search strategies for each database are described in Multimedia Appendix 2.

Study Selection
The Population, Concept, and Context framework is recommended by the Joanna Briggs Institute to identify the main concepts in primary review questions, guide the search strategy, and ensure application of the inclusion and exclusion criteria [20]. Therefore, we used the Population, Concept, and Context framework to regulate the scoping review process. Population was defined as patients who had undergone CVR surgery and received anticoagulant therapy postoperatively. Concept referred to the practical application and effectiveness of eHealth tools in the AM of patients after CVR. Context had no special restrictions, as eHealth tools can be applied in the patient’s home, hospital, anticoagulant therapy clinic, primary care center, and so forth. The inclusion criteria were as follows: (1) publication types including cross-sectional studies, longitudinal studies, cohort studies, case-control studies, pilot studies, and RCTs; (2) published in full text; and (3) published from the inception of each database to March 1, 2023. The exclusion criteria were as follows: (1) non-English publications; (2) publication types including empirical research, reviews, editorials, reports, case reports, letters, and conference proceedings or papers or abstracts; (3) qualitative studies reporting user experiences about eHealth tools in AM of patients after CVR; and (4) studies describing only the potential mechanisms or development process of eHealth tools, rather than their practical applications. The literature records retrieved were imported into Note Express software to screen for duplicate papers. Two researchers (YW and XW) independently conducted the initial screening of the titles and abstracts based on the inclusion and exclusion criteria. Full texts were then examined for secondary screening of potentially eligible papers. Any disagreement during the screening process was resolved through discussion with the third researcher (LC), and the final selection of papers was determined based on the established criteria.

Data Extraction and Analysis
Two researchers (YW and XW) independently extracted data from the included studies using standardized tables and cross-checking their findings. Any discrepancies were resolved through discussion with the third researcher (LC). The information extracted included authors, year, country, study design, study objective, participants’ characteristics, study location, content elements, outcome measures, and study conclusions. We summarized the literature on the services and effects of eHealth tools in the AM of patients after CVR and presented the main concepts and findings of the literature using data charts and tables.

Bibliometric Analysis
VOSviewer is a widely used tool by researchers for bibliometric analysis, providing effective visualization and revealing connections between research materials [21]. The process of scientometric analysis in this study included 2 parts. First, coauthorship analysis was performed to explore the relationship among researchers, research institutions, and countries, which contributed to understanding the trends of scientific cooperation. We chose “coauthorship” as the analysis type, selected “countries,” “organizations,” and “authors” as the units of analysis and “full counting” as the counting method. To achieve clearer effects, the minimum number for each project was set as 1. Second, keywords network visualization and overlay visualization were analyzed to explore the hotspots and frontiers. We chose the “author keywords” in the “co-occurrence” type for analysis, the counting method was “full counting,” and the minimum cluster size to 6. We merged keywords with the same meaning and deleted redundant keywords. The minimum number was set to 1, and we obtained 40 keywords in total.

Results
Characteristics of the Included Studies
We conducted a systematic search of 5 databases, which yielded 534 studies that may be relevant to the topic. After eliminating duplicates, screening 384 titles and abstracts, and reviewing 70 full texts, we found that 25 papers met the inclusion and exclusion criteria. Figure 1 shows the selection process flowchart based on PRISMA-ScR [22]. The literature included in this study comprised 12 RCTs [23-34], 8 cohort studies [35-42], 2 pilot studies [43,44], 1 longitudinal study [45], 1 cross-sectional study [46], and 1 cross-over study [47]. The 25 papers comprised 27,050 participants in total, with the sample size for each study ranging from 49 to 13,219, the duration of intervention was 1-24 months, and the follow-up period was 1-514 months. Most of the participants were older than 40 years of age and had undergone mechanical valve replacement surgery. A table in Multimedia Appendix 3 provides a summary of the study characteristics and participant demographics.

https://mhealth.jmir.org/2024/11/e48716
JMRI Mhealth Uhealth 2024 | vol. 12 | e48716 | p.72
(page number not for citation purposes)
Types of Anticoagulation Services Offered by eHealth Tools

Through the comprehensive understanding and interpretation of the included literature, we found that eHealth tools can be broadly classified into 4: computer-based support systems, electronic health records (EHRs), telemedicine platforms, and mobile apps. The application scenarios of eHealth tools are gradually expanding from large general hospitals to primary care centers and finally to patients’ homes. Most studies have reported the beneficial effects of eHealth tools on the rehabilitation outcome and anticoagulation quality in patients undergoing CVR, reflecting the potential and application prospect of eHealth tools in this field. Table 1 shows the specific types of eHealth tools and their application scenarios. Multimedia Appendix 3 shows the specific services provided by eHealth tools and their impacts on anticoagulation outcomes.
<table>
<thead>
<tr>
<th>Author, Year, Country</th>
<th>eHealth tools</th>
<th>Places</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitzmaurice et al (1996) [23], UK</td>
<td>Decision support systems (Anticoagulation Management Support System, Warwick)</td>
<td>Primary care clinic</td>
</tr>
<tr>
<td>Fitzmaurice et al (2000) [25], UK</td>
<td>Softop Information (Warwick, UK)</td>
<td>Primary care practices</td>
</tr>
<tr>
<td>Testa et al. (2006) [44], Italy</td>
<td>Electronic patient records (TaoNet, EDP-Progetti, Bolzano, Italy)</td>
<td>Peripheral health units</td>
</tr>
<tr>
<td>Poller et al (2008) [27], UK</td>
<td>PARMA5® (Instrumentation Laboratory, Milan, Italy) and DAWN AC (program 4S Dawn Clinical Software, Milnthorpe, UK)</td>
<td>Hospital</td>
</tr>
<tr>
<td>Ryan et al (2009) [28], Ireland</td>
<td>CoagCare (ZyCare Inc., Chapel Hill, NC, United States)</td>
<td>Home</td>
</tr>
<tr>
<td>Cafoilla et al (2011) [39], Italy</td>
<td>Automated computer systems coagulometer (HemoSense, San Jose, CA, United States)</td>
<td>Hospital</td>
</tr>
<tr>
<td>Bussey et al (2013) [43], United States</td>
<td>ClotFree system (ClotFree; Genesis Advanced Technologies, Lakehills, TX, United States)</td>
<td>Clinical research center</td>
</tr>
<tr>
<td>Ferrando et al (2015) [40], Spain</td>
<td>SintromacWebInternet-based system (Grifols, Barcelona, Spain)</td>
<td>Home</td>
</tr>
<tr>
<td>Cao et al (2018) [37], China</td>
<td>Digital anticoagulation clinic</td>
<td>Home and hospital</td>
</tr>
<tr>
<td>Zhu et al (2021) [32], China</td>
<td>A mobile user interface medical network follow-up platform</td>
<td>Home</td>
</tr>
<tr>
<td>Cao et al (2021) [42], China</td>
<td>Alfalfa</td>
<td>Home</td>
</tr>
<tr>
<td>Alanzzi et al (2022) [47], Saudi Arabia</td>
<td>WhatsApp</td>
<td>Home</td>
</tr>
<tr>
<td>Amruthlal et al (2022) [46], India</td>
<td>Smartphone app</td>
<td>Home</td>
</tr>
<tr>
<td>Ageno and Turpie (1998) [24], Canada</td>
<td>DAWN AC (Business Technology, Milnthorpe, Cumbria, the UK)</td>
<td>Hospital</td>
</tr>
<tr>
<td>Manotti et al (2001) [26], Italy</td>
<td>PARMA system (release 3.2, Parma, Italy)</td>
<td>Manotti et al (2001), Italy</td>
</tr>
<tr>
<td>O'Shea et al (2008) [35], United States</td>
<td>Management program (CoagCare, ZyCare Inc, Chapel Hill, NC, United States)</td>
<td>Home</td>
</tr>
<tr>
<td>Soliman Hamad et al (2009) [29], the Netherlands</td>
<td>Anticoagulant aid website</td>
<td>Home</td>
</tr>
<tr>
<td>Christensen et al (2011) [30], Denmark</td>
<td>Computer system (CSO/AC; IntraMed A/S, Værløse, Denmark)</td>
<td>Home</td>
</tr>
<tr>
<td>Thompson et al (2013) [33], United States</td>
<td>Automated computer systems coagulometer (HemoSense, San Jose, CA, United States)</td>
<td>Hospital</td>
</tr>
<tr>
<td>Koertke et al (2015) [34], Germany</td>
<td>The Institute for Applied Telemedicine</td>
<td>Heart centers</td>
</tr>
<tr>
<td>Talboom et al (2017) [38], the Netherlands</td>
<td>Portavita eHealth platform</td>
<td>Home</td>
</tr>
<tr>
<td>Jiang et al (2021) [41], China</td>
<td>Yixing App</td>
<td>Home</td>
</tr>
<tr>
<td>Jiang et al (2022) [36], China</td>
<td>Alfalfa</td>
<td>Home</td>
</tr>
<tr>
<td>Erba et al (2022) [45], Italy</td>
<td>PARMA GTS (Werfen, Milan, Italy) and WhatsApp</td>
<td>Anticoagulant clinic and home</td>
</tr>
</tbody>
</table>

*PARMA: Program for Archive, Refertation, and Monitoring of Anticoagulated patients.

Of the 25 included studies, 6 used computer-based support systems [23-27,39], 4 used EHR [33,38,43,44], 9 applied telemedicine platforms [28-32,34,35,37,40], and 6 used mobile apps [36,41,45-47]. Computer-based support systems are computer applications designed to aid clinicians in making diagnostic and therapeutic decisions in patient care [48]. Such support systems are typically installed in hospitals, large clinics for anticoagulant therapy, and primary health care centers. They can help doctors predict the appropriate dose of anticoagulants and the timing of the next international normalized ratio (INR) test based on the patient’s latest INR value [23-27,39]. EHRs are generated by 1 or more interactions between medical institutions and patients [49]. Doctors use these tools to record the patients’ medication information, laboratory measurement data, clinical history, or symptoms of discomfort for reference in the next visit. They use this information to make the best decision possible and summarize their clinical experience, which provides an important basis for improving the overall medical
level. EHR can be shared between primary health care clinics and hospitals through a central database to support CVR surgical follow-up [33].

A telemedicine platform is a means of receiving medical services remotely through various communication technologies [50]. It enables remote communication and data sharing between doctors and patients, thus providing digital medical services for patients after CVR, such as web-based consultation, real-time communication, and disease management [34,37,40,43]. Patients can use portable devices to measure INR at home and transmit the results to the telemedicine platforms. Doctors can then adjust the dosage of anticoagulants and the treatment plan remotely and determine the changes in anticoagulants and conditions of patients after CVR on time [28-30,35]. Moreover, telemedicine platforms can help to supervise the patient’s self-management regime at home for a long time, helping patients correctly understand and implement treatment plans and improving treatment compliance and outcomes [28,29,31,37].

Mobile apps are services that run on smartphones, tablets, or other mobile devices. Some have been developed specifically for AM [36,41,42,46], while others are smartphone-based messaging apps [45,47]. In the research we included, mobile apps mainly mentioned the following four services: (1) Medication assistant: the system automatically generated an oral warfarin regimen based on the patient’s latest INR value and previous warfarin dose, which were reviewed by specialists and sent to patients [42,46]. At the same time, the apps set intelligent reminders every day to urge the patient to take medicines regularly [36,41]. (2) Digital consultation: patients could communicate with doctors on the internet about their condition and anticoagulation treatment [47]. (3) Push health information: apps could send videos or pictures to help patients understand disease-related information [41,42]. (4) Patients’ interactive community: the patients could communicate with other postoperative patients, share their own rehabilitation experiences, and gain support and encouragement [41].

**Application Effects of eHealth Tools in AM**

Comprehensive considering the evaluation indexes of anticoagulation effects on patients after CVR can provide better scientific guidance for clinicians and patients. In the literature included, the evaluation of eHealth tools in AM mainly involved three aspects. (1) Clinical outcome: This indicator was reported in 24 papers, including TTR (n=19) [23-28,30,32,33,35-40,42-44,47], rate of achieving target INR (n=5) [27,29,31,42,43], bleeding or thrombotic events (n=18) [23,25,27,28,30,32,42,44,45], and mortality (n=1) [34]. Seven RCTs have demonstrated that using eHealth tools for AM in patients after CVR is a secure and efficacious approach that significantly enhances TTR (P<.05) [23,25,26,28,30,32,33]. Computer-based support systems dosing was found to be more effective in improving TTR than medical staff–monitored dosage at the majority of centers (P<.001) [24,27,42]. Meanwhile, eHealth tools had a significantly higher number of INR within the target range compared to the conventional administration group (P<.05) [27,29]. However, several studies showed no statistical difference (P>.05) in the occurrence of bleeding or thrombotic events between the eHealth tools group and the conventional group [23,25,27,30,33]. (2) Economic benefits: Five papers focused on the health economic benefits of eHealth tools, including the frequency of INR tests (n=3) [24,33,43] and cost-effectiveness (n=2) [23,25]. Research showed that the cost of using computerized decision support software was offset by not going to the hospital [23]. However, when considering the costs associated with establishing a nurse-led clinic, the total cost was higher than traditional treatment [25]. In addition, patient self-testing at home could accelerate INR results, but it is not statistically significant in reducing INR test frequency and generating economic benefits [33,34]. (3) Patient satisfaction: The conclusion of 2 studies revealed that most patients were satisfied with the use of eHealth tools (P<.001) [23,44], which also improved their quality of life [29,44]. These findings demonstrate the potential of eHealth tools in enhancing AM. However, further studies are warranted to explore the economic benefits of long-term anticoagulant therapy and determine whether it can reduce the incidence of clinical adverse events in patients after CVR.

**Scientometric Analysis**

**Publication Trends**

The earliest literature on using eHealth tools for AM in patients after CVR was published in 1996. Figure 2 shows the trend of publications, with the number of published papers gradually increasing after 2019. The purple bar above the year label represents the publication volume, and the blue line shows cumulative publication trends. China (n=5) and Italy (n=4) stand out as the 2 countries with the most published papers, the United States and the United Kingdom have each published 3 papers, and other nations have published only 1 or 2 papers. Developed countries had research published around 2000, while transitional countries, such as China and India, only paid attention to this topic around 2019, which may be influenced by cultural differences and medical and economic levels. The majority of the 25 papers in this review were published in thrombus-related journals. However, the application of eHealth tools in the AM of patients after CVR involves not only cardiology and antithrombotic therapy but also the intersection and interdisciplinary cooperation of medical informatics, electronic technology, and health sciences. Some of the papers included in this review were also published in internet medicine journals. Therefore, when seeking the latest progress, readers should not focus only on traditional thrombosis-related journals.
Collaborative Analysis of Countries, Research Institutions, and Authors

The 25 papers selected for this review originated from 12 countries and were associated with 46 institutions. Figure 3A depicts the academic collaboration among 5 of these countries, with China being the only transitional country represented. Each node represents a country, and the connection between nodes represents the cooperative relationship. Figure 3B shows that only 8 of 46 participating institutions show close cooperative relationships, and most of them are from Italy. Each node represents an institution, and the connection between nodes represents the cooperative relationship. It should be noted that there are deficiencies in the exchanges and cooperation between transitional and high-income countries, which poses challenges in enhancing the sharing and complementarity among international academic resources. Furthermore, 89 authors participated in the publication of 25 papers. Each node represents an author, the node size depends on the number of authorial publications, and the different colors of the line represent different cooperation networks (Figure 3C). Two collaborative networks have been formed centered on Palareti G, who is affiliated with the Orsola-Malpighi Hospital in Italy and plays a crucial role in the development of eHealth tools in the AM of patients after CVR.
Figure 3. Collaborative network among countries, research institutions, and authors in publications related to the use of eHealth tools for anticoagulation management (AM) in patients after cardiac valve replacement (CVR). (A) Cooperation map among countries. (B) Collaboration network among research institutions. (C) Cooperation network of authors.

Keyword Co-Occurrence Analysis and Cluster Analysis
To present the keyword map more clearly, a cluster analysis was conducted for keywords. Figure 4A indicates that current research on the application of eHealth tools in AM after CVR is mainly focused on 4 distinct areas. The node size indicates the frequency of keyword occurrence, and different colors represent different clusters. We found that 16 keywords in the red cluster were closely related to thrombus formation and bleeding, 10 keywords in the green cluster related to the innovation of INR test techniques (home testing, self-testing, etc), keywords in the blue cluster referred to artificial intelligence (AI) and machine learning, and yellow cluster included telemedicine, health care delivery, and remote consultation.
To further explore the hotspots and future directions, we superimposed time on keyword clustering to create keyword overlay visualization (Figure 4B). The larger the node size, the higher the frequency of keywords and the color of nodes corresponded to the average appearance time of keywords. According to high-frequency keyword analysis, anticoagulant drugs for patients undergoing CVR mainly included vitamin K antagonists led by warfarin. Additionally, the research hotspots of eHealth tools in AM have focused on AI, machine learning, and smartphone apps in recent years. Amruthal et al [46] constructed warfarin dosage prediction models by linear regression, support vector regression, logistic regression, and multilayer perceptron and found that the prediction model based on support vector regression showed the best predictive performance. Support vector regression was installed in a simple user-friendly Android mobile app. The introduction of advanced
machine learning algorithms provides more choice and support for AM, which may help to improve the efficiency and accuracy of anticoagulant treatment.

**Discussion**

**Summary**

This review provides a full evaluation and scientometric analysis of the use of eHealth tools in the AM of patients undergoing CVR surgery. We identified and included 25 studies investigating eHealth tools, such as computer-based support systems, EHRs, telemedicine platforms, and mobile apps. The development of eHealth tools is becoming increasingly important in promoting AM, and the emergence of machine learning and AI has introduced new opportunities for innovation and promotion. Over the past decade, there has been a lack of large-sample and long-term follow-up RCT studies on AM based on eHealth tools, highlighting the need for further research to verify their roles and effects. There remain many areas for improvement to promote the application of eHealth tools in the AM of patients after CVR.

**Key Developments and Benefits Provided by eHealth Tools for AM**

The AM of patients after CVR involves a large workload, complex patient information, high technical content, and a long follow-up period. The establishment and application of eHealth tools can improve the efficiency of clinical and management work [36,41,48,49]. The computer-based support system facilitates the standardization of anticoagulant recommendations and avoids differences in the dosage prescribed by medical staff, especially in primary health care institutions lacking anticoagulation therapy experience [51]. However, previously, due to technological limitations, these systems could not closely approximate doctors’ judgment and decision-making skills, and manual verification and intervention were needed to ensure accuracy [24]. With the development of machine learning, the use of deep learning, reinforcement learning, and ensemble learning are increasing gradually in warfarin management after CVR, which can further improve the accuracy of prediction [52,53]. Additionally, doctors can adjust anticoagulant doses of patients more accurately according to their EHR, thus avoiding excessive or insufficient therapy and reducing the occurrence of complications. Although these 2 eHealth tools are highly targeted, easy to operate, and improve work efficiency, their functions are relatively simple and not enough for more comprehensive management of patients undergoing anticoagulant therapy after CVR.

Telemedicine platforms and mobile apps offer a promising solution for expanding access to high-quality medical resources by providing a digital communication channel between clinicians and patients who have undergone CVR [32-36,38,39,41,44]. Through these 2 ways, doctors can provide electronic prescriptions, digital consultations, health education, and self-management support to not only help patients better understand postoperative rehabilitation but also develop their knowledge and skills and improve the compliance, safety, and effectiveness of AM. However, telemedicine platforms depend on high-quality communication devices to be compatible with other medical devices and systems to perform remote processing of medical data and real-time monitoring of patients. In contrast, mobile apps are more portable and can be used anywhere and anywhere, with a more intuitive interface and personalized services for anticoagulant therapy [39,44-47]. However, the older population, who comprise the majority of patients undergoing CVR surgery, may have difficulty learning and adapting to mobile apps and data on the internet. This may affect the accuracy of communication and information transfer between doctors and patients and limit postoperative follow-up. Additionally, practical limitations, such as network instability, data security, low degree of automation, and the need for medical institution support, could affect the effectiveness of telemedicine platforms and mobile apps.

**Research on Hotspots of eHealth Tools Based on Scientometric Analysis**

Research on the use of eHealth tools for AM after CVR surgery is lacking on a global scale. Developed nations have conducted more studies in this area, which may be attributed to their advanced economic status, superior medical quality, technical proficiency, and greater patient acceptance [54,55]. To promote the development of eHealth tools, it is suggested that countries provide a platform for exchanging and sharing the latest research results and invite scholars from different fields to participate in medical exchange programs, visiting scholar programs, medical conferences, and seminars. Moreover, universities can encourage students to effectively establish a cooperative culture, organize interdisciplinary research groups, conduct interdisciplinary training and education, and cultivate medical talents with an interdisciplinary background and international vision.

Based on the results of keyword clustering and overlay visualization, AI and machine learning have been adopted as research methods for developing eHealth tools and become a hotspot in recent years. These methods have provided valuable insights for future research. Genetic algorithms, backpropagation neural networks, and adapted neural-fuzzy inference system models have the potential to establish more accurate and stable prediction models of warfarin individual maintenance dose for patients after CVR [52,56]. Medical professionals can use biometric technology to visually identify the patient, the medication, and the confirmed ingestion, thereby reducing the risk of noncompliance in patients’ anticoagulant therapy [57]. To promote the wide application and sustainable development of eHealth tools, it is necessary to research intelligent auxiliary tools, combining sophisticated machine learning techniques and AI models to enhance the precision and dependability of prediction and innovate anticoagulant therapy tools.

**Future Directions for eHealth Tools in AM of Patients After CVR**

In 25 studies included, although the main outcome indicators of eHealth tools in AM covered most of the clinical indicators, they could not fully reflect the overall situation of patients from the perspectives of economy, society, psychology, and user experience. Therefore, it is suggested that improvements can be made in the following aspects: (1) Patient interaction indicators: Researchers can monitor the frequency of patients
logging in or using eHealth tools in the background, learn about their experience and needs, and then adjust the tool design and functions accordingly. For example, a user interface with voice support, large font, and novice guidance can help patients overcome the potential obstacles in using eHealth tools [58].

(2) Indicators of acceptability: Researchers should evaluate the acceptance of patients and health care providers using patient-physician satisfaction, use rate, and patient-medications compliance. (3) Anticoagulant dosage index: This can be used to evaluate the accuracy and consistency of manual administration and eHealth tool–assisted administration to ensure the correctness of AM decisions. (4) Sustainability indicators: Patients can evaluate the user experience of eHealth tools from engagement, function, esthetics, and information quality to ensure long-term use and promotion [59].

In addition to the aforementioned optimization metrics, several other important factors should be considered to facilitate eHealth tools. First, it is critical to ensure that the design of eHealth tools is compatible with other medical devices and systems that the patients may use (eg, blood clotting machines and electrocardiographs). To minimize errors and improve data accuracy, these tools should connect seamlessly, allowing data to be shared and integrated across different platforms. Lubitz et al [60] used compatible, wearable devices and Android or iOS smartphones–assisted electrocardiogram patch monitoring, which effectively identified undiagnosed atrial fibrillation at an early stage. Second, providing effective training and support to older or less educated patients is critical for increasing the acceptance and use of eHealth tools. Previous studies have shown that tailored coaching and education programs can help patients operate these tools effectively [61]. Third, the use of advanced machine learning algorithms and AI models helps further improve the accuracy and reliability of eHealth tools. Zeng et al [62] used various machine learning to construct a dynamic anticoagulant treatment scheme for hospitalized patients after CVR surgery. The results showed that the performance of reinforcement learning was significantly better than the other algorithms, and the quality of anticoagulation was significantly optimized [62]. Finally, implementing effective evaluation mechanisms for patients lost to follow-up is critical to ensure they receive appropriate anticoagulant treatment over the long term. Porter et al [63] conducted a 2-year follow-up of patients receiving oral anticoagulants and found that 12-week INR test intervals were feasible, which saved patient’s time and improved their compliance [63]. By addressing these factors and optimizing the application of eHealth tools in the AM of patients after CVR, it is possible to significantly improve the overall therapeutic effect and reduce the risk of complications.

Limitations
This review has some limitations. First, we did not assess the quality of the included studies because the study was to provide a comprehensive overview of the existing research landscape related to the application of eHealth tools in AM for patients after CVR. However, we believe that evaluating the quality of the literature will increase the significance of the study. Second, we did not include studies published in the form of other sources, such as qualitative studies, because we were unable to identify eligible studies. Third, concluding trends based on 25 papers are limited, and the conclusions of scientometric analysis should be interpreted with caution. The scientometric analysis in our research is influenced by several limitations, such as the omission of qualitative data, language barriers, and the potential for sample bias. To address the constraints associated with scientometric analysis, researchers can broaden their data sources, integrate qualitative data, and encompass papers published in various languages.

Conclusions
The application of AM based on eHealth tools is expected to truly reflect the social and economic benefits of digital intelligence, thus benefiting patients undergoing CVR. This study provides an overview of the scope, benefits, and future development of eHealth tools in AM for researchers, health care professionals, and post-CVR patients. However, the economic benefits and long-term impact of adverse events need further explored. It is suggested that future eHealth tools in AM should concentrate on enhancing patient contact, acceptance, cost-effectiveness, and sustainability while combining sophisticated algorithms to enhance the precision and dependability of eHealth tools.

Acknowledgments
The authors specially thank the National Natural Science Foundation of China, Hunan Provincial Natural Science Foundation, and Hunan Normal University. This work was supported by the National Natural Science Foundation of China (82100490 to LC), Hunan Provincial Natural Science Foundation (2023JJ30425 to LC), Hunan Normal University Students’ Innovation and Entrepreneurship Training Program (2022263 to MZ), and Hunan Normal University School of Medicine Open Foundation Project (KF2022043 to XW).

Data Availability
The data sets used and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
References


Abbreviations

AI: artificial intelligence
AM: anticoagulation management
CVR: cardiac valve replacement
EHR: electronic health record
INR: international normalized ratio
RCT: randomized controlled trial
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Review
TTR: time in therapy range
Dissemination Strategies for mHealth Apps: Systematic Review

Henri Claude Moungui¹, MPH, MSc; Hugues Clotaire Nana-Djeunga², MSc, PhD; Che Frankline Anyiang³, MSc; Mireia Cano⁴, MSc; Jose Antonio Ruiz Postigo⁵, MD, PhD; Carme Carrion⁴, MSc, PhD

¹Universitat Oberta de Catalunya, Barcelona, Spain
²Higher Institute for Scientific and Medical Research, Yaounde, Cameroon
³Texila American University, Georgetown, Guyana
⁴eHealth Lab Research Group, eHealth Center & School of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
⁵Prevention, Treatment and Care Unit, Department of Control of Neglected Tropical Diseases, World Health Organization, Geneva, Switzerland

Abstract

Background: Among the millions of mobile apps in existence, thousands fall under the category of mobile health (mHealth). Although the utility of mHealth apps has been demonstrated for disease diagnosis, treatment data management, and health promotion strategies, to be effective they must reach and be used by their target audience. An appropriate marketing strategy can ensure that apps reach potential users and potentially convert them to actual users. Such a strategy requires definitions of target end users, communication channels, and advertising content, as well as a timeline for effectively reaching and motivating end users to adopt and maintain engagement with the mHealth app.

Objective: The aim of this study was to identify strategies and elements that ensure that end users adopt and remain engaged with mHealth apps.

Methods: A systematic search of the PubMed, PsycINFO, Scopus, and CINAHL databases was conducted for suitable studies published between January 1, 2018, and September 30, 2022. Two researchers independently screened studies for inclusion, extracted data, and assessed the risk of bias. The main outcome was dissemination strategies for mHealth apps.

Results: Of the 648 papers retrieved from the selected databases, only 10 (1.5%) met the inclusion criteria. The marketing strategies used in these studies to inform potential users of the existence of mHealth apps and motivate download included both paid and unpaid strategies and used various channels, including social media, emails, printed posters, and face-to-face communication. Most of the studies reported a combination of marketing concepts used to advertise their mHealth apps. Advertising messages included instructions on where and how to download and install the apps. In most of the studies (6/10, 60%), instructions were oriented toward how to use the apps and maintain engagement with a health intervention. The most frequently used paid marketing platform was Facebook Ads Manager (2/10, 20%). Advertising performance was influenced by many factors, including but not limited to advertising content. In 1 (10%) of the 10 studies, animated graphics generated the greatest number of clicks compared with other image types. The metrics used to assess marketing strategy effectiveness were number of downloads; nonuse rate; dropout rate; adherence rate; duration of app use; and app usability over days, weeks, or months. Additional indicators such as cost per click, cost per install, and clickthrough rate were mainly used to assess the cost-effectiveness of paid marketing campaigns.

Conclusions: mHealth apps can be disseminated via paid and unpaid marketing strategies using various communication channels. The effects of these strategies are reflected in download numbers and user engagement with mHealth apps. Further research could provide guidance on a framework for disseminating mHealth apps and encouraging their routine use.

(JMIR Mhealth Uhealth 2024;12:e50293) doi:10.2196/50293
Introduction

Mobile Health Apps

Among the vast array of mobile apps currently available, health care apps serve various purposes, including disease diagnosis [1], health promotion, and disease prevention [2,3]. Such uses of mobile technology to provide patients with health care support or health service providers with technical support in a direct, low-cost, and engaging manner fall under the category of mobile health (mHealth) [4].

With approximately 200 new mHealth apps released every day, the number available now exceeds 300,000 [5]. One factor in this proliferation is the increasing use of mHealth technology by health service providers who not only seek advice from apps but also prescribe them to their patients [6]. In October 2020, Germany became the first country to cover the prescription costs of certain mHealth apps through statutory health insurance [7]. Moreover, because mHealth apps have the potential to replace a number of health provider tasks, it has been suggested that expertise in the use of mHealth-related technologies should be recognized as an essential competency for providers [1].

Dissemination of mHealth Apps to Users

The amount of academic research on mHealth apps has also increased, particularly in the areas of usability, effectiveness, adoption, and assessment. However, the highly important aspects of dissemination and marketing are as yet underexamined. App marketing refers to measures aimed at making a mobile app better known and acquiring users (ie, generating app downloads) and, moreover, contacting users and encouraging them to reach a specified goal [8].

An mHealth app is not a stand-alone product that can work effectively without human interaction, which cannot take place if users are unaware that a certain app exists and is accessible. Such “human touches,” although extraneous to the app itself, can be crucial for promoting use [9].

As users are key actors in mHealth adoption, it is critical to understand how they navigate the various stages from app discovery to frequent use. Google has created such a model [10]. It consists of four key stages: (1) discover (users come across an app and download it to their device), (2) onboard (the process of first use and registration), (3) engage (users start using the app regularly), and (4) embed (the desired outcome as users view the app as “vital” to their lives). Only a small proportion of users currently reach the embedment stage with any app [10]; for instance, the literature is sparse regarding the long-term integration and penetration of mobile interventions within mental health and other support service settings [11].

Increasing the chances of an app achieving embedment requires understanding users and placing them at the core of mHealth services. This process would start with persona definitions: fictional archetypes of actual product users. A persona enables program designers to create high-quality programs that effectively meet user needs [2]. In the marketing world, this also means segmenting users and locating them on the marketing funnel, which is a visual representation of the different phases in a customer’s journey toward conversion and their relationship with a product. By segmenting customers based on where they are located in the funnel, marketers target these groups much more effectively [12].

Positive customer experiences and journeys rely on ensuring that the consumer sees value in an app as a channel for accessing products and services and as a 2-way platform for seamless interactions. Although marketing strategies play a crucial role during the early stages of the customer journey, they have been subjected to very little analysis [13].

Marketing of Health Apps

A successful marketing strategy can ensure that an mHealth app reaches potential users and ease the adoption process. Such a strategy would clearly define target end users; determine the appropriate communication channels, content, and timelines to effectively reach users; and market the app as an attractive product, encouraging people to download it and become regular users. The strategy would include a mix of activities, depending upon the type of app and upon the stage of the launch period (from before the launch to after the launch), including email marketing, targeted advertising, and social media promotion [14].

Marketing services have evolved alongside information and communication technologies. In turn, digital marketing has provided a series of customized platforms for communicating with specific stakeholders using computers, smartphones, and tablet computers [15]. These channels enable information to be gathered and include websites as well as various social media platforms such as Facebook, YouTube, X (the platform formerly known as Twitter), Pinterest, TikTok, and LinkedIn. Traditional marketing also remains an option, with products being promoted on radio and television channels, as well as via printed posters in public spaces, flyers, and face-to-face conversations [16].

The cost of promoting an app will depend upon where money is spent, and those promotional activities that do not cost money will demand time. Although potential customers may be offered incentives such as money or supplemental products, the marketer or marketing firm involved in digital marketing can be offered a fee per click, download, or install. A small pilot trial of activities that cost money is recommended to assess results [14].

Specialized services are available for driving digital campaigns; for example, Facebook Ads Manager is a paid service that oversees paid digital marketing campaigns across the Facebook platform. Google Universal App Campaigns (UAC) is another paid service that promotes mobile apps by distributing marketing messages across Google formats and networks, such as the first page of applicable Google search results and small banner advertisements on relevant YouTube channels [17]. As Google
shares information among platforms, including Google Display Network, YouTube, and Google Play Store, Google UAC can capture the number of Google-driven impressions, clicks, and installs on Android devices.

The effectiveness of a marketing strategy can be observed not only through the number of app downloads but also by the effects of users interacting with the app. Referred to as mobile app engagement, this is defined by a variety of operational metrics, such as the number of log-ins, the number of days of use, the number of pages visited, and the number of tasks or modules completed [18]. Another gauge of marketing effectiveness is user onboarding. In the context of mobile apps, user onboarding is the process of providing instructions and highlighting key benefits and features via a set of example screens when the user first launches the app [19].

Objectives

The aim of this study was to review existing evidence on strategies and elements relevant to how the dissemination of mHealth apps is currently carried out and how these elements contribute to encourage user onboarding and engagement with mHealth apps.

Methods

Overview

This study was carried out following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [20] (Multimedia Appendix 1). The protocol of this review was registered with PROSPERO (CRD42022352369) [21].

Eligibility Criteria

Eligible sources were academic papers. All cohort studies, cross-sectional studies, and randomized controlled trials reporting on the dissemination and marketing of an mHealth app were included. The outcome expected from included studies was strategies and elements relevant to how the dissemination of mHealth apps is currently carried out and how these elements contribute to encourage user onboarding and engagement with mHealth apps.

Editorials, letters to the editor, scoping reviews, systematic reviews, meta-analyses, conference abstracts, commentaries, essays, book chapters, and study protocols were excluded, as were studies with any other study design (eg, bibliometric analysis, modeling study, systematic or web search or review of apps, landscape analysis, and scorecard analysis). We also excluded studies with participants aged <18 years and those not reporting the expected outcome. Search languages were limited to French and English.

Information Sources and Search Strategy

Searches were conducted on PubMed, PsycINFO, Scopus, and CINAHL databases over the 5-year period from January 1, 2018, to September 30, 2022 (refer to Multimedia Appendix 2 for the search strategy). The following search terms were used individually or combined according to Medical Subject Headings terms: “apps,” “mHealth,” “marketing,” “promotion,” and “dissemination.” Moreover, we conducted searches on JMIR and mHealth journals and cross-checked the reference lists of the selected studies to locate additional studies that met the inclusion criteria. The main outcome was dissemination strategies for mHealth apps.

Study Selection and Data Collection Process

All retrieved studies were imported into Rayyan (Rayyan Systems Inc) [22] and duplicate records eliminated. Screening consisted of blind peer review by 2 independent investigators. Any conflict was resolved through discussion or the adjudication of a third investigator.

Results

Selection of Studies

We identified 638 records through database searches (PubMed: n=215, 33.7%; PsycINFO: n=60, 9.4%; Scopus: n=283, 44.4%; and CINAHL: n=80, 12.5%) and 10 records through JMIR and mHealth journal searches. Of the total 648 records, 127 (19.6%) duplicates were removed. Next, of the remaining 521 articles, 502 (96.4%) were removed after title and abstract screening. The reasons for exclusion were unrelated outcome or outcome other than the subject of our review (309/502, 61.6%), study protocol (136/502, 27.1%), published review (48/502, 9.6%), study participants aged <18 years (6/502, 1.2%), and other study design (3/502, 0.6%). We then assessed the remaining 19 full texts for eligibility and excluded 9 (47%) for unrelated outcomes; thus, 10 (53%) papers were included in this review. A PRISMA-compliant flow diagram [20] of the paper selection process is shown in Figure 1. The characteristics of the studies selected are summarized in Table 1.
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the paper selection procedure.

Records identified: 638
- Database searched: 638
- Additional sources: 10

Duplicate records removed before screening (n=127)

Records excluded (n=502)
- Reasons for exclusion:
  - Unrelated outcome (n=309)
  - Protocol (n=116)
  - Review (n=48)
  - Participants aged <18 y (n=6)
  - Other study design (n=3)

Titles and abstracts screened (n=521)

Full texts assessed for eligibility (n=19)

Full texts excluded (n=9)
- Reason for exclusion:
  - Unrelated outcome (n=9)

Studies included (n=10)
<table>
<thead>
<tr>
<th>Authors, year; country</th>
<th>Population</th>
<th>Intervention</th>
<th>Outcomes</th>
<th>Design</th>
<th>Quality</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kvedarienė et al [23], 2019; Lithuania</td>
<td>N=149</td>
<td>Monitoring of allergic rhinitis and asthma in real life in Lithuanian MASK - air app users</td>
<td>High app engagement User retention rate was 107 days of use/user Patients are satisfied with the app overall</td>
<td>Longitudinal study</td>
<td>Low</td>
<td>Selection bias: patients recruited from an allergy clinic</td>
</tr>
<tr>
<td>Buss et al [24], 2022; Australia</td>
<td>N=46</td>
<td>Test the feasibility of an app-based intervention for cardiovascular and diabetes risk awareness and prevention</td>
<td>The app scored the highest for the information section and the lowest for the engagement section of the scale Nonuse and dropouts were too high, and adherence was too low to consider the intervention in its current form feasible Asking people aged ≥45 years to download the app and expect them to use it over 3 months without additional interaction was not feasible</td>
<td>Nonrandomized controlled trial (cohort study)</td>
<td>Low</td>
<td>Small sample size Nonuse and dropouts too high, adequacy too low</td>
</tr>
<tr>
<td>Arshanapally et al [17], 2022; United States</td>
<td>N=NI²</td>
<td>Investigate the outcomes of a paid digital marketing campaign to promote an mHealth³ app about parent-engaged developmental monitoring</td>
<td>Paid digital marketing can be an effective strategy to promote mHealth apps targeting parents of young children Google-driven marketing messages in English had a higher clickthrough rate than those in Spanish</td>
<td>Implementation study</td>
<td>Low</td>
<td>No relevant limitations</td>
</tr>
<tr>
<td>Resnick et al [25], 2021; United States</td>
<td>N=41</td>
<td>Assess the usability, acceptability, and user engagement of the Healthier Together mobile app</td>
<td>The app strongly engaged participants, with promising results on participants' knowledge of cancer prevention behaviors and success in achieving their cancer prevention behavioral goals</td>
<td>Mixed methods intervention</td>
<td>Low</td>
<td>Small sample size Nonprobabilistic purposive sample of non-Hispanic Black patients at 2 internal medicine primary care clinics Participants were incentivized US $40 to complete the baseline, in-person, 40-month enrollment process and interview and US $60 for completing the 45-month exit telephone interview</td>
</tr>
<tr>
<td>Authors, year; country</td>
<td>Population</td>
<td>Intervention</td>
<td>Outcomes</td>
<td>Design</td>
<td>Quality</td>
<td>Limitations</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------</td>
<td>--------------</td>
<td>----------</td>
<td>--------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>Zlotorzynska et al [26], 2021; United States</td>
<td>• N=NI</td>
<td>• Sex: 0% female</td>
<td>• Age: 18-24 years</td>
<td>• Other details: YMSM</td>
<td>Paid web-based recruitment campaign to recruit HIV-negative or unknown status YMSM for 4 randomized controlled trials of mHealth HIV prevention interventions</td>
<td>• Instagram advertisements yielded the highest proportions of eligible contacts who were racial or ethnic minority individuals and aged &lt;18 years</td>
</tr>
<tr>
<td>Rajani et al [27], 2021; United Kingdom</td>
<td>• N=154</td>
<td>• Sex: 38.8% female</td>
<td>• Age: 18-65 years</td>
<td>• Other details: smokers</td>
<td>Examining the association among perceived usefulness, perceived ease of use, and frequency of use of gamification features embedded in smoking cessation apps on self-efficacy and motivation to quit smoking</td>
<td>• The use of the apps was associated with increased self-efficacy and motivation-to-quit levels 4 weeks after app use compared with baseline</td>
</tr>
<tr>
<td>Roberts et al [28], 2019; United Kingdom</td>
<td>• N=32</td>
<td>• Sex: 68.8% female</td>
<td>• Age: ≥18 (mean 60, SD 11; range 37-78) years</td>
<td>• Other details: diagnosed with breast, prostate, or colorectal cancer</td>
<td>To seek opinions of survivors of breast, prostate, and colorectal cancer regarding using apps to promote PA</td>
<td>• Multiple factors affect engagement with PA apps, and this is highly personalized</td>
</tr>
<tr>
<td>Bidargaddi et al [29], 2018; United States and Australia</td>
<td>• N=1255</td>
<td>• Sex: NI</td>
<td>• Age: NI</td>
<td>• Other details: NI</td>
<td>To study the effect of time-varying push notifications on engagement in self-monitoring activity</td>
<td>• Pushing a notification with a tailored health message affects near-time proximal engagement with the self-monitoring activity in the app</td>
</tr>
<tr>
<td>Hui et al [30], 2018; United States</td>
<td>• N=101</td>
<td>• Sex: 87.1% female</td>
<td>• Age: ≥16 years</td>
<td>• Other details: patients with active asthma</td>
<td>The impact of different recruitment strategies and app features on adoption and continued use</td>
<td>• Adherence was dependent upon motivation derived from a sense that the health care professional and researcher were interested in the results and that using an app to support their self-management could improve their asthma control</td>
</tr>
<tr>
<td>Moungui et al</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Market tests of the CycleBeads app in 7 countries</td>
<td></td>
</tr>
</tbody>
</table>
Of the 10 selected studies, 6 (60%) were observational studies (longitudinal, cohort, implementation study, mixed methods, and cross-sectional), 1 (10%) reported on 4 randomized controlled trials [26], 1 (10%) reported on in-app microsurveys [31], and 1 (10%) reported on a microrandomized trial [29]. According to the Scottish Intercollegiate Guidelines Network (SIGN) [32] criteria, 8 (80%) of the 10 studies were of low quality, 1 (10%) of moderate quality, and 1 (10%) of high quality. A low-quality rating resulted from a small sample size, the study design (mostly observational), or possible selection bias. A summary of the design, quality, and limitations of the included studies can be found in Table 1.

The selected 10 studies covered 8 countries—Lithuania, the United Kingdom, Australia, the United States, Egypt, Ghana, India, and Jordan—with 50% (5/10) of the studies conducted in the United States and 20% (2/10) in the United Kingdom. The number of participants in the selected studies ranged from 32 to 1255. Of the 10 studies, 3 (30%) that recruited participants through social media and used impressions and clicks as a proxy measure of their number did not state the number of participants. In 6 (60%) of the 10 studies, 50% to 100% of the participants were female; sex information was not given in 2 (20%) of the 10 papers, 1 (10%) study targeted adolescent and young male individuals, and 1 (10%) targeted female individuals only.

Objectives of the Marketing Strategy

In each of the studies reviewed, we looked for the objective of the marketing strategy implemented through the lens of the Google model of mobile app user journey (discover, onboard, engage, and embed) [10] (Table 2). We found that 60% (6/10) of the studies aimed to drive people to discover their mHealth apps (come across the apps and download them), get onboard (start using the apps), and stay engaged by using the apps for a couple of weeks or months. Marketing strategy in 30% (3/10) of the studies targeted the discover and onboard stages, whereas 1 (10%) of the 10 studies aimed at the engage stage only.
Table 2. Summarized results of reviewed studies.

<table>
<thead>
<tr>
<th>Authors, year</th>
<th>Channels</th>
<th>Procedures of each dissemination strategy: How was it done?</th>
<th>Effects and results of each dissemination strategy</th>
<th>Objective with regard to the Google model for users’ journey with apps</th>
</tr>
</thead>
</table>
| Kvedarienė et al [23], 2019 | Face-to-face | • Health personnel trained patients how to use the app at a clinic | • The duration of app use in patients ranged from 1 to 680 (median 54, IQR 23-151) days  
• There was an increased frequency when the reported days were >200 (18%)  
• Adherence to mobile apps was higher when the app was promoted by physicians and when the users were taught how to use it | Discover, onboard, and engage |
| Buss et al [24], 2022 | Email | • Participants received an email and a user guide that included instructions to download an app from the app store on their mobile phones and then use it for 3 months  
• Researchers encouraged regular use  
• If they had questions or technical issues, participants could get in touch with researchers via email | • Of the 46 participants, 20 (43%) never used the app, and 15 (33%) dropped out  
• Adherence to app use (using the app at least once/week over 3 months) was 17% (8/46)  
• The median time between the first and last app use was 54 days  
• The research team did not actively engage with participants early in the study to verify that all participants could install the app  
• The intervention did not involve direct contact with health care professionals  
• The app did not contain enough interactive features | Discover, onboard, and engage |
| Arshanapally et al [17], 2022 | Social media | • Google- or Facebook-driven marketing message: when a user clicked on a marketing message, they were directed to either the Google Play Store or the Apple App Store, depending upon device type  
• Google Universal App Campaigns distributed marketing messages across several Google formats and networks  
• Facebook Ads Manager distributed marketing messages across the Facebook platform | • The Google-driven marketing messages garnered a total of 4,879,722 impressions and 73,956 clicks (clickthrough rate: 1.52%); from these clicks, there were 13,707 installs of the app on Google Play Store (18.53% install rate); the overall cost/install was US $0.93  
• The Facebook-driven marketing messages garnered a total of 2,434,320 impressions and 44,698 clicks (clickthrough rate: 1.84%); the average cost/install for Facebook could not be calculated because of the limitation of collecting Facebook-driven installs data  
• Animated graphics generated the greatest number of clicks among both English and Spanish audiences on Facebook when compared with other types of images | Discover, onboard, and engage |
| Resnick et al [25], 2021 | Face-to-face | • In-person information about the app, followed by installation of the mobile app on the participant’s mobile phone and instructions on how to select a goal, choose share settings, and invite other social ties  
• Weekly SMS text messages reminding participants to check in, share goal progress, and invite relatives | • Of the 171 participants, 41 (24%) downloaded the app  
• Engagement with the app (mean number of check-ins/participant out of 8 possible weekly check-ins during the study period) was 5.7  
• Of the 41 participants, 31 (76%) checked in at least 4 of the 8 weeks  
• The mean System Usability Scale score was 87 (SD 12; median 90, IQR 78-95) | Discover, onboard, and engage |
<table>
<thead>
<tr>
<th>Authors, year</th>
<th>Channels</th>
<th>Procedures of each dissemination strategy: How was it done?</th>
<th>Effects and results of each dissemination strategy</th>
<th>Objective with regard to the Google model for users’ journey with apps</th>
</tr>
</thead>
</table>
| Zlotorzynska et al [26], 2021 | Advertising on Grindr, Snapchat, Instagram, and Facebook | ● Advertising was purchased on Facebook, Instagram, Snapchat, Twitter, and Grindr; users who clicked on banner advertisements were taken directly to a study-specific eligibility screener and, if eligible, were asked to provide contact information for follow-up by respective study site staff  
● The Facebook Ads Manager proprietary algorithm allocated the distribution of advertisement placements (news feed, right-hand column, and Instagram feed and stories) that would provide the best performance  
● Advertising copy was written to give a very brief description of the study, identify the organization conducting study recruitment, mention the study incentive, and provide a call to action for potential participants seeking to learn more  
● Snapchat advertisements were created using Snap Publisher and were in the form of short videos up to 10 seconds long; these videos rotated through various stock photo images and superimposed text  
● Instagram placements were used to test advertisement performance  
● Images used in advertisements included stock photos, which were either purchased from web-based stock photo vendors (eg, Shutterstock) or accessed through Facebook’s free stock photo catalog integrated within the Facebook Ads Manager  
● Recruitment through various in-person and community outreach efforts included posting flyers, tabling at community events, reaching out to past study participants, and recruiting through clinics serving youth | ● Grindr and Snapchat advertisements produced the highest clickthrough rate compared with Facebook advertisements; these advertisements had the lowest proportions of users who initiated eligibility screeners  
● Facebook advertisements yielded the lowest cost/eligible contact, whereas advertisements on Twitter had the highest  
● Facebook or Instagram advertisements had much higher rates of screening and ultimately yielded much lower costs/eligible participant  
● The cost/eligible participant was markedly higher for the Instagram advertisements campaigns | Discover and onboard |
| Rajani et al [27], 2021 | Social media and paper posters | ● Participants were recruited via social media, and posters were displayed in public places in London, United Kingdom  
● Participants were provided instructions on the internet on how to download and start using the app | ● Of the 138 participants who installed the app, 116 (84.1%) completed all 4 weeks of the study  
● A 1-point increase in the average perceived frequency of gamification features was statistically significantly associated with a 3.35-point increase in self-efficacy from baseline to study end ($\beta$=3.35, 95% CI 0.31-6.40) | Discover and onboard |
<p>| Roberts et al [28], 2019 | Face-to-face, paper posters, email, and social media | | | Discover, onboard, and engage |</p>
<table>
<thead>
<tr>
<th>Authors, year</th>
<th>Channels</th>
<th>Procedures of each dissemination strategy: How was it done?</th>
<th>Effects and results of each dissemination strategy</th>
<th>Objective with regard to the Google model for users' journey with apps</th>
</tr>
</thead>
</table>
| Bidargaddi et al [29], 2018 | Push notifications | • Push notifications were sent at 1 of 6 chosen time points throughout the day, and a user could either receive or not receive a push notification at a chosen time point  
• At each considered time point, users were randomized to either receive or not receive a push notification containing a tailored health message with a 50% probability; once a time point was considered, the user was then considered unavailable for the remainder of the day  
• To mitigate the risk of users either turning off notifications or deleting the app owing to receiving too many push notifications, users were classified as either “available” or “unavailable” at each time point, and only those time points when users were “available” were considered for the push notification decision; several rules were applied to determine availability | • Of the 40 participants, 32 (80%) completed the study (dropout rate: 20%)  
• Reasons for dropping out were lack of time, family circumstances (eg, bereavement), and not wanting to update their smartphone’s operating system or register credit card details with Google Play  
• Factors affecting engagement included participants’ perceptions of the advantages and disadvantages of using apps to support physical activity, the relevance of the app to the user, the quality of the app, and the behavior change techniques used to promote physical activity | Engage |
| Hui et al [30], 2018 | Social media and face-to-face | • Researchers sent letters inviting people to use the app for up to 3 months  
• Practice nurses approached adults with active asthma to try out a prototype app  
• The app was also advertised via social media (Facebook and Twitter [now known as X])  
• Nurses were not expected to teach patients to download and use the app  
• Technological support was provided by the research team and the app developer  
• The messages on social media included brief information and a link to information for patients on how to download the app | • A total of 300 patients received information on the app and study from Facebook, Twitter, or organic searches  
• Only 135 patients downloaded the app, with 111 (82.2%) registering an account on the app  
• Social media attracted 87 users, but only 15 (17%) used the app for 30 days  
• A total of 24 patients were recruited, and 13 (54%) continued for 30 days  
• Successful adoption was dependent upon the ease of downloading and sufficient motivation  
• Some patients needed technological assistance with downloading the app and starting to use the features | Discover, onboard, and engage |
| Haile et al [31], 2018 | Social media and face-to-face | Series of culturally appropriate Facebook campaigns for each country of interest | | Discover and onboard |
Channels and Procedures of Each mHealth App Dissemination Strategy

As shown in Table 2, mHealth app promotion was conducted through various channels, including paper posters, face-to-face communications, email, and social media. The most reported channels were face-to-face [23,25,28,30,31] and social media [17,27,28,30,31], both of which appeared in 50% (5/10) of the studies. Messages were intended to inform people about the existence of the app, what it does and the benefits of using it, where to find the downloading link, downloading instructions, and information on how to use the app. Messages on these channels used text, voice-overs, and short videos. Messages on social media and email included a shortened URL to download the app. Face-to-face interactions [23,25,28,30,31] provided the opportunity for training support, including demonstrating app features, installing and using the app, and solving technical issues.

Google Play Store and Apple App Store were the main web repositories to which potential users were redirected for download. In addition, the social media pages of various organizations, such as the Facebook pages of cancer support groups [28], were also used to promote apps and provide download links.

Effects of Dissemination Strategies

Advertising performance was influenced by many factors, including but not limited to advertising content and creativity, competing advertisements within platforms, and emerging platforms that attracted users to new digital spaces [26]. The reviewed papers revealed that social media advertisements attracted substantial downloads over a short period with a large number of impressions [13] and prompted patients to respond to invitations to use the app. However, without the motivation provided by a trusted professional, users quickly ceased their use of the app; a dual promotion strategy was then used to increase app adoption: using social media for its reach and ease of response as well as promoting personal invitations from trusted health care professionals [25].

In 1 (10%) of the 10 studies, embedded push notifications were sent to users to keep them engaged with the app [24]. This effect was sustained over time, and push notifications containing tailored health messages reduced the rate at which users disengaged. The timing of notifications was a key factor, with the results suggesting that users were more likely to engage with an app within 24 hours when push notifications were sent at midday on a weekend. Animated graphics generated a greater number of clicks than other image types [17].

The success of mHealth app dissemination strategies in these studies was measured in terms of adherence, app use, the duration of app use, impressions, clicks, app downloads, and installs. Marketing messages with simple direct calls to action, such as “Track your child’s development,” generated high clickthrough rates (CTRs) [17].

Marketing Messages and Content of Communications

In some cases, promotional messages included questions prompting people to download and test the app, such as “Would you like to test an app that helps you manage your asthma better? [shortened URL]” [30]. One study included the inquiring headline “Ever think of testing at home?” and explanatory body text: “Fenway Health is looking for young men to help test out a mobile app to support sexual health” [26].

In a study promoting an mHealth app about parent-engaged developmental monitoring [17], marketing messages with animated graphic interchange format (GIF) images as well as images of younger children performed higher than messages without GIF images and images of older children.

Paid or Nonpaid Marketing Campaigns

The paid marketing campaign platforms that were used included Facebook Ads Manager and Google UAC; in addition, advertisements were placed on Grindr, Snapchat, and Instagram [17,26,31]. In 50% (5/10) of the studies, participants were offered incentives to participate [24-28]. Of these 5 studies, 4 (80%) were conducted in the United States and 1 (20%) in Australia. No study reported that paid marketing resulted in higher app adoption than unpaid campaigns.

Key Indicators of mHealth App Marketing Strategies

Various metrics were used in the reviewed studies to assess mHealth app marketing strategies. Key indicators included nonuse rate, dropout rate, and adherence rate, as well as median time between first and last app use, which was generally measured in days or weeks. Nonuse rate was defined as the proportion of participants who never used the app [24], and the duration of use was assessed by determining the reported number of impressions [13] and prompted patients to respond to invitations to use the app. However, without the motivation provided by a trusted professional, users quickly ceased their use of the app; a dual promotion strategy was then used to increase app adoption: using social media for its reach and ease of response as well as promoting personal invitations from trusted health care professionals [25].

In 1 (10%) of the 10 studies, embedded push notifications were sent to users to keep them engaged with the app [24]. This effect was sustained over time, and push notifications containing tailored health messages reduced the rate at which users disengaged. The timing of notifications was a key factor, with
of days of use [23]. The longest duration of app use was 680 days [23].

Dropout rate was defined as the proportion of participants who completely stopped using the app at least 14 days before they received an invitation to complete the end-of-study survey. Adherence rate was defined as the proportion of participants who used the app at least once every week over 3 months of the study [24].

Specific indicators for paid marketing strategies were cost per click (CPC), defined as the amount of money spent per click secured in each advertisement campaign or advertisement set; CTR, the number of clicks divided by impressions; and cost per impression [26]. Impression is the number of times an advertisement is viewed by a user on an advertising platform. For advertisements seen multiple times by the same user, each view is counted as an impression. Additional indicators were the number and percentage of clicks that took people through to the eligibility screener and of those who were eligible for the study [26].

People Implementing the Marketing Processes
In the studies reviewed, we found that promotional messages were mainly disseminated by researchers, trained nurses, and other health practitioners. Technological support was only provided by the research team and app developers. In some cases, recently enrolled patients contributed to promoting the app by sending download invitations to their family and friends [25].

Marketing Concepts Reported in the Reviewed Studies
Inspired by a book on how to successfully distribute apps [12], we screened the included studies to identify the marketing concepts that were used to promote mHealth apps in these studies (Table 3).
Table 3. Marketing concepts [12] reported in the reviewed studies.

<table>
<thead>
<tr>
<th>Authors, year</th>
<th>Incentivization</th>
<th>Personalization</th>
<th>Mobile app attribution</th>
<th>Loyalty marketing</th>
<th>Remarker-</th>
<th>A/B testing</th>
<th>Programmatic marketing</th>
<th>Predictive marketing</th>
<th>Thought-leadership marketing</th>
<th>Content marketing</th>
<th>Behavioral marketing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kvedarienė et al [23], 2019</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Buss et al [24], 2022</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Arshanapally et al [17], 2022</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Resnick et al [25], 2021</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Zlotorzynska et al [26], 2021</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rajani et al [27], 2021m</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Roberts et al [28], 2019m</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bidargaddi et al [29], 2018</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hui et al [30], 2018m</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Haile et al [31], 2018m</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

*a The incentivized model is the strategy of making a product, program, or other offering more attractive to customers by offering an incentive in exchange for buying or participating. In the app business, incentivization is normally used to quickly amass app installs. Of the 10 studies, half of the studies (n=5, 50%) used incentivization.
b Personalization involves customizing the timing and content of marketing messages to the target user based on their preferences, habits, and behavior patterns. Personalized messages refer to every user by name and entice the user to become more engaged with an app with the right kind of incentive based on their characteristics, such as age, gender, location, profession, and financial segment. Half of the studies (5/10, 50%) used personalization.
c Mobile app attribution is the process of recording and measuring the actions of app users, such as installs, level completions, in-app purchases, and other milestones. Mobile app attribution is essential to app marketing because it helps produce the data that are gathered and analyzed to measure how well marketing campaigns are working. The majority of the studies (7/10, 70%) used mobile app attribution.
d Loyalty marketing is a marketing strategy that focuses on nurturing existing customers rather than acquiring new ones. Only 1 (10%) of the 10 studies used loyalty marketing.
e Remarketing or retargeting targets every individual who has come into contact with the product but has not converted or who converted but later abandoned the app. It allows marketers to reconnect with these categories of users and “bring them back” or increase the time they spend engaging with the app. None of the studies used remarketing or retargeting.
f A/B testing involves the use of several versions of the same advertisement distributed to different groups with different designs, color coding, calls to action, and message content to determine which version produces the highest conversion rate. Of the 10 studies, only 2 (20%) used A/B testing.
g Programmatic marketing is the automated algorithm-based real-time buying and selling of advertising space through a bidding system, with the aim of reaching the right customers at the right time. None of the studies used programmatic marketing.
h Predictive marketing involves using data science based on customer behavior and habits to make smarter marketing decisions. By gathering and analyzing data about user behavior and identifying patterns, marketers can make forecasts about user behavior and make informed decisions about the likelihood of the success of their marketing content and offerings. Of the 10 studies, only 1 (10%) used predictive marketing.
i Thought-leadership marketing is the process of positioning a company as a leader in a specific domain by supplying customers with top-quality information. Only 1 (10%) of the 10 studies used thought-leadership marketing.
Content marketing is a marketing strategy that involves producing content that potential customers find useful, valuable, and relevant. Content marketing is highly effective at building a loyal user base and converting leads into customers. More than one-third of the studies (4/10, 40%) used content marketing.

Behavioral marketing involves segmenting the app’s user base based on user behavior with the aim of refining the marketing strategy and more effectively targeting users. More than one-third of the studies (4/10, 40%) used behavioral marketing.

These studies (n=6, 60%) used single-channel marketing, which involves reaching users through a single channel, eg, Facebook advertisements.

These studies (n=4, 40%) used multichannel marketing, which involves >1 channel, as opposed to an omnichannel marketing campaign, which attempts to reach users through all available channels.

In summary, most of the studies (7/10, 70%) reported using a combination of marketing concepts to advertise their mHealth apps. The most used concept was mobile app attribution (7/10, 70%). Mobile app attribution is essential to app marketing because it helps produce the data that are gathered and analyzed to measure how well marketing campaigns are working [12]. Furthermore, 40% (4/10) of the studies reported using at least 2 channels to market their mHealth apps.

Discussion

Principal Findings

We found that the marketing strategies used in almost all included studies (9/10, 90%) were aiming to drive at least app discovery and onboarding, with more than half (6/10, 60%) also targeting user engagement to mHealth apps. Social media, emails, television or radio, posters or flyers, and face-to-face communications were all used in the reviewed studies to inform people about the existence of mHealth apps, invite them to download, encourage them to use the apps, and maintain engagement. This is consistent with an integrative review of methods used to promote mobile apps, which also cited app store optimization via keywords and the inclusion of screenshots and videos for greater conversion rate, the use of push notifications, the promotion of apps via influencers, and the leveraging of user review and ratings [13].

Social media attracted many downloads over a short period, whereas emails were most often used for sharing instructions on how to download apps and interact with research teams for technical support.

The strategies used to promote mHealth apps included paid and unpaid marketing, and metrics such as CPC and CTR were used to measure effectiveness. Offering incentives to people to download and use the app did increase app downloads and use. However, it has been shown that people are less likely to keep using an app after incentivization, although the monetary value of the incentive could have a significant effect on the adherence [3]. Furthermore, a focused strategy is required to maintain a low app churn rate [12].

Factors Influencing Engagement, Onboarding, or Adherence to mHealth Apps

Reasons for User Engagement, Onboarding, or Adherence to the App

Onboarding was often reliant on the ease of downloading and sufficient motivation. In the study by Hui et al [30], adherence stemmed from awareness that a health care professional and a researcher were interested in the results and that using an app to support self-management could improve a participant’s control of their health condition. Additional factors in engagement included the perceptions of the advantages and disadvantages of using apps to support self-management specific to an individual’s health needs, the relevance of the app to the user, the quality of the app, and the behavior change techniques used to promote health [28]. All these factors are among those reported by a literature review that identified retention factors related to apps, such as feedback, appropriate reminders, and in-app support from peers or coaches [33]. Our findings also align with those of another systematic review that listed individualized reminders, user friendliness and technical stability of the apps, and personal support from health care professionals as intervention-related factors influencing adherence [3].

Push notifications and weekly SMS text messages inviting check-in were also used to influence user engagement [26], herein defined by number of check-ins. This is also consistent with the findings from a systematic review [34] that assessed 15 commercial apps for diabetes prevention and found that the app that included the notification features for activity tips, goal progress tips, goals adjustment, and completed goals had the highest engagement mean score (4.5 points out of 5). However, we argue that push notifications should not be a one-size-fits-all solution because marketing research suggests that among app users, just 50% accept push notifications from their favorite app, and 30% disable all push notifications [35].

In our review, user engagement was also influenced by the health goals that participants selected. This finding also aligns with that of a previous systematic review that further suggested that users could disengage at any time and re-engage at a later stage when needed. Thus, this feature might be particularly useful for addiction research targeting relapse prevention strategies [36].

Barriers to User Engagement, Onboarding, or Adherence to the App

From the studies included in this review, reasons for nonadoption included problems in app installation [24,30]; the use of other health apps that better suited participants’ needs and preferences; and other concerns, such as prioritizing COVID-19 over the condition addressed by the app [28]. Indeed, marketing can be influenced by competing health information targeting the same users. When people were concerned about contracting COVID-19 and seeking a pandemic-related app, they were less motivated to use an app being promoted to tackle other health issues [17,24]. This could lead to competition among mHealth apps for potential users or health care providers.

Other potential barriers to adoption included a lack of early active engagement with participants to verify that they could install the app, no direct contact with health care professionals,
and not enough interactive features. Indeed, as reported in another systematic review, being less informative and less interactive can lead to a very low engagement mean score [34]. The study by Roberts et al [28] reported that technical issues and concerns about data security reduced engagement. Similar concerns about data security with mHealth apps had been previously pointed out in the literature [37,38].

Return on Investment in Marketing

Although dependent upon the budget available, a decision threshold for the cost of marketing campaigns for mHealth apps to reach a certain level of engagement would be beneficial. According to a study on industry-specific Google benchmarks, for the health care sector overall, these costs amount to a CTR of 3.27% and a CPC of US$2.62 for Google Search and a CTR of 0.59% and a CPC of US$0.63 for the Google Display Network [39]. A similar study on industry-specific Facebook benchmarks produced a CTR of 0.83% and a CPC of US$1.32 for health care [40]. However, it is important to note that comparing CTR and CPC with industry-specific benchmarks for the entire health and medical field should not be the sole method of evaluating effectiveness because these benchmarks may be too broad [17].

People Involved in Marketing mHealth Apps

The most productive marketing team is a multitude of satisfied users championing the app on social media, encouraging their friends and colleagues to download it, and giving it 5-star reviews. Creating a strong user support system and feedback loop, regularly updating the app based on user feedback, and doing whatever it takes to keep users happy are the most important marketing tactics that can be deployed [12].

We found that in studies that investigated age and sex differences in engagement with apps, the age of research participants did not predict app engagement [24,25]. However, there were statistically significant differences in sex and app use, with more male participants using the apps in question than female participants, but not in the duration of app use [24,30]. This result is contradictory with that of another study that found that female sex positively influenced adherence [3]. Therefore, we could not draw conclusions on the effect of sex and age on app adherence and thus leave it to further research.

Marketers of mHealth apps should always consider the motivations of the app audience; for example, the study by Roberts et al [28] stated that apps promoting walking can be appealing to survivors of cancer. Similar findings were highlighted elsewhere [3]. In addition, consideration must be given to the timing of interventions intended to maintain engagement. Data should be collected to predict the moments that users will be available and receptive to in-app notifications.

Limitations

We have noted confusion surrounding the term user engagement. Engagement with digital health interventions and engagement with mHealth apps are not clearly separated. This study focuses on the latter, defined as a set of actions by a user within an mHealth app [18]. This differs from user engagement with digital health interventions, which is conceptualized in terms of both experience and behavior [41]. This confusion may explain why many of the studies identified during the database searches had to be excluded: they addressed outcomes related to changes in health behavior; for example, most of the excluded papers reported levels of user engagement with a health intervention, such as increased physical activity, but not engagement with an mHealth app. However, we recognize that these definitions are interlinked: research has shown that app engagement can motivate behavior change [25]. We have noted the same confusion with the term user embedment. In the 10 reviewed studies, only 1 (10%) referred to embedment as integrating a functionality within the app.

Most of the studies we reviewed (9/10, 90%) were conducted in high-income countries, with half being carried out in the United States (5/10, 50%). This may limit generalizability in low- and middle-income countries. Moreover, we only searched for papers written in French and English. We also note that the filters we applied with our search terms combination to avoid noise could have excluded some potentially useful papers. This could explain why the reviewed studies mostly reported research conducted in the United States and Europe (8/10, 80%).

Our findings may also be subject to observer bias [42] because in every reviewed study the research team members were involved in the diffusion processes. In some of the studies (2/10, 20%), participants received in-person physical assistance with app installation. This would be impossible for users in many settings. The generalizability of these findings is also limited by the fact that none of the included studies covered the dissemination of mHealth apps among health care personnel.

Implications and Future Research

Integration of mHealth Apps Into Routine Clinical Practice

None of the studies we reviewed aimed to address the embedment of mHealth apps in routine practice as part of their marketing strategy. To tackle the issue of the embedment of mHealth apps, researchers have proposed a framework for prescribing apps and outlined the key issues that need to be addressed to enable app dissemination in clinical care. This includes education and awareness, the creation of digital formulations, workflow and electronic health record integration, payment models, and patient or provider support [43]. As suggested by this framework, a starting point for the integration of mHealth apps into routine clinical practice would be education and awareness, meaning the promotion of mHealth apps, the aim of which would be to create a base of users downloading the apps because the number of app downloads and interactions over time also provides an indication of sustained uptake over time [11].

At this critical point of creating a user base, inspired by the synthesis of our findings, we offer a set of recommended uses of different channels (Table 4). This would first be applicable during the launching phase of the app and to lead users through their app onboarding stage. The use of these channels could change depending upon the objective of the marketing strategy.
### Table 4. Recommended uses of channels to promote mobile health apps.

<table>
<thead>
<tr>
<th>Channel and recommended use</th>
<th>Targeted people</th>
<th>Supportive marketing concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Email</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inform about the existence of the app, link for download, functionalities, app release notes and app updates, and general information regarding the app (developer, brand owner, and sponsor)</td>
<td>High-level users (influencers and decision-makers) and target base users (health care providers and day-to-day users of the app)</td>
<td>A/B testing</td>
</tr>
<tr>
<td>Give instructions on how to download and install and use the app</td>
<td>High-level users (influencers and decision-makers) and target base users (health care providers and day-to-day users of the app)</td>
<td>A/B testing</td>
</tr>
<tr>
<td>Provide technical support and answers to users’ questions; share user guide and tips</td>
<td>Effective users</td>
<td>A/B testing</td>
</tr>
<tr>
<td><strong>Social media</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inform about the existence of the app, and share the link for download</td>
<td>Potential target base users (health care providers and day-to-day users of the app)</td>
<td>A/B testing, incentivization, loyalty marketing, thought-leadership marketing, and content marketing</td>
</tr>
<tr>
<td>Engage in direct interactions to provide technical support and answers to users’ questions</td>
<td>Potential target base users (health care providers and day-to-day users of the app)</td>
<td>A/B testing, incentivization, loyalty marketing, thought-leadership marketing, and content marketing</td>
</tr>
<tr>
<td><strong>Television</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short promotional video report on the app and its functionalities, as well as 1- to 3-minute video spots with speech by high-level users (influencers and decision makers) recommending the apps</td>
<td>Potential target base users (health care users and day-to-day users of the app)</td>
<td>Incentivization and thought-leadership marketing</td>
</tr>
<tr>
<td><strong>Posters or flyers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infographics and key text message to inform about the existence of the app and its main value; include a QR code and text to indicate link to download</td>
<td>Potential target base users (health care providers and day-to-day users of the app)</td>
<td>Predictive marketing and incentivization</td>
</tr>
<tr>
<td><strong>Face-to-face interaction: in-person training or meeting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inform about the existence of the app, link for download, functionalities, app release notes and app updates, and general information regarding the app (developer, brand owner, and sponsor)</td>
<td>High-level users (influencers and decision-makers) and potential target base users (health care providers and day-to-day users of the app)</td>
<td>A/B testing</td>
</tr>
<tr>
<td>Give instructions on how to download, install, or use the app; share and explain user guide and discuss tips</td>
<td>High-level users (influencers and decision-makers) and potential target base users (health care providers and day-to-day users of the app)</td>
<td>A/B testing</td>
</tr>
<tr>
<td><strong>Face-to-face interaction: in-person ad hoc (unplanned) encounter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engage in direct interactions to provide technical support and answers to users’ questions</td>
<td>Health care providers and day-to-day users of the app</td>
<td>Remarketing or retargeting and personalization</td>
</tr>
<tr>
<td>Sell the app (highlight its main value) and manage to install it on users’ devices; explain how to use it and discuss tips</td>
<td>Health care providers and day-to-day users of the app</td>
<td>Remarketing or retargeting and personalization</td>
</tr>
</tbody>
</table>

Social media could be a beneficial entry point for motivating people to download an app, and human interaction is key during the engagement phase. Therefore, mHealth app promoters should provide users with training and support to start and continue using the apps. This can be done by maintaining communication through social media, including app-dedicated pages. At this point, content marketing—producing content that potential customers find useful—is valuable. It has been demonstrated that content marketing is highly effective in building a loyal user base and converting leads into customers [12].

It is important to note that although social media marketing also tends to attract people who are not the intended audience, communication through email requires a list of targeted email addresses. This entails contacting people directly and requesting their addresses or interacting with someone who will reveal potential users’ email addresses or share an app link with potential users; for instance, a hospital director may share information about an app with hospital staff or share an attendance list containing the email addresses of hospital staff. Future public health campaigns targeting the parents of young children should consider crafting marketing messages for social media campaigns with animated GIF images as well as images of young children.

As time-varying push notifications have been shown to contribute to mHealth app user engagement, developers should...
interact with health care providers to implement this strategy. One approach to this is to apply mobile app attribution: the process of recording and measuring the actions of app users, such as installs, level completions, and in-app purchases [12].

Future Research
Finally, future research could be dedicated to developing a framework on how to disseminate mHealth apps. Such a framework, in addition to various marketing concepts presented in this review, should take into account additional considerations that are specific to mHealth apps, such as data confidentiality and privacy, and segment users on the marketing funnel [12] based on the best available evidence on engaging users with mHealth apps. One issue impeding the dissemination of apps that emerged in our study was the existence of competing apps. Some researchers have suggested that digital formularies or app libraries could help to address this. Digital formularies provide a short list of available apps, and providers could search these formularies and know what is available for a specific diagnosis or purpose [43]. Further research could explore and expand on the effectiveness of digital formularies as a dissemination channel for mHealth apps and the enablers of embedment of mHealth apps into routine practice. Finally, further research could aim to address the gap in identifying specific marketing strategies that would effectively drive the embedment of mHealth apps into routine practice.

Conclusions
The dissemination of mHealth apps takes place via face-to-face interactions, email, and printed posters and social media channels with diverse results. The effects of these strategies are reflected in download numbers and user engagement with mHealth apps. The results of this study will serve to guide future research and guide the marketing of mHealth apps for their routine use within the health sector.

The development of a framework for health care designers to promote their apps within health systems would be immensely beneficial. Such a framework would help systematize the dissemination of mHealth apps and guide the impact assessment of the dissemination strategies.

Acknowledgments
The study received funding from the Universitat Oberta de Catalunya, Barcelona, Spain, for proofreading and article processing fees.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 74 KB - mhealth_v12i1e50293_app1.doc ]

Multimedia Appendix 2
Search strategy.
[DOC File , 57 KB - mhealth_v12i1e50293_app2.doc ]

References
2. Be he@lthy, be mobile: personas toolkit. World Health Organization. 2019. URL: https://apps.who.int/iris/handle/10665/329947 [accessed 2023-04-22]


27. Rajani NB, Mastellos N, Filippidis FT. Impact of gamification on the self-efficacy and motivation to quit of smokers: observational study of two gamified smoking cessation mobile apps. JMIR Serious Games 2021 Apr 27;9(2):e27290 [Full text] [doi: 10.2196/27290] [Medline: 33904824]


37. Tobias G, Spanier AB. Developing a mobile app (iGAM) to promote gingival health by professional monitoring of dental selfies: user-centered design approach. JMIR Mhealth Uhealth 2020 Aug 14;8(8):e19433 [FREE Full text] [doi: 10.2196/19433] [Medline: 32795985]


Abbreviations

CPC: cost per click
CTR: clickthrough rate
GIF: graphic interchange format
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SIGN: Scottish Intercollegiate Guidelines Network
UAC: Universal App Campaigns

©Henri Claude Moungui, Hugues Clotaire Nana-Djeunga, Che Frankline Anyiang, Mireia Cano, Jose Antonio Ruiz Postigo, Carme Carrion. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR
mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Effects of mHealth-Based Lifestyle Interventions on Gestational Diabetes Mellitus in Pregnant Women With Overweight and Obesity: Systematic Review and Meta-Analysis

Yirong He¹²*, MSc; Chuanya Huang¹²*, PhD; Qiuyang He²³*, MSc; Shujuan Liao¹², PhD; Biru Luo¹², PhD

¹Department of Nursing, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China
²Key Laboratory of Birth Defects and Related Diseases of Women and Children, Sichuan University, Ministry of Education, Chengdu, Sichuan, China
³Department of Obstetric Nursing, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China
*these authors contributed equally

Corresponding Author:
Biru Luo, PhD
Department of Nursing
West China Second University Hospital
Sichuan University
#No. 20, Section 3, People’s South Road
Chengdu, Sichuan, 610041
China
Phone: 86 88570307
Email: luomr@scu.edu.cn

Abstract

Background: The increasing incidence of gestational diabetes mellitus (GDM) is a global health problem that is more likely to occur in pregnant women with overweight or obesity. Adhering to a healthy lifestyle is associated with a reduced risk of GDM. With the development of IT, mobile health (mHealth) interventions have become widely available in health care. However, there are no definitive conclusions on the effectiveness of mHealth-based lifestyle interventions in preventing GDM.

Objective: This study aims to evaluate the impact of mHealth-based lifestyle interventions on GDM and other pregnancy outcomes in pregnant women with overweight or obesity.

Methods: A systematic literature search was conducted in 5 English databases (MEDLINE, Embase, Web of Science, CENTRAL, and CINAHL) and 4 Chinese databases (CBM, CNKI, Vip, and Wanfang) to identify randomized controlled trials (RCTs) on the effectiveness of mHealth-based interventions for GDM from inception to January 10, 2023. In total, 2 authors independently screened the studies and extracted the data. The quality of the included studies was examined using the Cochrane risk-of-bias tool. Data synthesis was conducted using Review Manager (version 5.4; The Cochrane Collaboration).

Results: A total of 16 RCTs with 7351 participants were included in this study. The included studies were published between 2014 and 2021 and were conducted in China, the United States, Australia, New Zealand, the United Kingdom, Ireland, and Norway. The sample sizes of the studies ranged from 75 to 2202, and the duration of the mHealth-based lifestyle interventions ranged from 4 to 28 weeks. Compared with usual care, mHealth-based lifestyle interventions significantly reduced the incidence of GDM (odds ratio [OR] 0.74, 95% CI 0.56-0.96; \(P=0.03; \hat{I}^2=65\%\)), preterm birth (OR 0.65, 95% CI 0.48-0.87; \(P=0.004; \hat{I}^2=25\%\)), macrosomia (OR 0.59, 95% CI 0.40-0.87; \(P=0.008; \hat{I}^2=59\%\)), and gestational weight gain (mean difference= −1.12 kg, 95% CI −1.44 to −0.80; \(P<0.001; \hat{I}^2=43\%\)). The subgroup analysis showed that interventions delivered via apps (OR 0.55, 95% CI 0.37-0.83; \(P=0.004; \hat{I}^2=44\%\)), provided by obstetricians (OR 0.69, 95% CI 0.51-0.93; \(P=0.02; \hat{I}^2=60\%\)), and targeted at Asian populations (OR 0.44, 95% CI 0.34-0.58; \(P<0.001; \hat{I}^2=0\%\)) and that used the International Association of Diabetes and Pregnancy Study Groups diagnostic criteria (OR 0.58, 95% CI 0.39-0.86; \(P=0.007; \hat{I}^2=69\%\)) showed a statistically significant reduction in the risk of GDM.

Conclusions: mHealth-based lifestyle interventions had a favorable impact on the prevention of GDM in pregnant women with overweight and obesity. Future studies need to further explore the potential of mHealth-based interventions for GDM through better design and more rigorous large-scale RCTs.
Introduction

Background

Gestational diabetes mellitus (GDM) is defined as a carbohydrate intolerance of varying severity with onset or first recognition during pregnancy [1]. GDM is one of the most common obstetric complications, with the prevalence varying from 7.5% to 27% in different areas [2]. GDM is associated with substantial adverse pregnancy outcomes such as neonatal hypoglycemia and macrosomia as well as long-term metabolic risk in pregnant women and their offspring [3]. Risk factors for GDM include age, ethnicity, overweight or obesity, and family history of diabetes [4]. As obesity becomes a global epidemic, perinatal overweight and obesity are also a growing concern [5]. In recent decades, the rates of overweight and obesity among women of reproductive age have increased dramatically [6]. In the United States, 60% of women are overweight or obese during pregnancy compared with 30% in Europe and 10% in Asia [7]. In pregnant women with overweight and obesity, it is estimated that the risk of GDM is more than twice that of other pregnant women [8]. Unhealthy lifestyle behaviors are a critical factor that affects overweight and obesity during the perinatal period [9]. Numerous randomized controlled trials (RCTs) have attempted to reduce the incidence of GDM through diet [10], physical activity [11], or combination interventions [12]. However, when lifestyle interventions are provided in a personalized way, they are commonly expensive and may lack scalability from the perspective of public health [13]. Therefore, there is an urgent need for effective real-world solutions that address the demand of pregnant women seeking personalized support, information, and guidance to help reduce the risk of GDM.

As IT advances and more people use electronic devices, mobile health (mHealth) has developed rapidly [14]. mHealth is commonly defined as health care services provided by health care professionals using telecommunications technology [15]. mHealth has been applied in many areas of perinatal health care, with medical staff providing health care services to pregnant women through multimedia methods such as mobile apps, software, SMS text messages, email, web-based diaries, and integrated systems combining various components of digital communications technologies [16]. Studies have shown that mHealth care can reduce gestational weight gain (GWG); improve pregnant women’s health behaviors; and reduce the number of medical visits, thereby decreasing financial burden [13]. However, the impact of mHealth interventions on pregnancy outcomes in women with overweight or obesity is uncertain. More and more systematic reviews have found an effect of lifestyle interventions based on mHealth technology on diabetes prevention among adults with overweight and obesity [17,18]; however, little is known about their effectiveness in the perinatal population. Several previous reviews have attempted to synthesize the results of mHealth-based lifestyle programs for pregnant women, but none have evaluated the quantitative effects of these programs [19,20]. There is still no consensus on the impact of mHealth lifestyle interventions on preventing GDM and other pregnancy outcomes in women with overweight or obesity.

Objectives

Therefore, we conducted a systematic review and meta-analysis of RCTs to summarize mHealth interventions delivered in different ways and assess the effectiveness of mHealth-based lifestyle interventions in reducing the risk of GDM.

Methods

This systematic review and meta-analysis follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [21] and is presented in Multimedia Appendix 1. It was registered in PROSPERO on November 18, 2021, with registration number CRD42021286995.

Search Strategy

A systematic literature search was conducted in 5 English databases (MEDLINE, Embase, Web of Science, CENTRAL, and CINAHL) and 4 Chinese databases (CBM, CNKI, Vip, and Wanfang) for studies published in English and Chinese. A systematic search was conducted combining Medical Subject Headings (MeSH) and free-text terms, including overweight (MeSH) OR obesity (MeSH); pregnant* OR prenatal OR antenatal OR maternal OR gestational; gestational diabetes mellitus OR gestational diabetes OR GDM OR T2DM OR impaired fasting glucose OR impaired glucose tolerance; telemedicine (MeSH) OR telerehabilitation (MeSH) OR telecommunications (MeSH) OR electronic health OR eHealth OR eHealthcare OR mHealth OR mobile health OR telecare OR eHealthcare OR mcare OR telemonitor* OR telerehab* OR telemannagement OR mobile communication OR remote consult OR mobile technolog* OR mobile devic* OR internet (MeSH) OR web* OR online OR smartphone (MeSH) OR telephone (MeSH) OR cell phone (MeSH) OR cellular phone (MeSH) OR mobile phone OR messag* OR SMS. The searches were unlimited by time up to January 10, 2023, and were limited to RCTs. The full details of the search strategy for each database are provided in Multimedia Appendix 2. We complemented this strategy by manually searching the reference lists of included studies and related reviews.

Inclusion and Exclusion Criteria

In total, 2 researchers (YH and CH) independently screened the titles and abstracts and selected the studies in accordance with...
the eligibility criteria. Any disagreement was resolved through discussion with a third researcher (QH). Studies were included if they met the following inclusion criteria: (1) pregnant women with overweight (BMI ≥ 25 kg/m²) or obesity (BMI ≥ 30 kg/m²; population); (2) mHealth interventions including pregnancy nutrition, physical activity, weight management, and health behavior education delivered via the internet, websites, telephone, app, SMS text message, email, or other types of information and communications technologies (intervention); (3) usual care, routine care, conventional care, or standard care without mHealth (comparison); (4) incidence of GDM, postpartum hemorrhage, preterm birth, cesarean delivery, pregnancy-induced hypertension, macrosomia, neonatal gestational age, and GWG (outcome); (5) RCTs (study design); and (6) English or Chinese (language). There were no restrictions regarding the year of publication. We excluded studies that (1) included women with either type 1 or type 2 diabetes mellitus before pregnancy or with existing GDM; (2) contained incomplete data; (3) lacked data related to GDM; or (4) were study protocols, comments, editorials, and conference abstracts.

Study Selection and Data Extraction

The reference management program EndNote X9 (Clarivate Analytics) was used for data management. The studies were imported into EndNote after an extensive database search. After removing duplicates, 2 authors independently reviewed the titles and abstracts according to the eligibility criteria. Disagreements were resolved through discussion or consultation with a third researcher. Data were extracted by an independent researcher (YH) using the predesigned data collection forms, and the extracted data were verified by a second researcher (CH). Disagreements were resolved through consensus. The extracted data included the authors, year, country, study design, sample size, participant characteristics, intervention, control, GDM criteria, and outcomes.

Quality Assessment

In total, 2 researchers (YH and CH) independently assessed the studies’ risk of bias in accordance with the Cochrane Handbook for Systematic Reviews of Interventions. This tool consists of 6 items: selection bias, performance bias, detection bias, attrition bias, reporting bias, and other biases. Each item was judged as “low risk,” “high risk,” or “unclear risk.” A third researcher (QH) was available if there was a difference in opinion in assessing the risk of bias.

Data Synthesis and Analysis

Statistical analyses were conducted using Review Manager (version 5.4; The Cochrane Collaboration). The overall effect difference was considered statistically significant if the 2-tailed P value was < .05. Continuous variables were presented using the mean difference, and dichotomous variables were described using the odds ratio (OR) with a 95% CI. Heterogeneity was assessed using the Cochran Q test and the I² statistic. We considered < 25%, 25% to 50%, 50% to 75%, and > 75% as low, moderate, high, and severe heterogeneity between the studies, respectively. If I² ≤ 50% and the P value was > .10, a fixed-effects model was considered; otherwise, a random-effects model was used. The sources of heterogeneity were explored using subgroup analysis. A funnel plot was constructed to check for potential publication bias.

Results

Study Selection

Figure 1 shows a PRISMA flowchart of the study selection process. A total of 1725 records were retrieved from 9 electronic databases. After removing duplicates (393/1725, 22.78% of the studies), 1332 studies were included for screening. Of these 1332 studies, we then excluded 1234 (92.64%) based on the relevance of the abstract and title, and the remaining 98 (7.36%) studies were assessed for eligibility. After full-text review, 84% (82/98) of the studies were excluded for the reasons outlined in Figure 1. Finally, 16 studies were included in the review and meta-analysis.
Characteristics of the Included Studies

Study Characteristics

The characteristics of the included studies are shown in Table 1. The 16 studies included in this review and meta-analysis were RCTs, 4 (25%) of which were multicenter RCTs. The included studies were published between 2014 and 2021 and were conducted in China (8/16, 50%) [22-29], the United States (2/16, 12%) [30,31], Australia (2/16, 12%) [32,33], New Zealand (1/16, 6%) [34], the United Kingdom (1/16, 6%) [35], Ireland (1/16, 6%) [36], and Norway (1/16, 6%) [37]. A total of 7351 participants were included in the studies, and the sample sizes varied from 74 [34] to 2153 [32]. In total, 38% (6/16) of the studies included pregnant women with obesity [23,24,26,28,33,35], and 62% (10/16) included both pregnant women with overweight and pregnant women with obesity [22,25,27,29,30,31,32,34,36,37]. A total of 81% (13/16) of the articles reported diagnostic criteria for GDM. The incidence of GDM was determined by screening pregnant women using an oral glucose tolerance test. The diagnostic criteria for GDM were inconsistent among the included studies. In total, 56% (9/16) of the RCTs used the International Association of Diabetes and Pregnancy Study Groups (IADPSG) criteria (2010) [22,23,25,27-29,34-36], 12% (2/16) used the World Health Organization 2013 criteria [33,37], 6% (1/16) used the Australasian Diabetes in Pregnancy Society criteria [32], and 6% (1/16) of the RCTs used the Carpenter-Coustan criteria [30]. All the RCTs (16/16, 100%) reported more than 1 outcome.
<table>
<thead>
<tr>
<th>Author, year, and country</th>
<th>Study design</th>
<th>Sample size</th>
<th>Participant characteristics</th>
<th>Intervention</th>
<th>Control</th>
<th>GDM criteria</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dodd et al [32], 2014, Australia</td>
<td>Multicenter RCT</td>
<td>2153; IG 1080; CG 1073 (1:1)</td>
<td>Women with a singleton pregnancy between 10 and 20 wk of gestation and a BMI of ≥25 kg/m²</td>
<td>Phone, in person</td>
<td>Standard care</td>
<td>ADPSC</td>
<td>LGA, macrosomia, PIH, pre-eclampsia, GDM, PTB, CS, and PPH</td>
</tr>
<tr>
<td>Sagedal et al [37], 2015, Norway</td>
<td>RCT</td>
<td>591; IG: 296; CG: 295 (1:1)</td>
<td>Women with a singleton pregnancy at ≤20 wk of gestation who had a prepregnancy BMI of ≥25 kg/m²</td>
<td>Phone, website, in person</td>
<td>Standard care</td>
<td>WHO</td>
<td>GWG, GDM, LGA, SGA, pre-eclampsia, PTB, PPH, and NW</td>
</tr>
<tr>
<td>Seneviratne et al [34], 2015, New Zealand</td>
<td>RCT</td>
<td>74; IG: 37; CG: 37 (1:1)</td>
<td>Women aged 18-40 y with a BMI of ≥25 kg/m² and a singleton pregnancy of &lt;20 wk of gestation</td>
<td>Software, device, home-based, in person</td>
<td>Usual care</td>
<td>IADPSG</td>
<td>NW, hypoglycemia, GWG, GDM, PIH, PTB, CS, PPH, MVPA, and pre-eclampsia</td>
</tr>
<tr>
<td>Poston et al [35], 2015, United Kingdom</td>
<td>Multicenter RCT</td>
<td>1280; IG: 629; CG: 651 (1:1)</td>
<td>Women aged ≥16 y with a BMI of ≥30 kg/m² and a singleton pregnancy between 15 and 18 wk plus 6 d of gestation</td>
<td>Phone, email, DVD, pedometer, logbook, in person</td>
<td>Standard care</td>
<td>IADPSG</td>
<td>GDM, FBG, pre-eclampsia, CS, PPH, GWG, HOMA-IR, GL, GI, fat, MVPA, LGA, and NW</td>
</tr>
<tr>
<td>Simmons et al [33], 2017, Australia</td>
<td>Multicenter RCT</td>
<td>192; IG: 92; CG: 100 (1:1)</td>
<td>Women with a BMI of ≥29 kg/m², ≤19 (~6 to , 6) d of gestation, and a singleton pregnancy</td>
<td>Email, phone, pedometer, device, in person</td>
<td>Usual care</td>
<td>WHO</td>
<td>GDM, FBG, HOMA-IR, insulin, NW, LGA, SGA, MVPA, and sugar intake</td>
</tr>
<tr>
<td>Chen [22], 2017, China</td>
<td>RCT</td>
<td>160; IG: 80; CG: 80 (1:1)</td>
<td>Women with a prepregnancy BMI of ≥24 kg/m² and a singleton pregnancy between 8 and 12 wk of gestation</td>
<td>Mobile apps, SMS text messages, in person</td>
<td>Usual care</td>
<td>IADPSG</td>
<td>BMI, GDM, CS, and NW</td>
</tr>
<tr>
<td>Kennelly et al [36], 2018, Ireland</td>
<td>RCT</td>
<td>498; IG: 241; CG: 257 (1:1)</td>
<td>Singleton pregnant women between 10 and 15 wk of gestation with a BMI between 25.0 and 39.9 kg/m²</td>
<td>Mobile app, email, in person</td>
<td>Usual care</td>
<td>IADPSG</td>
<td>GDM, BWG, GL, GI, PIH, CS, HOMA-IR, NW, LGA, and SGA</td>
</tr>
<tr>
<td>Li [23], 2018, China</td>
<td>RCT</td>
<td>1000; IG: 500; CG: 500 (1:1)</td>
<td>Women with a BMI of ≥28 kg/m², aged &gt;18 y, and with a singleton pregnancy between 8 and 12 wk of gestation</td>
<td>Phone, in person</td>
<td>Usual care</td>
<td>IADPSG</td>
<td>GDM, BWG, NW, and macrosomia</td>
</tr>
<tr>
<td>Tang et al [24], 2019, China</td>
<td>RCT</td>
<td>136; IG: 68; CG: 68 (1:1)</td>
<td>Pregnant women with a BMI of ≥28 kg/m² aged &gt;18 y</td>
<td>Mobile app, software</td>
<td>Standard care</td>
<td>N/A</td>
<td>Macrosomia, CS, hypoglycemia, and GDM</td>
</tr>
<tr>
<td>Ferrara et al [30], 2020, United States</td>
<td>Multicenter RCT</td>
<td>389; IG: 195; CG: 194 (1:1)</td>
<td>Women at 8-15 wk of gestation with singletons, with a prepregnancy BMI of 25-40 kg/m², and aged ≥18 y</td>
<td>Phone, device, in person</td>
<td>Usual care</td>
<td>Carpenter and Cousins criteria</td>
<td>GWW, caloric intake, MVPA, FBG, HOMA-IR, GDM, PIH, pre-eclampsia, CS, NW, and macrosomia</td>
</tr>
<tr>
<td>Author, year, and country</td>
<td>Study design</td>
<td>Sample size</td>
<td>Participant characteristics</td>
<td>Intervention</td>
<td>Control</td>
<td>GDM criteria</td>
<td>Outcomes</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>--------------</td>
<td>---------</td>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td>Cao [25], 2020, China</td>
<td>RCT</td>
<td>96; IG: 48; CG: 48 (1:1)</td>
<td>Women aged 22-38 y with a BMI of ≥24 kg/m² and a singleton pregnancy between 8 and 12 wk of gestation</td>
<td>Mobile apps, SMS text messages, in person</td>
<td>Usual care</td>
<td>IADPSG</td>
<td>FBG, GDM, BMI, CS, NW, and LGA</td>
</tr>
<tr>
<td>Wu and Guang [26], 2020, China</td>
<td>RCT</td>
<td>140; IG: 70; CG: 70 (1:1)</td>
<td>Pregnant women with a BMI of ≥28 kg/m² aged &gt;18 y</td>
<td>Mobile app, in person</td>
<td>Usual care</td>
<td>N/A</td>
<td>BMI, NW, GDM, PIH, CS, PTB, macrosomia, and PPH</td>
</tr>
<tr>
<td>Liu et al [31], 2021, United States</td>
<td>RCT</td>
<td>217; IG: 112; CG: 105 (1:1)</td>
<td>Pregnant women aged 18-44 y with a gestational age of ≤16 wk and a prepregnancy BMI of ≥25 kg/m²</td>
<td>Phone, software, SMS text messages, in person</td>
<td>Standard care</td>
<td>N/A</td>
<td>GWG, PTB, LBW v, macrosomia, SGA, GDM, PIH, CS, and NW</td>
</tr>
<tr>
<td>Zhou et al [27], 2021, China</td>
<td>RCT</td>
<td>104; IG: 52; CG: 52 (1:1)</td>
<td>Singleton pregnant women with a BMI of ≥24 kg/m² and aged &gt;18 y</td>
<td>Mobile app, in person</td>
<td>Usual care</td>
<td>IADPSG</td>
<td>FBG, 2-hour BG w, GDM, CS, PTB, and macrosomia</td>
</tr>
<tr>
<td>Kang and Sung [28], 2021, China</td>
<td>RCT</td>
<td>106; IG: 53; CG: 53 (1:1)</td>
<td>Women with a BMI of ≥28 kg/m², aged &gt;18 y, and with a singleton pregnancy between 12 and 20 wk of gestation</td>
<td>Mobile app, in person</td>
<td>Usual care</td>
<td>IADPSG</td>
<td>GDM, FBG, 2-hour BG, HbA1c x, CS, PTB, and macrosomia</td>
</tr>
<tr>
<td>Ding et al [29], 2021, China</td>
<td>RCT</td>
<td>215; IG: 104; CG: 111 (1:1)</td>
<td>Pregnant women with a BMI of ≥24 kg/m² at the onset of pregnancy, aged &lt;35 y, and at &lt;12 wk of gestation</td>
<td>Mobile app, in person</td>
<td>Usual care</td>
<td>IADPSG</td>
<td>Energy intake, GDM, FBG, GWG, CS, PTB, PIH, pre-eclampsia, PPH, NW, and macrosomia</td>
</tr>
</tbody>
</table>

aGDM: gestational diabetes mellitus.
bRCT: randomized controlled trial.
cIG: intervention group.
dCG: control group.
eADPSC: Australasian Diabetes in Pregnancy Society criteria.
fLGA: large-for-gestational-age infant.
PiPH: pregnancy-induced hypertension.
PTB: preterm birth.
CS: cesarean section.
PPH: postpartum hemorrhage.
WHO: World Health Organization.
GWG: gestational weight gain.
SGA: small-for-gestational-age infant.
NW: neonatal weight.
iAIDPSG: International Association of Diabetes and Pregnancy Study Groups.
MVPA: moderate-to-vigorous physical activity.
FBG: fasting blood glucose.
HOMA-IR: homeostatic model assessment of insulin resistance.
GL: glycemic load.
GI: glycemic index.
N/A: not applicable.
LBW: low birth weight.
BG: blood glucose.
HbA1c: glycated hemoglobin.
**Characteristics of the mHealth Interventions**

The details of the mHealth interventions are presented in Multimedia Appendix 3 [22-37]. The interventions were divided into 3 groups: exercise only, diet only, and mixed interventions. Regarding the delivery mode, 12% (2/16) of the studies [23,32] provided mHealth interventions through phone counseling, 25% (4/16) of the studies [26-29] only provided the interventions through mobile apps, and 62% (10/16) of the studies [22,24,25,30,31,33-37] adopted a combination of methods. The duration of the mHealth interventions ranged from 4 to 28 weeks. The mHealth interventions were delivered by various types of personnel, including a single provider in 12% (2/16) of the studies [29,32] and multidisciplinary prenatal care providers in 88% (14/16) of the studies [22-28,30,31,33-37].

**Characteristics of the Comparators**

Most studies (14/16, 88%) briefly described the control group, which generally included regular maternity visits and education on diet and exercise during pregnancy. In the control groups, all studies provided standard or usual care for participants, which was based on the different countries’ perinatal practices and local hospital guidelines.

**Risk of Bias**

The Cochrane risk-of-bias tool was used to assess the risk of bias in each study, and the results are shown in Figure 2 [22-37]. A total of 75% (12/16) of the studies described the details of the randomization scheme, and the risk of random sequence generation was low. Only 31% (5/16) of the studies reported allocation concealment, with a low risk of selection bias. Blinding of participants or staff was difficult because of the nature of the mHealth interventions; thus, 25% (4/16) of the studies, which did not blind the participants or staff, were rated as having a high risk of bias, and 62% (10/16) of the studies were rated as unclear regarding the risk of bias. A total of 12% (2/16) of the studies did not report the outcome assessment and were rated as having a high risk of detection bias. In total, 81% (13/16) of the included studies described details of participant dropout and had a low risk of bias. A total of 56% (9/16) of the studies followed preregistered protocols for analysis and outcome reporting, and the remaining 44% (7/16) did not provide information on published protocols or registrations and had an unclear risk of reporting bias. The funnel plot showed no publication bias in any of the included studies (Multimedia Appendix 4).

![Figure 2. Results of the risk-of-bias assessment of the included studies [22-37].](https://mhealth.jmir.org/2024/1/e49373)
Meta-Analysis Results

Primary Outcome: the Incidence of GDM

The pooled analysis of the 16 RCTs with 7351 participants showed that the mHealth interventions significantly decreased the incidence of GDM in women with overweight and obesity (OR 0.61, 95% CI 0.29-1.28; P=.19, $I^2=61$%), exercise (OR 1.11, 95% CI 0.57-2.17; P=.76, $I^2=0$%), or a combination of the 2 (OR 0.71, 95% CI 0.51-0.99; P=.05; $I^2=74$%) were not statistically significantly associated with reduced risk of GDM. When compared with the control groups, app-based mHealth interventions were significantly effective in reducing GDM (OR 0.55, 95% CI 0.37-0.83; P=.004; $I^2=44$%). Another subgroup analysis conducted based on different providers indicated that interventions provided by obstetricians and nurses were effective in reducing GDM (OR 0.69, 95% CI 0.51-0.93; P=.02; $I^2=60$%). Interventions targeted at Asian populations (OR 0.44, 95% CI 0.34-0.58; P<.001; $I^2=0$%) and using the IADPSG as a diagnostic criterion for GDM showed a reduction in GDM compared with the control groups (OR 0.58, 95% CI 0.39-0.86; P=.007; $I^2=69$%).

Table 2 provides the results of the subgroup analyses, and the forest plots are presented in Multimedia Appendix 5 [22-37]. The subgroup analysis conducted based on the different interventions showed that diet (OR 0.61, 95% CI 0.29-1.28; P=.19; $I^2=61$%), exercise (OR 1.11, 95% CI 0.57-2.17; P=.76; $I^2=0$%), or a combination of the 2 (OR 0.71, 95% CI 0.51-0.99; P=.05; $I^2=74$%) were not statistically significantly associated with reduced risk of GDM. When compared with the control groups, app-based mHealth interventions were significantly effective in reducing GDM (OR 0.55, 95% CI 0.37-0.83; P=.004; $I^2=44$%). Another subgroup analysis conducted based on different providers indicated that interventions provided by obstetricians and nurses were effective in reducing GDM (OR 0.69, 95% CI 0.51-0.93; P=.02; $I^2=60$%). Interventions targeted at Asian populations (OR 0.44, 95% CI 0.34-0.58; P<.001; $I^2=0$%) and using the IADPSG as a diagnostic criterion for GDM showed a reduction in GDM compared with the control groups (OR 0.58, 95% CI 0.39-0.86; P=.007; $I^2=69$%).

Figure 3. Forest plot of the effect of mobile health interventions on gestational diabetes mellitus [22-37]. IV: inverse variance.

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Experimental Events</th>
<th>Control Events</th>
<th>Weight</th>
<th>Odds Ratio</th>
<th>IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cao, 2020</td>
<td>11</td>
<td>15</td>
<td>5.3</td>
<td>0.45</td>
<td>[0.19, 1.10]</td>
</tr>
<tr>
<td>Chem, 2017</td>
<td>10</td>
<td>14</td>
<td>6.8</td>
<td>0.44</td>
<td>[0.22, 0.87]</td>
</tr>
<tr>
<td>Ding et al, 2021</td>
<td>25</td>
<td>31</td>
<td>5.5</td>
<td>0.52</td>
<td>[0.29, 0.94]</td>
</tr>
<tr>
<td>Dodd et al, 2014</td>
<td>148</td>
<td>1000</td>
<td>11.1</td>
<td>1.26</td>
<td>[0.98, 1.63]</td>
</tr>
<tr>
<td>Ferrara et al, 2020</td>
<td>10</td>
<td>19</td>
<td>8.5</td>
<td>0.59</td>
<td>[0.28, 2.09]</td>
</tr>
<tr>
<td>Kang and Sung, 2021</td>
<td>5</td>
<td>53</td>
<td>4.0</td>
<td>0.22</td>
<td>[0.11, 0.48]</td>
</tr>
<tr>
<td>Kennedy et al, 2018</td>
<td>37</td>
<td>241</td>
<td>8.7</td>
<td>1.11</td>
<td>[0.68, 1.83]</td>
</tr>
<tr>
<td>Li et al, 2019</td>
<td>25</td>
<td>62</td>
<td>8.6</td>
<td>0.37</td>
<td>[0.23, 0.60]</td>
</tr>
<tr>
<td>Liu et al, 2021</td>
<td>8</td>
<td>112</td>
<td>5.0</td>
<td>0.54</td>
<td>[0.22, 1.37]</td>
</tr>
<tr>
<td>Pouton et al, 2015</td>
<td>100</td>
<td>172</td>
<td>11.1</td>
<td>0.95</td>
<td>[0.74, 1.22]</td>
</tr>
<tr>
<td>Saeid et al, 2017</td>
<td>32</td>
<td>206</td>
<td>8.1</td>
<td>1.31</td>
<td>[0.76, 2.27]</td>
</tr>
<tr>
<td>Sennermatt et al, 2015</td>
<td>4</td>
<td>37</td>
<td>2.0</td>
<td>2.12</td>
<td>[0.36, 12.36]</td>
</tr>
<tr>
<td>Simmons et al, 2017</td>
<td>18</td>
<td>92</td>
<td>6.6</td>
<td>1.04</td>
<td>[0.51, 2.13]</td>
</tr>
<tr>
<td>Tang et al, 2019</td>
<td>18</td>
<td>68</td>
<td>6.1</td>
<td>0.79</td>
<td>[0.37, 1.72]</td>
</tr>
<tr>
<td>Xu and Guan, 2020</td>
<td>0</td>
<td>70</td>
<td>0.0</td>
<td>0.14</td>
<td>[0.01, 2.70]</td>
</tr>
<tr>
<td>Zhou et al, 2021</td>
<td>1</td>
<td>62</td>
<td>1.4</td>
<td>0.11</td>
<td>[0.01, 0.90]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>3657</td>
<td>3694</td>
<td>100.0</td>
<td>0.74</td>
<td>[0.56, 0.96]</td>
</tr>
</tbody>
</table>

Total events: 524
Heterogeneity: Tau² = 0.16; Chi² = 43.00, df = 15 (P = 0.0002); I² = 65%
Test for overall effect: Z = 2.22 (P = 0.03)
### Table 2. Subgroup analyses of the included studies (n=16).

<table>
<thead>
<tr>
<th>Subgroup analysis</th>
<th>References</th>
<th>Studies, n (%)</th>
<th>Sample size</th>
<th>Effect estimates (95% CI)</th>
<th>P value</th>
<th>$I^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Diet              | • Tang et al [24]  
• Zhou et al [27]  
• Kang and Sung [28]  
• Kennelly et al [36] | 4 (25) | 844 | 0.61 (0.29-1.28) | .19 | 61 |
| Exercise          | • Ferrara et al [30]  
• Seneviratne et al [34] | 2 (12) | 463 | 1.11 (0.57-2.17) | .76 | 0 |
| Both              | • Chen [22]  
• Li [23]  
• Cao [25]  
• Wu and Guang [26]  
• Ding et al [29]  
• Liu et al [31]  
• Dodd et al [32]  
• Simmons et al [33]  
• Poston et al [35]  
• Sagedal et al [37] | 10 (62) | 6044 | 0.71 (0.51-0.99) | .05 | 74 |
| **mHealth® technology** |            |                |             |                           |         |           |
| Phone             | • Ferrara et al [30]  
• Simmons et al [33]  
• Poston et al [35]  
• Sagedal et al [37]  
• Liu et al [31]  
• Dodd et al [32] | 6 (38) | 4822 | 1.08 (0.92-1.26) | .37 | 1 |
| App               | • Chen [22]  
• Tang et al [24]  
• Cao [25]  
• Wu and Guang [26]  
• Zhou et al [27]  
• Kang and Sung [28]  
• Ding et al [29]  
• Kennelly et al [36] | 8 (50) | 1455 | 0.55 (0.37-0.83) | .004 | 44 |
| Computer          | • Li [23]  
• Seneviratne et al [34] | 2 (12) | 1074 | 0.72 (0.14-3.74) | .69 | 71 |
| **Provider**      |            |                |             |                           |         |           |
| Dietitian         | • Ding et al [29]  
• Dodd et al [32] | 2 (12) | 2368 | 0.84 (0.36-2.00) | .70 | 86 |
| Exercise physiologists | • Seneviratne et al [34] | 1 (6) | 74 | 2.12 (0.36-12.36) | .40 | N/A |
| Obstetricians     | • Chen [22]  
• Li [23]  
• Tang et al [24]  
• Cao [25]  
• Wu and Guang [26]  
• Zhou et al [27]  
• Kang and Sung [28]  
• Ferrara et al [30]  
• Liu et al [31]  
• Simmons et al [33]  
• Poston et al [35]  
• Kennelly et al [36]  
• Sagedal et al [37] | 13 (81) | 4909 | 0.69 (0.51-0.93) | .02 | 60 |
| **Duration**      |            |                |             |                           |         |           |

---

https://mhealth.jmir.org/2024/1/e49373
<table>
<thead>
<tr>
<th>Subgroup analysis</th>
<th>References</th>
<th>Studies, n (%)</th>
<th>Sample size</th>
<th>Effect estimates (95% CI)</th>
<th>P value</th>
<th>I² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short term</td>
<td>Zhou et al [27]</td>
<td>3 (19)</td>
<td>1601</td>
<td>0.60 (0.26-1.39)</td>
<td>.24</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Liu et al [31]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poston et al [35]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium to long term</td>
<td>Chen [22]</td>
<td>11 (69)</td>
<td>5518</td>
<td>0.77 (0.53-1.10)</td>
<td>.15</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Li [23]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wu and Guang [26]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kang and Sung [28]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ding et al [29]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ferrara et al [30]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dodd et al [32]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simmons et al [33]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seneviratne et al [34]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kennelly et al [36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sagedal et al [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long term</td>
<td>Tang et al [24]</td>
<td>2 (12)</td>
<td>232</td>
<td>0.62 (0.35-1.12)</td>
<td>.11</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Cao [25]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>Chen [22]</td>
<td>8 (50)</td>
<td>1957</td>
<td>0.44 (0.34-0.58)</td>
<td>&lt;.001</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Li [23]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tang et al [24]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cao [25]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wu and Guang [26]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zhou et al [27]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kang and Sung [28]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ding et al [29]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Ferrara et al [30]</td>
<td>8 (50)</td>
<td>5394</td>
<td>1.09 (0.93-1.26)</td>
<td>.29</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Liu et al [31]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dodd et al [32]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simmons et al [33]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seneviratne et al [34]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poston et al [35]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kennelly et al [36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sagedal et al [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**GDM² diagnostic criteria**
### Secondary Outcomes: Maternal and Neonatal Outcomes

#### Effect on GWG

A total of 50% (8/16) of the studies examined the effects of mHealth-based interventions on GWG [23,29-31,34-37]. There were statistically significant differences in decreases in GWG between the mHealth intervention groups and the control groups (mean difference = -1.12 kg, 95% CI -1.44 to -0.80; \( P < .001; I^2 = 43\%\); Figure 4 [23, 29-31,34-37]). The effects of the mHealth-based lifestyle interventions on maternal and neonatal outcomes are shown in Table 3.

#### Figure 4. Forest plot of the effect of mobile health interventions on gestational weight gain [23, 29-31,34-37]. IV: inverse variance.
Table 3. Effectiveness of mobile health–based lifestyle interventions on maternal and neonatal outcomes.

<table>
<thead>
<tr>
<th>Maternal and neonatal outcomes</th>
<th>References</th>
<th>Studies, n (%)</th>
<th>Sample size</th>
<th>Statistical method</th>
<th>Effect estimates (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maternal outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postpartum hemorrhage (%)</td>
<td>Wu and Guang [26] • Ding et al [29] • Dodd et al [32] • Seneviratne et al [34] • Poston et al [35] • Sagedal et al [37]</td>
<td>6 (38)</td>
<td>4664</td>
<td>Odds ratio (IV; fixed)</td>
<td>1.06 (0.90 to 1.24)</td>
<td>.49</td>
</tr>
<tr>
<td>Cesarean delivery (%)</td>
<td>Chen [22] • Tang et al [24] • Cao [25] • Wu and Guang [26] • Ferrara et al [30] • Dodd et al [32] • Seneviratne et al [34] • Poston et al [35] • Kennelly et al [36]</td>
<td>9 (56)</td>
<td>5224</td>
<td>Odds ratio (IV; fixed)</td>
<td>0.89 (0.79 to 1.00)</td>
<td>.05</td>
</tr>
<tr>
<td>Preeclampsia or PIH (%)</td>
<td>Wu and Guang [26] • Ding et al [29] • Ferrara et al [30] • Liu et al [31] • Dodd et al [32] • Seneviratne et al [34] • Poston et al [35] • Kennelly et al [36] • Sagedal et al [37]</td>
<td>9 (56)</td>
<td>5829</td>
<td>Odds ratio (IV; fixed)</td>
<td>0.96 (0.80 to 1.15)</td>
<td>.63</td>
</tr>
<tr>
<td>GWG (kg)</td>
<td>Li [23] • Ding et al [29] • Ferrara et al [30] • Liu et al [31] • Seneviratne et al [34] • Poston et al [35] • Kennelly et al [36] • Sagedal et al [37]</td>
<td>8 (50)</td>
<td>4133</td>
<td>Mean difference (IV; fixed)</td>
<td>-1.12 (-1.44 to -0.80)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Neonatal outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preterm birth (%)</td>
<td>Wu and Guang [26] • Zhou et al [27] • Kang and Sung [28] • Ding et al [29] • Liu et al [31] • Dodd et al [32] • Seneviratne et al [34]</td>
<td>7 (44)</td>
<td>2998</td>
<td>Odds ratio (IV; fixed)</td>
<td>0.65 (0.48 to 0.87)</td>
<td>.004</td>
</tr>
<tr>
<td>Macrosomia (%)</td>
<td>Li [23] • Tang et al [24] • Wu and Guang [26] • Zhou et al [27] • Kang and Sung [28] • Ding et al [29] • Ferrara et al [30] • Liu et al [31] • Dodd et al [32] • Seneviratne et al [34]</td>
<td>9 (56)</td>
<td>4449</td>
<td>Odds ratio (IV; random)</td>
<td>0.59 (0.40 to 0.87)</td>
<td>.008</td>
</tr>
<tr>
<td>LGA &gt;90th percentile (%)</td>
<td>Dodd et al [32] • Simmons et al [33] • Poston et al [35] • Kennelly et al [36] • Sagedal et al [37]</td>
<td>5 (31)</td>
<td>4966</td>
<td>Odds ratio (IV; random)</td>
<td>0.80 (0.60 to 1.06)</td>
<td>.12</td>
</tr>
</tbody>
</table>
### Maternal and neonatal outcomes

<table>
<thead>
<tr>
<th>SGA&lt;sup&gt;e&lt;/sup&gt; &lt;10th percentile (%)</th>
<th>Liu et al [31]</th>
<th>Simmons et al [33]</th>
<th>Kennelly et al [36]</th>
<th>Sagedal et al [37]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies, n (%)</td>
<td>4 (25)</td>
<td>1529</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical method</td>
<td>Odds ratio (IV; fixed)</td>
<td>1.10 (0.78 to 1.55)</td>
<td>.60</td>
<td></td>
</tr>
</tbody>
</table>

*IV: inverse variance.

**Effect on Cesarean Delivery**

A total of 56% (9/16) of the studies explored the effects of mHealth-based interventions on cesarean delivery [22,24-26,30,32-36]. The pooled results showed no significant difference between the mHealth intervention and control groups (OR 0.89, 95% CI 0.79-1.00; \( P = .05; I^2 = 39\% \)).

**Forest plots of maternal and neonatal outcomes are shown in Multimedia Appendix 6 [22-37].**

**Effect on Pregnancy-Induced Hypertension**

A total of 56% (9/16) of the studies assessed the effects of mHealth-based interventions on the risk of pregnancy-induced hypertension [26,29-32,34-37]. Compared with the control groups, the groups with mHealth-based interventions did not show statistically significant decreases in pregnancy-induced hypertension (OR 0.96, 95% CI 0.80-1.15; \( P = .63; I^2 = 31\% \)).

**Effect on Postpartum Hemorrhage**

A total of 38% (6/16) of the studies assessed the effects of mHealth-based lifestyle interventions on the risk of postpartum hemorrhage [26,29,31,32,34-37]. There were no statistically significant decreases in postpartum hemorrhage between the groups with mHealth-based interventions and the control groups (OR 1.06, 95% CI 0.90-1.24; \( P = .49; I^2 = 17\% \)).

**Effect on Preterm Birth**

A total of 44% (7/16) of the studies explored the effects of mHealth-based interventions on the risk of preterm birth [26-29,31,32,34]. Statistically significant differences in decreases in preterm birth were found between the mHealth intervention groups and the control groups (OR 0.65, 95% CI 0.48-0.87; \( P = .004; I^2 = 25\%; Figure 5 [26-29,31,32,34] \)).

**Effect on Macrosomia**

A total of 56% (9/16) of the studies evaluated the effects of mHealth-based interventions on the risk of macrosomia [23,24,26-32]. Statistically significant differences in decreases in macrosomia were found between the mHealth intervention groups and the control groups (OR 0.59, 95% CI 0.40-0.87; \( P = .008; I^2 = 59\%; Figure 6 [23,24,26-32] \)).
### Effect on Large-for-Gestational-Age Infants

A total of 31% (5/16) of the studies explored the effects of mHealth-based lifestyle interventions on large-for-gestational-age infants [32,33,35-37]. Compared with the control groups, the mHealth intervention groups did not show statistically significant decreases in large-for-gestational-age infants (OR 0.80, 95% CI 0.60-1.06; P = .12; I² = 51%).

### Effect on Small-for-Gestational-Age Infants

In total, 25% (4/16) of the studies assessed the effects of mHealth-based lifestyle interventions on small-for-gestational-age infants [31,33,36,37]. Compared with the control groups, the mHealth intervention groups did not show statistically significant decreases in small-for-gestational-age infants (OR 1.10, 95% CI 0.78-1.55; P = .60; I² = 0%).

### Discussion

#### Methodological Quality of the Included Studies

The methodological quality of the included studies was assessed using the Cochrane risk-of-bias tool. A total of 25% (4/16) of the studies did not report details of randomization methods, and 75% (12/16) of the studies did not report allocation concealment and were at risk of selection bias. In total, 25% (4/16) of the studies did not blind participants and researchers. Therefore, there was a high risk of performance bias in these studies. A total of 12% (2/16) of the studies did not blind the outcome assessors and were rated as having a high risk of detection bias. The included studies reported complete data with low attrition bias. Finally, the funnel plot showed no substantial publication bias among the included studies. Therefore, the overall methodological quality was moderate, and larger samples and well-designed randomized trials are needed in the future.

### Summary of Principal Findings

#### Effectiveness of mHealth Interventions

A total of 16 RCTs were included in this systematic review and meta-analysis. All of these studies offered mHealth lifestyle interventions for perinatal women with overweight or obesity. Pooled results showed that the mHealth interventions reduced the incidence of GDM, which was consistent with the results of a previous study [38]. Various strategies have been proposed to prevent GDM effectively, and the primary intervention strategy is to change the lifestyle. Lifestyle interventions, including dietary guidance, physical exercise, weight management, health education, and blood glucose self-monitoring, are first-line strategies for GDM prevention [39]. The benefits of lifestyle interventions are mediated by mechanisms that improve glycemic variables and outcomes in type 4 and type 2 diabetes by increasing insulin sensitivity and reducing oxidative stress, which has been demonstrated in studies in other populations [40]. Compared with conventional lifestyle interventions, mHealth-based lifestyle interventions can make health education more attractive by providing more intuitive and vivid education and consultation with the help of electronic devices. In addition, pregnant women need to maintain close and continuous contact with the medical team during pregnancy and communicate effectively in the event of pregnancy complications and conscious fetal abnormalities. The advantage of mHealth technology is its ability to provide time-sensitive connectivity and high-quality health care for pregnant women in all regions [41]. mHealth lifestyle interventions enable pregnant women to maintain a healthy diet and engage in appropriate physical activity on a daily basis to encourage a healthier lifestyle during pregnancy, thus reducing the incidence of GDM.

#### Subgroup Analysis

In this review, we classified mHealth lifestyle interventions into 3 groups based on delivery approach: app-based interventions, phone-based interventions, and mHealth interventions delivered via a computer. Subgroup analysis demonstrated that app-delivered mHealth interventions were highly effective in reducing the risk of GDM in women with overweight or obesity. mHealth telephone-based interventions and interventions delivered via a computer had no significant effect on the prevention of GDM. Perinatal women usually have different problems and specific needs during pregnancy. Personalized applications can focus on individual characteristics and tailor goals and actions to diverse populations [42]. In this way, medical staff can leverage app-based mHealth interventions to provide health care tailored to the specific needs of pregnant women with overweight or obesity. In contrast, the applications help medical staff remotely monitor real-time parameters related...
to the health of pregnant women who are overweight or obese during pregnancy, observe whether they are adhering to a healthy and appropriate lifestyle, facilitate communication with them, and help them control their blood sugar and reduce the incidence of GDM [43].

Regarding the different providers of the interventions, the results of the subgroup analysis reported that mHealth interventions delivered by medical staff, including obstetricians and nurses, were effective in preventing GDM. It has been suggested that different models of care provided by various intervention providers may influence outcomes. mHealth interventions delivered by dietitians may focus on the food needs of women during pregnancy [44], whereas exercise physiologists are more concerned with physical activity during pregnancy [45]. Medical staff usually focus on the overall health of pregnant women, providing nutrition, exercise, weight management, and other comprehensive knowledge of pregnancy health care [46].

Subgroup analysis showed that mHealth interventions combining diet and exercise were effective in preventing gestational diabetes in women with overweight or obesity. Combined interventions with diet and exercise appeared to have a greater impact on GDM than interventions with diet or exercise alone, and the effects of diet and exercise on GDM were indistinguishable [47]. Type 2 diabetes has been proven to be preventable through combined diet, exercise, and weight loss interventions [48]. Another subgroup analysis found that mHealth-based lifestyle interventions were effective in preventing GDM in Asian populations. However, the pooled effects of interventions on the risk of GDM in predominantly White populations did not reach statistical significance. The disease burden of GDM varies by race because of multiple factors, including socioeconomic status, lifestyle, and culture, and the prevalence of GDM is significantly higher among Asian and Hispanic populations than among White populations [49].

The results of the subgroup analysis showed that, in studies using the IADPSG diagnostic criteria for GDM, mHealth-based lifestyle interventions had a significantly greater effect on GDM. The prevalence of GDM varies widely worldwide, at least in part because of a lack of consistency in screening and diagnostic criteria. Using lower glucose level thresholds as recommended by the IADPSG, significantly higher numbers of women with GDM were identified compared with using other diagnostic criteria [50].

Maternal and Neonatal Outcomes

Our meta-analyses revealed that the combined effect of mHealth interventions reduced the incidence of preterm birth, macrosomia, and excessive GWG. A review summarized the evidence regarding the influence of maternal diet before and during pregnancy on preterm birth. The results indicated that better maternal diet quality during pregnancy, characterized by a high intake of vegetables, fruits, whole grains, dairy products, and protein, played a significant role in reducing the risk of premature birth [51]. In addition, this study was in line with the findings of Fair and Soltani [52], who conducted a systematic review of the effectiveness of lifestyle interventions on weight gain in women with overweight or obesity during pregnancy and found that lifestyle interventions slightly reduced weight gain during pregnancy by 0.3 to 2.4 kg compared with standard care. There is no robust evidence that mHealth-based interventions are associated with a lower prevalence of postpartum hemorrhage, pregnancy-induced hypertension, cesarean sections, or any alteration in gestational age, consistent with a previous systematic review [53]. One reason for this nonsignificant effect may be the insufficient power of the combined sample size. Another possibility is the short duration of the interventions (median 18 wk), which might have been inadequate to affect some obstetric complications and neonatal outcomes. Further studies are needed to actively explore the associations between mHealth interventions and maternal and neonatal outcomes.

Strengths and Limitations

This systematic review and meta-analysis has several strengths. We combined MeSH terms and keywords covering pregnancy and mHealth to conduct a comprehensive search in 5 primary English electronic databases and 4 main Chinese electronic databases to minimize the possibility of publication bias. We used a robust 3-step search strategy to include databases containing published and unpublished RCTs. To minimize bias, the review methods were preregistered in accordance with the PRISMA statement. In addition, some of the studies included in this meta-analysis were conducted in Western, high-income countries, and some were conducted in Eastern, lower-income countries. Our meta-analyses provided an excellent synthesis of the responses of participants from different cultural backgrounds to mHealth interventions.

However, there were also some limitations to this review. First, we only retrieved studies published in English or Chinese owing to language limitations. Studies published in other languages were not included in the review, which may have led to some publication bias. Second, the methodological quality of the included studies was not optimal, and some studies had a risk of performance bias and detection bias. Third, the included studies varied in sample size, participant characteristics, components of the mHealth interventions, and intervention implementation methods, which may have led to high heterogeneity. Moreover, key variables such as the exact start of the intervention and intensity of physical activity were missing and incomplete in some studies and may have biased the pooled effects. Finally, not all the included studies reported the safety and cost-effectiveness of the mHealth interventions.

Conclusions

This systematic review and meta-analysis demonstrated that mHealth-based lifestyle interventions have a favorable impact on the prevention of GDM in pregnant women with overweight and obesity. mHealth interventions are a convenient and effective way to support pregnant women with overweight and obesity in out-of-hospital self-management in the context of rapid advances in IT and faster transmission speed. However, the potential of mHealth-based interventions for GDM needs to be further explored with better design and more rigorous large-scale RCTs.
Authors’ Contributions
All authors made substantial contributions to the study. YH initiated the study. CH and QH conducted the data extraction and analyses. YH wrote the first draft of the manuscript. BL critically reviewed and revised the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 28 KB - mhealth_v12i1e49373_app1.docx ]

Multimedia Appendix 2
Search strategies.
[DOCX File , 22 KB - mhealth_v12i1e49373_app2.docx ]

Multimedia Appendix 3
The details of the mobile health interventions.
[DOCX File , 23 KB - mhealth_v12i1e49373_app3.docx ]

Multimedia Appendix 4
Funnel plot.
[DOCX File , 29 KB - mhealth_v12i1e49373_app4.docx ]

Multimedia Appendix 5
Forest plots of subgroup analysis.
[DOCX File , 172 KB - mhealth_v12i1e49373_app5.docx ]

Multimedia Appendix 6
Forest plots of secondary outcomes.
[DOCX File , 99 KB - mhealth_v12i1e49373_app6.docx ]

References


15. mHealth: new horizons for health through mobile technologies. World Health Organization. 2011. URL: https://www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technology [accessed 2023-12-10]


32. Dodd JM, Turnbull D, McPhee AJ, Deussens AR, Grivell RM, Yelland LN, LIMIT Randomised Trial Group. Antenatal lifestyle advice for women who are overweight or obese: LIMIT randomised trial. BJM 2014 Feb 10;348:g1285 [FREE Full text] [doi: 10.1136/bmj.g1285] [Medline: 24513442]


Abbreviations

GDM: gestational diabetes mellitus
GWG: gestational weight gain
IADPSG: International Association of Diabetes and Pregnancy Study Groups
MeSH: Medical Subject Headings
mHealth: mobile health
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

©Yirong He, Chuanya Huang, Qiuyang He, Shujuan Liao, Biru Luo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Effectiveness of Telecare Interventions on Depression Symptoms Among Older Adults: Systematic Review and Meta-Analysis

Man Wu¹, BSN; Chaoyang Li¹, BSN; Ting Hu¹, BSN; Xueyang Zhao¹, BSN; Guiyuan Qiao¹, BSN; Xiaolian Gao¹, BSN; Xinhong Zhu¹, PhD; Fen Yang¹,², PhD

¹School of Nursing, Hubei University of Chinese Medicine, Wuhan, China
²Hubei Shizhen Laboratory, Wuhan, China

Corresponding Author:
Fen Yang, PhD
School of Nursing
Hubei University of Chinese Medicine
North District, No 16, Huangjiahu West Road
Hongshan District
Wuhan, 430065
China
Phone: 86 18062003261
Email: fenyang@hbtcm.edu.cn

Abstract

Background: Depression is the most common psychiatric disorder among older adults. Despite the effectiveness of pharmacological and psychological therapies, many patients with late-life depression (LLD) are unable to access timely treatment. Telecare has been shown to be effective in addressing patients' psychosocial issues, while its effectiveness in serving patients with LLD remains unclear.

Objective: This study aimed to evaluate the effectiveness of telecare in reducing depression and anxiety symptoms and improving quality of life (QoL) in patients with LLD.

Methods: Databases including the Cochrane Library, Web of Science, PubMed, Embase, and EBSCO were searched for randomized controlled trials (RCTs) evaluating the effectiveness of telecare for LLD from database establishment to December 28, 2022.

Results: A total of 12 RCTs involving 1663 participants were identified in this study. The meta-analysis showed that (1) telecare significantly reduced depressive symptoms in patients with LLD compared to those in usual care (UC; standardized mean difference [SMD]=−0.46, 95% CI −0.53 to −0.38; P<.001), with the best improvement observed within 3 months of intervention (SMD=−0.72, 95% CI −1.16 to −0.28; P<.001); (2) other scales appeared more effective than the Patient Health Questionnaire-9 for LLD in telecare interventions (SMD=−0.65, 95% CI −0.96 to −0.35; P<.001); (3) telecare was more effective than telephone-based interventions for remote monitoring of LLD (SMD=−1.13, 95% CI −1.51 to −0.76; P<.001); (4) the reduction of depressive symptoms was more pronounced in patients with LLD with chronic conditions (SMD=−0.67, 95% CI −0.89 to −0.44; P<.001); (5) telecare was more effective for LLD in Europe and the Americas than in other regions (SMD=−0.73, 95% CI −0.99 to −0.47; P<.001); (6) telecare significantly reduced anxiety symptoms in patients with LLD (SMD=−0.53, 95% CI −0.73 to −0.33; P=.02); and (7) there was no significant improvement in the psychological components of QoL in patients with LLD compared to those receiving UC (SMD=0.30, 95% CI 0.18–0.43; P=.80).

Conclusions: Telecare is a promising modality of care for treatment, which can alleviate depression and anxiety symptoms in patients with LLD. Continued in-depth research into the effectiveness of telecare in treating depression could better identify where older patients would benefit from this intervention.

(JMIR Mhealth Uhealth 2024;12:e50787) doi:10.2196/50787

KEYWORDS
telecare; depression; anxiety; quality of life; older adults; meta-analysis
Introduction

Statistics show that the world’s population older than 60 years will double between 2015 and 2050, increasing from 12.0% to 22.0% [1]. With the rapid growth of the older population, late-life depression (LLD) has gradually emerged as a hot topic in the field of geriatric medical research. LLD refers to depressive disorders occurring in adults older than 60 years [2,3]. Research findings indicate a global prevalence of LLD of 28.4% [4], which could potentially be higher among individuals with concurrent physical ailments. As a geriatric syndrome with multifactorial etiology, LLD is highly associated with medical problems that pervade later life, including diabetes, hypertension, and dementia [2,5]. LLD is often chronic or recurrent and is associated with functional impairment, diminished health-related quality of life (QoL), and impaired social-psychological functioning [3,6]. A study confirmed that health care costs for patients with LLD were 43.0% to 52.0% higher for outpatient services and 47.0% to 51.0% higher when outpatient and inpatient services were combined, compared to those for individuals without LLD [7].

Despite its high prevalence and severe adverse outcomes, LLD is often overlooked and inadequately treated due to other complications resulting from aging-related issues. Psychopharmacotherapy and psychotherapy have been demonstrated to be effective for people with depression [6]; however, these treatments still have limitations, such as medical side effects and poor treatment adherence [8,9]. Due to mobility issues, geographic isolation, stigma associated with mental illness, and negative beliefs about treatment, older adults have limited access to health care or may be unwilling to seek help from health care institutions [10-12]. Additionally, underuse of professional mental health services, including low detection rates by health care providers and the lack of awareness among older patients regarding the severity of their condition [13,14], is also one of the factors that impede the treatment of LLD. Limited by these factors, only a minority of older adults receive appropriate treatment for depression. Therefore, there is an urgent need to study the clinical effectiveness of alternative therapies for depression, which are more socially acceptable and easily available.

In recent years, there has been increasing attention toward using telecare to support the management and well-being of mental health [15]. Telecare refers to the delivery of health care directly to users, typically in their own homes, supported by information and communication technologies such as telephone, videoconferencing, and applications [16,17]. Health care professionals can remotely provide consultation, assessment, and intervention services to patients [18]. These services include, but are not limited to, lifestyle monitoring, remote monitoring of vital signs for diagnosis, as well as long-distance assessment and education. The benefits of telecare are evident. Evidence suggests that as a promising strategy, telecare services can serve as a medium to overcome certain barriers, thereby enhancing mental health care and increasing opportunities to access evidence-based care under different conditions [19]. Particularly, telecare benefits older adults who are socially isolated or physically frail due to illness, disability, or other familial roles [17,20]. Currently, telecare has been widely used in the management of various chronic conditions among older adults, such as diabetes, hypertension, Parkinson disease, etc, yielding positive outcomes [21-23]. Depression is a commonly observed chronic condition among older adults, closely associated with an approximate 50% increase in chronic disease-related health care costs [24]. Given the significant impact of LLD on patients’ QoL and its potential consequences on decreased productivity or suicide, ensuring continuity of care is imperative. Telecare has been proposed as an effective alternative to help bridge this treatment problem. Considering the complexity and severity of LLD, it is necessary to further explore whether telecare is effective in improving health outcomes for patients with LLD.

Previous reviews have assessed the evidence related to the use of telecare for managing mental health issues [11,25]. In the field of psychiatry, telecare has been found to significantly impact mental health outcomes in older adults, including reducing emergency visits and hospitalizations, as well as improving cognitive function [11]. However, the efficacy of telecare for depression is inconsistent. Some studies suggest the effectiveness of telecare in reducing symptoms of depression [11,26], while others indicate that the impact of telecare on improving depressive symptoms is limited, even yielding contradictory results [27,28]. Previous meta-analyses examining the effectiveness of telecare on depression have mostly focused on adult populations [25-27]. However, compared to other age groups, LLD is considered to be different [14]. Differences in study design, intervention methods, and treatment intensity may contribute to varying clinical outcomes in telecare treatments for LLD. Despite recent meta-analyses demonstrating significant efficacy of telemedicine in alleviating depressive symptoms among older adults, the evaluation of its evidence remains limited [29]. Due to inherent heterogeneity in inclusion criteria, interpretation of these results should be approached cautiously. The severe clinical outcomes and interfering factors often pose significant challenges in the treatment of LLD. Determining whether telecare management is effective for LLD is critical. It is unclear how effective telecare is in improving depression, anxiety symptoms, and QoL in patients with LLD. Therefore, this systematic review and meta-analysis explored the efficacy of telecare for LLD.

Methods

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [30].

Search Strategy

We conducted searches in Cochrane Library, Web of Science, PubMed, Embase, and EBSCO for randomized controlled trials (RCTs) published from the inception of the databases up to December 28, 2022, without any language restriction. MeSH (Medical Subject Headings) and free search terms were both used in the literature search. The search terms included “cell Phones,” “telemedicine,” “smartphone,” “mobile applications,” “mobile phone*,” “telephone*,” “telehealth,” “tele-healthy,” “electronic health*,” “application*,” “m-health,” “messaging,” “depression,” “depressive disorder,” “depress*,” etc.
“Major depression,” “sadness,” “late-life depression,” “LLD,” “aged,” “elder*,” “geriatric,” “senior people,” “RCTs,” etc. All titles, keywords, and abstracts have been reviewed in accordance with our search criteria. In this study, these research articles were exclusively published in English, focusing on telecare interventions for LLD. The specific search strategy is shown in Multimedia Appendix 2.

**Study Selection and Data Exclusion**

The inclusion criteria were the following: (1) studies were RCTs reported in full text with their title and abstract; (2) the average age of the study population was at least 60 years; (3) participants were diagnosed with depression in accordance with any established diagnostic criteria or with a score above a cutoff of any established depression rating scale at baseline; (4) the studies compare telecare (mobile phone, telephone, app, video, etc) participants with the control group receiving usual care (UC; routine, offline, or standard care); and (5) any health care professional providing care (ie, psychiatrists, family physicians, nurses, psychologists, etc). Exclusion criteria were the following: (1) patients with manic or psychotic episodes or symptoms; (2) studies not related to the objective of this review and insufficient data, such as failure to report depression scale scores; and (3) books and studies without full text and studies in the format of abstracts of conference papers.

**Data Extraction**

Two authors independently reviewed all the databases, with specific search strategies for the relevant articles (MW and CYL). The software EndNote X9 (Clarivate) was used to import all the references and remove duplicates. After removing duplicates, the relevance of the title and abstract of the articles was evaluated. Any disagreements were discussed until a consensus was reached. After screening the title and abstract, the articles were selected for the next step of a full-text review. The 2 authors screened the full-text articles independently (MW and CYL). Finally, eligible articles included in the study were processed based on inclusion and exclusion criteria. Any discrepancies that arose during the assessment were resolved by a third reviewer (FY). Two authors independently extracted data from the included studies and entered them into a predesigned data extraction form. Data extracted for this study included the following: first author, year of publication, country, sample size, mean age, intervention approach, duration, presence or absence of comorbid chronic conditions, depression degree, and outcome measurement tools (Table 1).
Table 1. Basic characteristics of the included studies (N=12; all are randomized controlled trials).

<table>
<thead>
<tr>
<th>First author (year); country</th>
<th>Sample size, N (TC(^a)/UC(^b))</th>
<th>Age (years), mean (SD)</th>
<th>Duration</th>
<th>Comorbid chronic diseases</th>
<th>Depression degree</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rollman (2009) [31]; United States</td>
<td>302 (150/152)</td>
<td>TC: 64 (10.8); UC: 64 (11.2)</td>
<td>Baseline, 8 months</td>
<td>Yes</td>
<td>Moderate</td>
<td>HAM-D(^c) and SF-36(^d)</td>
</tr>
<tr>
<td>Aburizik (2013) [32]; United States</td>
<td>52 (29/23)</td>
<td>TC: 66.4 (7.9); UC: 64.1 (10.5)</td>
<td>Baseline, 10 weeks</td>
<td>Yes</td>
<td>Mild</td>
<td>PHQ-9(^e) and BDI(^f)</td>
</tr>
<tr>
<td>Lee (2014) [23]; Korea</td>
<td>25 (12/13)</td>
<td>TC: 66.7 (7.9); UC: 65.4 (8.6)</td>
<td>Baseline, 6 months</td>
<td>Yes</td>
<td>Mild, moderate</td>
<td>CES-D(^g)</td>
</tr>
<tr>
<td>Villani (2014) [33]; Italy</td>
<td>80 (40/40)</td>
<td>TC: 71 (4); UC: 73 (5)</td>
<td>Baseline, 12 months</td>
<td>Yes</td>
<td>Moderate, severe</td>
<td>PHQ-9 and STAI-6(^b)</td>
</tr>
<tr>
<td>Pickett (2014) [34]; United States</td>
<td>124 (60/64)</td>
<td>TC: 69.1 (10.9); UC: 68.6 (10.7)</td>
<td>Baseline, 12 weeks</td>
<td>No</td>
<td>Mild</td>
<td>PHQ-9</td>
</tr>
<tr>
<td>O'Neil (2014) [35]; Australia</td>
<td>121 (61/60)</td>
<td>TC: 61.0 (10.2); UC: 58.9 (10.7)</td>
<td>Baseline, 6 months</td>
<td>Yes</td>
<td>Mild, moderate</td>
<td>PHQ-9, CDS(^i), and SF-12(^j)</td>
</tr>
<tr>
<td>Gellis (2014) [36]; United States</td>
<td>94 (46/48)</td>
<td>TC: 80.1 (7.8); UC: 78.3 (6.9)</td>
<td>Baseline, 3 months, and 6 months</td>
<td>Yes</td>
<td>Mild, moderate</td>
<td>PHQ-9, HAM-D, and SF-12</td>
</tr>
<tr>
<td>Yang (2019) [37]; China</td>
<td>212 (107/105)</td>
<td>TC: 61.25 (8.60); UC: 60.85 (10.80)</td>
<td>Baseline, 12 months</td>
<td>Yes</td>
<td>Mild, moderate</td>
<td>HADS-D(^k) and SDS(^l)</td>
</tr>
<tr>
<td>Naik (2019) [21]; United States</td>
<td>225 (136/89)</td>
<td>61.9 (8.3)</td>
<td>Baseline, 6 months, and 12 months</td>
<td>Yes</td>
<td>Moderate</td>
<td>PHQ-9</td>
</tr>
<tr>
<td>Dobkin (2020) [22]; United States</td>
<td>72 (37/35)</td>
<td>TC: 65.62 (9.76); UC: 64.80 (9.62)</td>
<td>Baseline, 3 months, and 6 months</td>
<td>Yes</td>
<td>Moderate</td>
<td>HAM-D, BDI, HAM-A(^m), and SF-36</td>
</tr>
<tr>
<td>Almeida (2021) [38]; Australia</td>
<td>200 (79/121)</td>
<td>≥65</td>
<td>Baseline, and 52 weeks</td>
<td>No</td>
<td>Mild, moderate</td>
<td>PHQ-9, GAD-7(^n), and SF-12</td>
</tr>
<tr>
<td>Koehler (2021) [39]; Germany</td>
<td>156 (79/77)</td>
<td>TC: 68.30 (9.13); UC: 64.34 (11.35)</td>
<td>Baseline, 12 months</td>
<td>Yes</td>
<td>Moderate</td>
<td>PHQ-9 and SF-36</td>
</tr>
</tbody>
</table>

\(^a\)TC: telecare.  
\(^b\)UC: usual care.  
\(^c\)HAM-D: Hamilton Depression Rating Scale.  
\(^d\)SF-36: 36-Item Short Form Survey.  
\(^e\)PHQ-9: Patient Health Questionnaire-9.  
\(^f\)BDI: Beck Depression Inventory.  
\(^g\)CES-D: Center for Epidemiological Survey, Depression Scale.  
\(^h\)STAI-6: Spielberger’s State Trait Anxiety Inventory.  
\(^i\)CDS: Cardiac Depression Scale.  
\(^j\)SF-12: 12-Item Short Form Survey.  
\(^k\)HADS-D: Hospital Anxiety and Depression Scale.  
\(^l\)SDS: Zung Self-Rating Depression Scale.  
\(^m\)HAM-A: Hamilton Anxiety Rating Scale.  
\(^n\)GAD-7: 7-item Generalized Anxiety Disorder Scale.

Quality Assessment
Two authors (MW and CYL) independently assessed the quality of the studies using the Cochrane Risk of Bias tool [40]. The assessment tool included 7 items (random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias), and authors judged each item individually as “low risk,” “high risk,” and “unclear risk.” The study was considered to be of high quality with a low risk score for at least 4 domains, of which 3 key areas had to be included (random sequence generation, allocation concealment, and incomplete outcome data). Consensus was reached by 2 authors (MW and CYL) through discussion with a third evaluator (FY).

Statistical Analysis
Data were analyzed using Stata (version 16.0; StataCorp) and Review Manager (version 5.3; The Cochrane Collaboration). Intervention effects were estimated by calculating Cohen’s d standardized mean differences (SMDs) and 95% CIs [41]. All studies reported outcomes as continuous data. The Cochran Q
statistic and $I^2$ statistic were used to assess the statistical heterogeneity between selected studies. Random-effects models were used when study heterogeneity was high ($P<.10; I^2>50\%$); otherwise, a fixed-effects model would be used. When heterogeneity identified across studies was high, we further performed subgroup analyses to explore possible explanations for heterogeneity. Publication bias was measured using a funnel plot and Egger linear regression analysis, and $P<.05$ on the Egger test indicated statistically significant publication bias [42].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. RCT: randomized controlled trial.

Results

Literature Search

The database search yielded 15,265 articles, of which 14,249 publications were excluded. A total of 1016 full-text articles were assessed for eligibility. Finally, only 12 studies were eligible for inclusion in this meta-analysis [21-23,31-39], all of which were RCTs published between 2009 and 2021. The PRISMA flow diagram is shown in Figure 1.

Risks of Bias and Quality Assessment

Overall, the quality of the included studies was moderate, of which 5 (41.7%) were of high quality. These studies show that the main bias in the blinding of participants and personnel may be caused by the nature of the intervention measures. All 12 articles reported adequate random sequence generation and, therefore, had a low risk of bias in this regard. In addition, 5 studies reported allocation concealment, which is a low risk of bias. As for detection bias, the assessors were blinded in 7 studies, the presence of blinding was unclear in 3 studies, and 2 studies were not blinded. The risks of study attrition bias and reporting bias were both low. Other risks of bias were also low but were unclear in 1 study. The specific risk of bias and quality assessment results are shown in Figures 2 and 3 [21-23,31-39].
Study and Patient Characteristics

The characteristics of the studies included are summarized in Table 1. A total of 1663 patients with LLD were involved, with an average age of over 60 years in each group. The sample size ranged from 25 [23] to 302 [31] participants. Studies were carried out across 6 countries, including the United States (n=6) [21,22,31,32,34,36], Korea (n=1) [23], Italy (n=1) [33], Australia (n=2) [35,38], China (n=1) [37], and Germany (n=1) [39]. Nine of these used telephone-based interventions, while the remaining studies used remote monitoring systems. Durations ranged from 10 weeks to 52 weeks. Depression, anxiety symptoms, and QoL were substantial influencing factors of treatment for older adults. Therefore, our primary outcome of interest was depression, and secondary outcomes were anxiety symptoms and QoL. Depression was evaluated using the Hamilton Depression Rating Scale, Patient Health Questionnaire-9 (PHQ-9), Beck Depression Inventory, Center for Epidemiological Survey, Depression Scale, Cardiac Depression Scale, Hospital Anxiety and Depression Scale, and Zung Self-Rating Depression Scale. Anxiety symptoms were assessed using Spielberger’s State Trait Anxiety Inventory, Hamilton Anxiety Rating Scale, and the 7-item Generalized Anxiety Disorder Scale. QoL was assessed using the 12-Item Short Form Survey and the 36-Item Short Form Survey. A higher score on the scales indicated better QoL and
greater severity of depression and anxiety symptoms. The specific interventions are available in Multimedia Appendix 3.

**Depression Symptoms**

A total of 12 RCTs involving 1663 participants were identified in this meta-analysis to calculate the effectiveness of telecare on depression, anxiety symptoms, and QoL in patients with LLD.

To compare the effects of telecare and UC in improving LLD, we included data from 12 of these studies. Our results show that telecare significantly reduced depressive symptoms in patients with LLD compared to those in UC (SMD=–0.46, 95% CI –0.53 to –0.38; \(P<.001\)). Fixed-effects model analysis revealed significant heterogeneity among the 12 included studies (\(I^2=83.16\%; P<.001; \) Figure 4) [21-23,31-39].

**Figure 4.** Forest plot for primary outcomes: depression. a: Hamilton Depression Rating Scale; b: Patient Health Questionnaire-9; c: Beck Depression Inventory; d: Center for Epidemiological Survey, Depression Scale; e: Cardiac Depression Scale; f: Hospital Anxiety and Depression Scale; g: Zung Self-Rating Depression Scale.

To address high heterogeneity, we performed subgroup analyses grouped by the type of scale (PHQ-9 or others), duration time (≤3 months or >3 months), device type (telephone-based or remote monitoring system), comorbid chronic diseases (presence or absence), and region (Europe and the Americas or others).

Random-effects models indicated that telecare significantly reduced depressive symptoms in patients with LLD compared to the UC participants (SMD=–0.59, 95% CI –0.80 to –0.38; \(P<.001\)). Results of subgroup analysis by duration showed that short-term (≤3 months) interventions (SMD=–0.72, 95% CI –1.16 to –0.28; \(P<.001\)) were more effective than long-term (>3 months) interventions (SMD=–0.52, 95% CI –0.75 to –0.29; \(P<.001\)); other scales (SMD=–0.65, 95% CI –0.96 to –0.35; \(P<.001\)) were more effective than the PHQ-9 (SMD=–0.53, 95% CI –0.83 to –0.22; \(P<.001\)); the remote monitoring system (SMD=–1.13, 95% CI –1.51 to –0.76; \(P<.001\)) was more effective than telephone-based interventions (SMD=–0.38, 95% CI –0.56 to –0.20; \(P<.001\)); the effect on patients with LLD with chronic diseases (SMD=–0.67, 95% CI –0.89 to –0.44; \(P<.001\)) was better than that on patients with LLD without comorbid chronic diseases (SMD=–0.10, 95% CI –0.41 to 0.20; \(P=.07\)); and telecare was more effective in Europe and the Americas (SMD=–0.73, 95% CI –0.99 to –0.47; \(P<.001\)) than in other regions (SMD=–0.22, 95% CI –0.35 to –0.09; \(P=.42\); Table 2).
Table 2. Subgroup meta-analysis for patients with late-life depression.

<table>
<thead>
<tr>
<th>Subgroups</th>
<th>Cohen $d$ SMD$^a$</th>
<th>95% CI</th>
<th>$P$ value</th>
<th>Heterogeneity ($I^2$; %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>-0.59</td>
<td>-0.80 to -0.38</td>
<td>&lt;.001</td>
<td>86.42</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3 months</td>
<td>-0.72</td>
<td>-1.16 to -0.28</td>
<td>&lt;.001</td>
<td>86.91</td>
</tr>
<tr>
<td>&gt;3 months</td>
<td>-0.52</td>
<td>-0.75 to -0.29</td>
<td>&lt;.001</td>
<td>84.72</td>
</tr>
<tr>
<td><strong>Type of scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9$^b$</td>
<td>-0.53</td>
<td>-0.83 to -0.22</td>
<td>&lt;.001</td>
<td>87.13</td>
</tr>
<tr>
<td>Others</td>
<td>-0.65</td>
<td>-0.96 to -0.35</td>
<td>&lt;.001</td>
<td>86.21</td>
</tr>
<tr>
<td><strong>Device type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telephone-based</td>
<td>-0.38</td>
<td>-0.56 to -0.20</td>
<td>&lt;.001</td>
<td>75.30</td>
</tr>
<tr>
<td>Remote monitoring system</td>
<td>-1.13</td>
<td>-1.51 to -0.76</td>
<td>&lt;.001</td>
<td>78.32</td>
</tr>
<tr>
<td><strong>Comorbid chronic diseases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence</td>
<td>-0.67</td>
<td>-0.89 to -0.44</td>
<td>&lt;.001</td>
<td>85.31</td>
</tr>
<tr>
<td>Absence</td>
<td>-0.10</td>
<td>-0.41 to 0.20</td>
<td>.07</td>
<td>61.45</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe and the Americas</td>
<td>-0.73</td>
<td>-0.99 to -0.47</td>
<td>&lt;.001</td>
<td>86.38</td>
</tr>
<tr>
<td>Others</td>
<td>-0.22</td>
<td>-0.35 to -0.09</td>
<td>.42</td>
<td>0.00</td>
</tr>
</tbody>
</table>

$^a$SMD: standardized mean difference.

$^b$PHQ-9: Patient Health Questionnaire-9.

Meta-regression analysis showed that heterogeneity may not be related to the year of publication ($P=.42$), total sample size ($P=.21$), study area ($P=.35$), comorbid chronic disease ($P=.47$), duration ($P=.75$), and outcome measurement tools ($P=.29$). However, only the intervention device ($P=.004$) may have contributed to the heterogeneity.

**Sensitivity Analysis and Publication Bias**

The stability and reliability of the results of this meta-analysis and potential factors contributing to heterogeneity were explored by sensitivity analysis to assess the effect of the data of each study on the combined effect value (ie, SMD). The results of the sensitivity analysis showed that excluding each study individually had no significant effect on the combined effect value, and the study results were stable and reliable (Figure 5) [21-23,31-39]. Publication bias was assessed using funnel plots and Egger test indicators. The funnel plot was symmetrically distributed on both sides (Figure 6), and the Egger test showed no significant publication bias ($P=.05$).
Anxiety Symptoms

To examine the efficacy of telecare in reducing anxiety compared with that of UC, we included 3 articles on patients with LLD. The results showed that telecare significantly reduced anxiety symptoms in patients with LLD (SMD = –0.53, 95% CI –0.73 to –0.33; \( P = .02 \); Figure 7) [22,33,38].
**QoL**

Six studies assessed the mental components of QoL by using the Medical Outcomes Study Short Form survey. Our meta-analysis shows that the QoL of patients with LLD improved, but, overall, it was not significant (SMD=0.30, 95% CI 0.18-0.43; \(P=0.80\); Figure 8) [22,31,35,36,38,39].

**Discussion**

**Principal Findings**

This meta-analysis shows that compared with UC, telecare significantly reduces symptoms of depression and anxiety but has no significant effect on improving QoL in patients with LLD.

**Primary Outcome Measures**

The pooled results show that telecare has a significant effect on reducing depressive symptoms in patients with LLD, which is consistent with the findings of previous studies [25,26,29]. Apart from dealing with depression itself, the increased severity of LLD is also related to factors such as aging, chronic disease, and socioeconomic stress [5]. Telecare offers unique and innovative opportunities for treating depression symptoms in older adults. Patients with LLD can leverage the advantages of telecare to connect with health care professionals, overcoming geographical distance and physical limitations, thereby reducing the psychological burden of coping with the disease [43]. Furthermore, professional psychological support is crucial for patients with LLD, and it can encourage patients to express their feelings and release stress [44]. However, it is worth noting that despite telecare offering more possibilities for treating LLD, the complexity of the medical population makes it challenging. Telecare can provide greater coverage for health care, yet considerations such as individual needs of older patients or environmental backgrounds need to be factored in [45]. Currently, offering targeted telecare services to a large population of older adults in rural, remote, or underserved areas remains a challenge [10]. In particular, older adults face significant barriers in using telephone and internet connections during the COVID-19 pandemic [46]. As a result, telecare management may not be as effective for this population as for others. The size of the research effect will depend on the nature of the intervention and the quality of the study [47]. High-quality telemedicine will help older adults benefit both physically and mentally. Further investigation and more research are necessary. Subgroup analysis indicates that the effectiveness of telecare in treating LLD can be influenced by measurement tools.
durations, intervention devices, comorbid chronic conditions, and regions involved. In terms of depression measurement tools, other scales appear to be more effective than PHQ-9 (0.65% vs 0.53%), which may be related to measurement errors caused by differences in specific items and the generalizability of different measurement tools [48]. Results from durations of ≤3 and >3 months showed a reduction in depressive symptoms in patients with LLD, with short-term interventions proving to be more effective (0.72% vs 0.52%). Short-term interventions focus more on addressing specific issues or symptoms, producing immediate effects. For older adults, short-term interventions might be more readily accepted as long-term treatments could induce fatigue or a lack of patience. Our findings differ slightly from those of another study [49], which implemented more targeted interventions based on different treatment responses, confirming the more significant effectiveness of long-term interventions. Therefore, there is insufficient evidence to conclusively establish that telecare is necessarily superior in short-term intervention efficacy for LLD compared to long-term interventions. In fact, for depression management, a combination of short-term and long-term interventions is often required to deliver comprehensive and enduring support and management [50].

Subgroup analysis also found that remote monitoring systems appear to be more effective than telephone-based management (1.13% vs 0.38%). The remote monitoring system ensures timely and accurate transmission of patients’ symptom information and data to health care professionals, enabling patients to receive effective treatments [51]. Telecare was more effective in patients with LLD with comorbid chronic conditions compared to those without such comorbidities (0.67% vs 0.10%). Co-occurrence of chronic medical conditions and depression is common. Evidence suggests that older adults with chronic illness are more likely to be affected by depressive symptoms than those without chronic illness [2,5,7]. Older adults with chronic conditions are more likely to seek medical care and adhere to treatment [52]. Therefore, while actively treating chronic conditions, there might be a degree of alleviation in depressive symptoms among older adults. Telecare was more effective in Europe and the Americas in improving depressive symptoms in patients with LLD compared than in other regions (0.73% vs 0.22%). The health care systems in Europe and the Americas are generally more developed, which may lead to more comprehensive support for telecare [53]. In low- and middle-income countries, the resources available for geriatric mental health care are considered severely inadequate [54]. Nevertheless, telecare is beginning to have an important impact on many aspects of health care in transitional countries [55].

Secondary Outcome Measures

Telecare has a positive effect on improving anxiety symptoms of patients with LLD. This result is consistent with findings from other studies [56]. Telecare offers a more convenient access method, allowing patients to receive treatment at home, thereby circumventing the inconvenience and anxiety associated with hospital visits [16,17]. Health care professionals can engage with patients more frequently through telecare, gaining insights into their symptoms and emotional fluctuations. This allows for adjustments in the treatment plan to effectively meet the unique needs of this population [18,57]. Additionally, the symptoms of anxiety and depression are often co-occurring [58], particularly among older adults. Due to the similarity between depression and anxiety symptoms, many treatment approaches are shared between the two. A recent meta-analysis suggests that psychotherapy delivered remotely is as effective as face-to-face therapy for anxiety disorder [59]. This evidence is based on outcomes obtained from different age groups. It may be more challenging to create a trusting relationship remotely than in person [60]. Older adults have negative views about health IT performing accurately and dependably, which will have a significant impact on the acceptance of telecare [61]. In brief, when using telecare for addressing emotional disorders in older adults, closer supervision and guidance might be necessary. Health care professionals need to distinguish the appropriateness of using telecare for communication and, in turn, individually tailor patient care.

We found that the mental component of QoL in patients with LLD improved after using telecare; however, this difference was not significant compared to that with the use of UC. This finding aligns with results from other studies [62,63]. Improving QoL is a comprehensive issue that includes not only improvements in health care but also social, psychological, and emotional factors [64]. Influenced by these factors, it is difficult to compare the results of QoL considering different contexts. Several results from RCTs with older adults using telemonitoring systems showed an improvement in the participants’ QoL [65,66]; other telemonitoring RCTs could not achieve congruent results [67]. Improvements in QoL often require deeper interactions and personalized care. In particular, participants with mental disorders may benefit from individual and tailored solutions provided by general practitioners [68]. When using telecare, it is crucial to acknowledge that each subpopulation of marginalized older adults has differing strengths and needs. The studies we included focused more on managing the disease itself, which may weaken overall effectiveness. It is not easy to present telemedicine to the older population. The limitations inherent in older adults may lead to difficulties in receiving telecare, including the lack of technical literacy, equipment access barriers, cognitive function, etc [11]. These reasons could explain why telecare is not significant in improving the QoL of patients with LLD. The potential value of telecare in maintaining the QoL for individuals with LLD warrants further exploration. While this study did not reveal a positive impact of telecare on the QoL for patients with LLD, it has been established that telecare can assist patients with LLD in gaining more information about health services.

Limitations

This study still had some limitations. First, most of the studies included in the review lacked sufficient measure detail, leading to irreversible bias. Our study mainly included 2 interventions based on telephone and remote monitoring to reduce this bias. Second, the measurement tools used in this study lacked standardization and heavily relied on self-reports from participants, introducing a degree of subjectivity and concealment that is not as rigorous as structured interviews. However, we attempted to validate the effectiveness of the results by using authoritative scales. Third, differences in the

https://mhealth.jmir.org/2024/1/e50787
specific intervention methods, frequency, and content among the included studies may lead to clinical heterogeneity across different studies.

**Conclusions**

Our meta-analysis shows that telecare has a positive impact on depressive and anxiety symptoms, despite high heterogeneity in depression symptoms. Therefore, studies with larger sample sizes and homogeneity were required to determine the effects of telecare in patients with LLD. Future research can continue to refine telecare systems and assess the specific needs of older vulnerable populations for more accurate evidence.

---

**Acknowledgments**

This project was jointly supported by the Hubei Provincial Natural Science Foundation and the Innovative Development of Chinese Medicine of China (2023AFD160).

**Conflicts of Interest**

None declared.

---

**References**


https://mhealth.jmir.org/2024/1/e50787 Jmir MHEALTH AND UHEALTH 2024 | vol. 12 | e50787 p.135 (page number not for citation purposes)


Abbreviations

LID: late-life depression  
MeSH: Medical Subject Headings  
PHQ-9: Patient Health Questionnaire-9  
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
QoL: quality of life  
RCT: randomized controlled trial  
SMD: standardized mean difference  
UC: usual care

Edited by L Buis; submitted 12.07.23; peer-reviewed by W He, X Tan, S Peng; comments to author 15.11.23; revised version received 01.12.23; accepted 22.12.23; published 17.01.24.

Please cite as:

URL: https://mhealth.jmir.org/2024/1/e50787  
doi:10.2196/50787  
PMID:38231546

©Man Wu, Chaoyang Li, Ting Hu, Xueyong Zhao, Guiyuan Qiao, Xiaolian Gao, Xinhong Zhu, Fen Yang. Originally published in JMIR MHealth and uHealth (https://mhealth.jmir.org), 17.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,
distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Abstract

Background: Musculoskeletal diseases affect 1.71 billion people worldwide, impose a high biopsychosocial burden on patients, and are associated with high economic costs. The use of digital health interventions is a promising cost-saving approach for the treatment of musculoskeletal diseases. As physical exercise is the best clinical practice in the treatment of musculoskeletal diseases, digital health interventions that provide physical exercises could have a highly positive impact on musculoskeletal diseases, but evidence is lacking.

Objective: This systematic review aims to evaluate the impact of digital physical health exercises on patients with musculoskeletal diseases concerning the localization of the musculoskeletal disease, patient-reported outcomes, and medical treatment types.

Methods: We performed systematic literature research using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The search was conducted using the PubMed, BISP, Cochrane Library, and Web of Science databases. The Scottish Intercollegiate Guidelines Network checklist was used to assess the quality of the included original studies. To determine the evidence and direction of the impact of digital physical health exercises, a best-evidence synthesis was conducted, whereby only studies with at least acceptable methodological quality were included for validity purposes.

Results: A total of 8988 studies were screened, of which 30 (0.33%) randomized controlled trials met the inclusion criteria. Of these, 16 studies (53%) were of acceptable or high quality; they included 1840 patients (1008/1643, 61.35% female; 3 studies including 197 patients did not report gender distribution) with various musculoskeletal diseases. A total of 3 different intervention types (app-based interventions, internet-based exercises, and telerehabilitation) were used to deliver digital physical health exercises. Strong evidence was found for the positive impact of digital physical health exercises on musculoskeletal diseases located in the back. Moderate evidence was found for diseases located in the shoulder and hip, whereas evidence for the entire body was limited. Conflicting evidence was found for diseases located in the knee and hand. For patient-reported outcomes, strong evidence was found for impairment and quality of life. Conflicting evidence was found for pain and function. Regarding the medical treatment type, conflicting evidence was found for operative and conservative therapies.

Conclusions: Strong to moderate evidence was found for a positive impact on musculoskeletal diseases located in the back, shoulder, and hip and on the patient-reported outcomes of impairment and quality of life. Thus, digital physical health exercises could have a positive effect on a variety of symptoms of musculoskeletal diseases.

(JMIR Mhealth Uhealth 2024;12:e50616) doi:10.2196/50616

KEYWORDS
mobile health; mHealth; electronic health; eHealth; digital health applications; DiGA; musculoskeletal; MSK; home-based; PROM; disorder; mobile phone
**Introduction**

**Background**

A total of 1.71 billion people are affected by musculoskeletal diseases worldwide [1]. They are characterized by chronic pain, functional disability, impairment, and reduced quality of life [1,2]. The most commonly affected body regions are the lower back and neck, with a period prevalence over the last 12 months of up to 61.3% and 45.7% [3], respectively, and a common disease is osteoarthritis, with a prevalence of up to 17.9% [4]. In addition to the high biopsychosocial burden [5], the evident increase in the incidence of musculoskeletal diseases over the last decades [6] results in high economic costs because of lost workdays and conservative or operative medical treatments [5].

To overcome such undesirable consequences, evidence-based, effective, and cost-saving health interventions are required. Therefore, the use of digital health interventions is a promising approach.

Digital health interventions aim to manage a wide range of diseases and health issues using digital devices such as smartphones, tablets, computers, or wearables, including mobile apps, telerehabilitation and web-based physician visits, web-based interactive programs, or tracking tools [7]. The use of mobile apps is increasing, with common intervention types categorized as physical exercise and fitness, lifestyle and stress, diet and nutrition, or medication reminders and educational materials [7]. In some countries, such as Germany, so-called digital health applications are also supported by health insurers after being evaluated as medical devices [8]. However, owing to their cost-saving potential and the increasing number of commercially available digital health interventions [7], further research is needed to evaluate the impact of different types of digital health interventions on specific diseases.

Previous systematic reviews have extensively evaluated the impact of digital health interventions on internal diseases. Positive effects have been demonstrated in treating chronic obstructive pulmonary disease [9], cardiovascular disease [10], and diabetes [11]. These effects encompass improvements in clinically relevant outcomes such as quality of life, health-related impairments, amelioration of risk factors and their consequences, as well as the control and management of HbA1c levels. For musculoskeletal diseases, only 2 previous systematic reviews have evaluated the impact of digital health interventions as a primary outcome. One review [12] showed that there are substantial clinical benefits in the management of musculoskeletal diseases for the patient-reported outcomes of pain (9 out of 19 studies) and functional disability (10 out of 16 studies). The results show that digital health interventions as adjuncts and as stand-alone treatments are not inferior but partly superior compared with interventions based on standard therapy, nondigital self-management, noninteractive digital measures, or no intervention. However, in this previous review, no evidence synthesis was performed. In addition, a further review [13] conducted a meta-analysis and showed moderate-quality evidence that digital health interventions are effective in reducing pain and improving function and self-management in patients with musculoskeletal disease. The included studies considered digital health interventions as interventions that are to be used only at home and as adjuncts to standard clinical care, compared with standard care, noninteractive digital interventions, or no intervention. Taken together, the use of digital health interventions as an adjunct to regular therapy could have positive health-related effects for both internal and musculoskeletal diseases, although less evidence is available for the latter.

However, little is known about the relationship between clinically relevant factors, such as the localization of the musculoskeletal diseases, patient-reported outcomes, or the type of applied conservative or operative medical treatments, and the effects of different types of digital health interventions in the treatment of musculoskeletal diseases. In terms of evidence-based medicine, such relationships must first be clarified when using digital health interventions as a regular treatment option for specific musculoskeletal diseases. Because of the increasing number of original studies, more systematic research is needed to review and assess the existing evidence. Previous systematic reviews [12,13] have included all types of digital health interventions, providing a comprehensive overall result across all biopsychosocial domains. As physical exercise is the best clinical practice for the treatment of musculoskeletal diseases [14], digital physical health exercises could have a highly positive impact on musculoskeletal diseases. However, little is known about how the effects of digital physical health exercises are related to the aforementioned clinically relevant factors.

**Objective**

Therefore, this systematic review aimed to evaluate the impact of digital physical health exercises on patients with musculoskeletal diseases concerning the localization of the musculoskeletal disease, patient-reported outcomes, and medical treatment types. In addition, a best-evidence synthesis was conducted to estimate the direction and strength of the existing evidence.

**Methods**

**Research Design and Eligibility Criteria**

The systematic review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15]. Eligibility criteria according to the population, intervention, comparison, outcome, study design (PICOS) scheme [16] were applied. Table 1 presents the inclusion criteria according to the PICOS scheme. Textbox 1 presents the search line.

The corresponding keywords are also presented. Studies were not reviewed if they did not report on a specific musculoskeletal disease, if the digital health intervention included no physical exercises, if no control group was considered, or if none of the included patient-reported outcomes were assessed as a primary outcome. All methodological steps were performed by 1 author and validated by a second author. Uncertainties were discussed until consensus was reached. Because of the literary nature of this study, ethics approval was not required.
Table 1. PICOS\(^a\) scheme for the definition of the inclusion criteria and the presentation of the corresponding keywords.

<table>
<thead>
<tr>
<th>Population</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcome</th>
<th>Study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with any musculoskeletal disease according to the definition of the WHO(^c)</td>
<td>Any digital health intervention using home-based physical exercises</td>
<td>Any conventional or no therapy</td>
<td>Patient-reported outcomes pain, function, disability, and quality of life assessed by established and validated clinical questionnaires or scales</td>
<td>Randomized controlled trials</td>
</tr>
</tbody>
</table>

**Inclusion criteria\(^b\)**

- "Musculoskeletal disease” OR “Musculoskeletal disorder” OR “Musculoskeletal pain” OR “Chronic pain” OR “Acute pain” OR “Overuse pain” OR “Chronic injury” OR “Chronic injuries” OR “Acute injury” OR “Overuse injury” OR “Overuse injuries” OR “Chronic disease” OR “Acute disease” OR “Overuse disease” OR “Osteoporosis” OR “Osteoarthritis” OR “Rheumatoid arthritis” OR “Tendinopathy” OR “Tendinopathies” OR “Rotator cuff” OR “Lower extremity” OR “Lower extremities” OR “Upper extremity” OR “Upper extremities” OR “Hip” OR “Knee” OR “Foot” OR “Hand” OR “Ankle” OR “Wrist” OR “Elbow” OR “Low back” OR “Neck” OR “Back” OR “Spine” OR “Shoulder” OR “Arm” OR “Leg” OR “Muscle” OR “Tendon” OR “Ligament”

**Keywords**

- “Digital movement therapy” OR “Digital movement therapies” OR “Mobile health” OR “eTherapy” OR “eTherapies” OR “Web-based intervention” OR “Digital intervention” OR “Computer-based intervention” OR “App-based intervention” OR “Digital health application” OR “Technology-assisted therapy” OR “Technology-assisted therapies” OR “Internet-based intervention” OR “Computer-assisted therapy” OR “Computer-assisted therapies” OR “health app” OR “mobile application” OR “Smartphone” OR “Mobile phone” OR “ehealth” OR “mhealth” OR “telerhabilitation” OR “Telemedicine” OR “online intervention” OR “internet-delivered intervention”

- “Osteopathy” OR “movement therapy” OR “movement therapies” OR “physical therapy” OR “physical therapies” OR “therapeutic exercise” OR “medical gymnastic” OR “traditional therapy” OR “traditional therapies” OR “manual therapy” OR “manual therapies” OR “physiotherapy” OR “No therapy” OR “No therapies” OR “conventional therapy” OR “conventional therapies” OR “no treatment” OR “no intervention” OR “wait-and-watch” OR “wait-and-see” OR “watch and wait” OR “wait and see”

\(^a\)PICOS: Population, intervention, comparison, outcome, and study design.

\(^b\)Others: Studies in English or German language with free full access were included.

\(^c\)WHO: World Health Organization.

\(^d\)N/A: Not applicable.

Textbox 1. Search line.

"(Musculoskeletal disease OR musculoskeletal disorder OR musculoskeletal pain OR chronic pain OR acute pain OR overuse pain OR chronic injury OR acute injuries OR acute injuries OR overuse injury OR overuse injuries OR chronic disease OR acute disease OR overuse disease OR osteoporosis OR osteoarthritis OR rheumatoid arthritis OR tendinopathy OR tendinopathies OR rotator cuff OR lower extremity OR lower extremities OR upper extremity OR upper extremities OR hip OR knee OR foot OR hand OR ankle OR wrist OR elbow OR low back OR neck OR back OR spine OR shoulder OR arm OR leg OR muscle OR tendon OR ligament) AND (digital movement therapy OR digital movement therapies OR mobile health OR eTherapy OR eTherapies OR web-based intervention OR digital intervention OR computer-based intervention OR app-based intervention OR digital health application OR technology-assisted therapy OR technology-assisted therapies OR internet-based intervention OR computer-assisted therapy OR computer-assisted therapies OR health app OR mobile application OR smartphone OR mobile phone OR ehealth OR mhealth OR telerhabilitation OR telemedicine OR online intervention OR internet-delivered intervention) AND (osteopathy OR movement therapy OR movement therapies OR physical therapy OR physical therapies OR therapeutic exercise OR medical gymnastic OR traditional therapy OR traditional therapies OR manual therapy OR manual therapies OR physiotherapy OR no therapy OR no therapies OR conventional therapy OR conventional therapies OR no treatment OR no intervention OR watch-and-wait OR wait-and-see OR watch and wait OR wait and see) AND (randomized controlled trials)."
Literature Search, Study Selection, and Risk of Bias

The literature search was performed on July 21, 2022, using the PubMed (MEDLINE), BISp (Federal Institute of Sport Science), Cochrane Library, and Web of Science databases. The search line included terms presented in Table 1. The “outcomes” category was not included in the search strategy but was considered in the subsequent study inclusion and selection process. No filters or other restrictions were used. The retrieved records were exported to a reference manager (EndNote 20, Clarivate). All duplicates were identified using the software and were removed after a manual review. On the basis of the defined eligibility criteria, studies were included or excluded by reviewing the titles, abstracts, and full texts. Full texts were accessed via public or open access and university accounts. If the full texts were not accessible, the authors were contacted. The study quality and the associated risk of bias were assessed using the Scottish Intercollegiate Guidelines Network checklist for randomized controlled trials [17]. The checklist consisted of 10 items related to the internal validity and 2 items related to the overall assessment of the studies. For each included study, all items were answered with “yes,” “no,” “can’t say,” or “not applicable.” The study quality was then finally rated throughout the “Scottish Intercollegiate Guidelines Network checklist for randomized controlled trials: Notes for completion of checklist” as “not acceptable,” “borderline,” “acceptable,” and “high,” as previously done [18]. The definitions of these quality classifications are presented in Textbox 2.

Textbox 2. Definitions for ratings of the overall methodological study quality.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Study quality</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>High quality</td>
<td>Most criteria met. Little or no risk of bias. Results unlikely to be changed by further research.</td>
<td></td>
</tr>
<tr>
<td>Acceptable quality</td>
<td>Most criteria met. Some flaws in the study with an associated risk of bias. Conclusions may change in the light of further studies.</td>
<td></td>
</tr>
<tr>
<td>Borderline quality</td>
<td>Crude effect estimates have been presented or have been calculated (thus no confounders have been considered), but the study is otherwise acceptably sound with respect to other possible biases.</td>
<td></td>
</tr>
<tr>
<td>Not acceptable quality</td>
<td>Either most criteria not met, or significant flaws relating to key aspects of study design. Conclusions likely to change in the light of further studies.</td>
<td></td>
</tr>
</tbody>
</table>

Note: Definitions according to Asker et al [18].

Data Extraction and Synthesis of Results

Data extraction was performed according to the PICOS scheme. A best-evidence synthesis was conducted to clarify the evidence for digital physical health exercises on clinically relevant factors clustered as (1) localization of the musculoskeletal diseases, (2) patient-reported outcomes (according to the eligibility criteria), and (3) medical treatment types (conservative vs operative). Within these clusters, the study results were individually classified as positive, negative, or equal for each clinically relevant factor. In accordance with a previous study [14], the study results were classified as positive or negative if the intervention or control group showed statistically better significant study results than the other group for >50% of the outcome parameters that were used to examine the respective clinically relevant factors. If no statistically significant differences were reported between the intervention and control groups, the study results were classified as equal. With regard to the best-evidence synthesis, the established criteria [18] are summarized in Table 2, and to increase the validity, only studies with at least acceptable study quality were included [19].

Table 2. Criteria of best-evidence synthesis according to Asker et al [18].

<table>
<thead>
<tr>
<th>Rating</th>
<th>Study quality</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong evidence</td>
<td>≥2 high-quality studies</td>
<td>≥75% consistent findings in these studies</td>
</tr>
<tr>
<td>Moderate evidence</td>
<td>1 high-quality study and/or ≥2 acceptable-quality studies</td>
<td>≥75% consistent findings in these studies</td>
</tr>
<tr>
<td>Limited evidence</td>
<td>1 acceptable-quality study and/or ≥1 borderline-quality study</td>
<td>N/Aa</td>
</tr>
<tr>
<td>Conflicting evidence</td>
<td>Conflicting results in several studies of any quality</td>
<td>&lt;75% of studies reported concordant results</td>
</tr>
<tr>
<td>No evidence</td>
<td>No admissible studies were found</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aN/A: Not applicable.
Results

Literature Search, Study Selection, and Risk of Bias

Figure 1 shows the flowchart of the literature search including the study selection process according to the PRISMA guidelines. On the basis of the eligibility criteria, of 10,441 records, 30 (0.29%) studies were finally included in the risk of bias assessment. Although 1453 studies were identified as duplicates, 8958 studies that did not meet the inclusion criteria addressed no musculoskeletal diseases, were not randomized controlled trials, or addressed other outcomes. Table 3 summarizes the results of the risk of bias assessment.

There were 10 studies with high quality [20-29], 6 studies with acceptable quality [30-35], 10 studies with borderline quality [36-45], and 4 studies with not acceptable quality [46-49]. Thus, because of the not acceptable and borderline qualities of 14 studies, 16 studies were further analyzed and finally included in the best-evidence synthesis.

Figure 1. Flowchart of the literature search including the study selection process according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. RCT: randomized controlled trial.
Table 3. Results of the 30 studies checked for the risk of bias assessment using the Scottish Intercollegiate Guidelines Network checklist.

<table>
<thead>
<tr>
<th>Study</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8 (%)</th>
<th>1.9</th>
<th>2.1</th>
<th>2.2</th>
<th>2.3</th>
<th>Total</th>
<th>Study quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abadiyan et al [20]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>3</td>
<td>CS</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Allen et al [21]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>13.1</td>
<td>Yes</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>Blanquero et al [22]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
<td>Yes</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>Choi et al [23]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>CS</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
<td>Yes</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>Fatoye et al [24]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>16</td>
<td>CS</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Fleischman et al [25]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>16.6</td>
<td>Yes</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>Moffet et al [26]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>6.3</td>
<td>Yes</td>
<td>Yes</td>
<td>++</td>
<td>CS</td>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>Nelligan et al [27]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>12.6</td>
<td>Yes</td>
<td>Yes</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>11</td>
</tr>
<tr>
<td>Nelson et al [28]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>CS</td>
<td>Yes</td>
<td>Yes</td>
<td>1</td>
<td>Yes</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Özden et al [29]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>7</td>
<td>CS</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>Bennell et al [30]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>10.1</td>
<td>Yes</td>
<td>N/A</td>
<td>++</td>
<td>Yes</td>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>Chhabra et al [31]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
<td>Yes</td>
<td>N/A</td>
<td>+</td>
<td>Yes</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Hardt et al [32]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>CS</td>
<td>Yes</td>
<td>10</td>
<td>Yes</td>
<td>N/A</td>
<td>+</td>
<td>Yes</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Hernandez-Garrio et al [33]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>18</td>
<td>Yes</td>
<td>N/A</td>
<td>+</td>
<td>Yes</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Rodriguez-Sanchez-Laulhe et al [34]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>16</td>
<td>Yes</td>
<td>N/A</td>
<td>+</td>
<td>Yes</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Tousignant et al [35]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>CS</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>15</td>
<td>CS</td>
<td>CS</td>
<td>+</td>
<td>Yes</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Anan et al [36]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>CS</td>
<td>Yes</td>
<td>25.6</td>
<td>Yes</td>
<td>N/A</td>
<td>_c</td>
<td>CS</td>
<td>Yes</td>
<td>7</td>
</tr>
<tr>
<td>Bäcker et al [37]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>42</td>
<td>No</td>
<td>N/A</td>
<td>_c</td>
<td>CS</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Bossen et al [38]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>24.6</td>
<td>Yes</td>
<td>N/A</td>
<td>_c</td>
<td>CS</td>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>Correia et al [39]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>36</td>
<td>No</td>
<td>N/A</td>
<td>_c</td>
<td>CS</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>del Pozo-Cruz et al [40]</td>
<td>Yes</td>
<td>CS</td>
<td>CS</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>_c</td>
<td>Yes</td>
<td>CS</td>
<td>7</td>
</tr>
</tbody>
</table>
Study Characteristics

Table 4 presents the study characteristics of the 16 included studies according to the PICOS scheme.

The publication period ranged from 2011 [35] to 2022 [29,34], whereby 3 studies were published each in 2018 [21,31,32], 2019 [22,23,25], and 2021 [20,27,33]. The most common publication country was Australia, with 3 studies [27,28,30], followed by 2 studies each published by Spain [22,34], Canada [26,35], and the United States [21,25]. Across the 16 studies, 1840 patients were investigated, and the sample sizes ranged from 34 [33] to 350 patients [21]. The reported dropout rate was up to 18% (6/34) [33]. The average age of the patients varied from 38.5 [20] to 66 years [26,35], and the average female proportion across all studies reported was 61.35% (1008/1643) and varied from 51% (148/290) [25] to 100% (34/34) [33]. Regarding the localization of the musculoskeletal diseases, 7 studies were on knee-specific diseases such as total knee arthroplasty [25,26,32,35], knee osteoarthritis [21,27], and chronic knee pain [30]. This was followed by 4 studies on back-specific diseases such as low back pain [24,29,31] and chronic neck pain [20]. A total of 2 studies were on hand-specific diseases [22,34], whereas only 1 study was found for each full body [33], shoulder-specific diseases [23], and hip-specific diseases [28].

Regarding the patient-reported outcomes, 14, 12, 6, and 5 studies investigated pain [20-23,25-27,29-35], function [21,23,25-30,32-35], disability [20,22,24,29,31,34], and quality of life [20,27-30], respectively. In the 16 included studies, 26 different patient-reported outcomes were investigated. With regard to the digital health interventions, 7, 5, and 4 studies used app-based [20,22,23,28,31,32,34], web-based [21,25,27,29,30], and telerehabilitation-based physical exercises [24,26,33,35], respectively, whereby the duration of the digital health interventions ranged from 7 days [32] to 12 months [21]. As control groups, 9 studies used physiotherapy [20,21,23,25,26,28,31,32,35]; 4 studies used paper-based exercises [22,25,29,34]; 2 studies used internet-based information material [27,30]; and 1 study each used global postural re-education [20], waiting list [21], clinic-based McKenzie therapy [24], and no therapy [33].
<table>
<thead>
<tr>
<th>Study</th>
<th>Population and setting</th>
<th>Intervention and assessment</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abadiyan et al [20]</td>
<td>• Sample size: n=60&lt;br&gt;• Average age: 38.5 y&lt;br&gt;• Female: 55%&lt;br&gt;• Disease: chronic neck pain&lt;br&gt;• Country: Iran</td>
<td>• I (&quot;Seeb&quot; app+ GPR): n=20&lt;br&gt;• C1 (GPR alone): n=20&lt;br&gt;• C2 (conventional PT): n=20&lt;br&gt;• Duration: 8 wk&lt;br&gt;• Survey dates: baseline, 8 wk</td>
<td>• Drop out: 3%&lt;br&gt;• I: 5%, C1: 5%, and C2: 0%&lt;br&gt;• Pain: app+GPR&gt;PT&lt;br&gt;• Neck disability index: app+GPR&gt;PT&lt;br&gt;• GPR&gt;PT&lt;br&gt;• Quality of life: app+GPR&gt;PT&lt;br&gt;• GPR&gt;PT</td>
</tr>
<tr>
<td>Allen et al [21]</td>
<td>• Sample size: n=350&lt;br&gt;• Average age: 65.3 y&lt;br&gt;• Female: 71.7%&lt;br&gt;• Disease: knee osteoarthritis&lt;br&gt;• Country: United States</td>
<td>• I (IBET): n=142&lt;br&gt;• C1 (conventional PT): n=140&lt;br&gt;• C2 (WL): n=68&lt;br&gt;• Duration: 12 mo&lt;br&gt;• Survey dates: baseline, 4, and 12 mo</td>
<td>• Drop out: 13.1%&lt;br&gt;• I: 21.1%, C1: 7.9%, and C2: 7%&lt;br&gt;• Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and other functional tests&lt;br&gt;• IBET=PT=WL</td>
</tr>
<tr>
<td>Bennell et al [30]</td>
<td>• Sample size: n=148&lt;br&gt;• Average age: 61.2 y&lt;br&gt;• Female: 56.1%&lt;br&gt;• Disease: chronic knee pain&lt;br&gt;• Country: Australia</td>
<td>• I: internet-based education material supported by videoconferences with physiotherapist for home exercises (n=74)&lt;br&gt;• Control group: internet-based education material only (n=74)&lt;br&gt;• Duration: 9 mo&lt;br&gt;• Survey dates: baseline, 3, and 9 mo</td>
<td>• Drop out: 10.1%&lt;br&gt;• I: 11% and C: 10%&lt;br&gt;• Pain and function: education+PT&gt;education&lt;br&gt;• Quality of life: education+PT&gt;education</td>
</tr>
<tr>
<td>Blanquero et al [22]</td>
<td>• Sample size: n=50&lt;br&gt;• Average age: 50.0 y&lt;br&gt;• Female: 82%&lt;br&gt;• Disease: carpal tunnel release&lt;br&gt;• Country: Spain</td>
<td>• I: ReHand app for physical home training (n=25)&lt;br&gt;• Control group: paper and home-based physical exercise program (n=25)&lt;br&gt;• Duration: 4 wk&lt;br&gt;• Survey dates: baseline, 4 wk</td>
<td>• Drop out: 0%&lt;br&gt;• Hand disability and pain: app based&gt;paper based</td>
</tr>
<tr>
<td>Chhabra et al [31]</td>
<td>• Sample size: n=93&lt;br&gt;• Average age: 41.2 y&lt;br&gt;• Female: not reported&lt;br&gt;• Disease: chronic low back pain&lt;br&gt;• Country: India</td>
<td>• I: Snapcare app for physical home training (n=45)&lt;br&gt;• Control group: conventional therapy (n=48)&lt;br&gt;• Duration: 12 wk&lt;br&gt;• Survey dates: baseline, 12 wk</td>
<td>• Drop out: 0%&lt;br&gt;• Pain: app based=conventional&lt;br&gt;• Disability: app based=conventional&lt;br&gt;• Current Symptom Score: app based&gt;conventional</td>
</tr>
<tr>
<td>Choi et al [23]</td>
<td>• Sample size: n=84&lt;br&gt;• Average age: 54.5 y&lt;br&gt;• Female: 68%&lt;br&gt;• Disease: frozen shoulder&lt;br&gt;• Country: Korea</td>
<td>• I: app (no name given) for physical home training (n=42)&lt;br&gt;• Control group: conventional home-based self-exercises (n=42)&lt;br&gt;• Duration: 3 mo&lt;br&gt;• Survey dates: baseline, 4, 8, and 12 wk</td>
<td>• Drop out: 0%&lt;br&gt;• Pain and range of motion: app based=conventional</td>
</tr>
<tr>
<td>Fatoye et al [24]</td>
<td>• Sample size: n=56&lt;br&gt;• Average age: 48.7 y&lt;br&gt;• Female: not reported&lt;br&gt;• Disease: chronic low back pain&lt;br&gt;• Country: Nigeria</td>
<td>• I: telerehabilitation home-based McKenzie therapy (TBMT): n=24&lt;br&gt;• Control group: clinic-based McKenzie therapy (CBMT): n=32&lt;br&gt;• Duration: 8 wk&lt;br&gt;• Survey dates: baseline, 4, and 8 wk</td>
<td>• Drop out: 16%&lt;br&gt;• I: 13% and C: 19%&lt;br&gt;• Disability: TBMT=CBMT</td>
</tr>
<tr>
<td>Study</td>
<td>Population and setting</td>
<td>Intervention and assessment</td>
<td>Outcomes</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Fleischman et al [25]  | • Sample size: n=290  
• Average age: 65.0 y  
• Female: 51%  
• Disease: total knee arthroplasty  
• Country: United States | • I: web-based PT at home (n=96)  
• C1: paper-based PT at home (n=97)  
• C2: formal outpatient PT (n=97)  
• Duration: 6 mo  
• Survey dates: baseline, 4-6 wk, 6 mo | • Drop out: 15.9%  
• I: 17%, C1: 27%, and C2: 6%  
• Knee flexion and Knee Injury and Osteoarthritis Outcome Score (KOOS): Web PT= paper PT=PT |
| Hardt et al [32]       | • Sample size: n=60  
• Average age: 65.9 y  
• Female: 57%  
• Disease: total knee arthroplasty  
• Country: Germany | • I: PT+“GenuSport” app (PT+app; n=33)  
• Control group: PT (n=27)  
• Duration: 7 d  
• Survey dates: daily for 7 d | • Drop out: 10%  
• I: 15% and C: 7%  
• Active range of motion, pain, function, KOOS, and Knee Society Score: PT+app>PT |
| Hernando-Garjo et al [33] | • Sample size: n=34  
• Average age: 53.4 y  
• Female: 100%  
• Disease: fibromyalgia  
• Country: Mexico | • I: telerehabilitation with home-based aerobic exercises (n=17)  
• Control group: no additional intervention (n=17)  
• Duration: 15 wk  
• Survey dates: baseline, 15 wk | • Drop out: 18%  
• I: 18% and C: 18%  
• Physical function: telerehabilitation=nothing |
| Moffet et al [26]      | • Sample size: n=205  
• Average age: 66.0 y  
• Female: 51.2%  
• Disease: total knee arthroplasty  
• Country: Canada | • I: home-based telerehabilitation (n=104)  
• Control group: home-visiting PT (n=101)  
• Duration: 2 mo  
• Survey dates: baseline, 2, and 4 mo | • Drop out: 6.3%  
• I: 9.6% and C: 2.9%  
• WOMAC, KOOS, function, and range of motion: telerehabilitation=PT |
| Nelligan et al [27]    | • Sample size: n=206  
• Average age: 60.0 y  
• Female: 61.2%  
• Disease: knee osteoarthritis  
• Country: Australia | • I: website (information+active exercises) and text messages (n=103)  
• Control group: website with information only (n=103)  
• Duration: 24 wk  
• Survey dates: baseline, 24 wk | • Drop out: 12.6%  
• I: 12.6% and C: 12.6%  
• Pain, WOMAC, KOOS, quality of life: website information+exercise>website information only |
| Nelson et al [28]      | • Sample size: n=70  
• Average age: 64.5 y  
• Female: 63%  
• Disease: total hip replacement  
• Country: Australia | • I: telerehabilitation and technology-based home exercise (n=35)  
• Control group: PT and paper-based home exercise (n=35)  
• Duration: 6 wk  
• Survey dates: baseline, 6 wk, 6 mo | • Drop out: 1%  
• I: 3% and C: 0%  
• Quality of life and function: telerehabilitation+exercise=PT+exercise |
| Özden et al [29]       | • Sample size: n=50  
• Average age: 41.3 y  
• Female: 60%  
• Disease: low back pain  
• Country: turkey | • I: telerehabilitation with Fizyoweb software (n=25)  
• Control group: same exercises with paper-based instructions (n=25)  
• Duration: 8 wk  
• Survey dates: baseline, 8 wk | • Drop out: 7%  
• I: 7% and C: 7%  
• Pain, function, disability, and quality of life: telerehabilitation>paper based |
| Rodríguez Sánchez-Laulhé et al [34] | • Sample size: n=36  
• Average age: 59.8 y  
• Female: 61%  
• Disease: rheumatoid arthritis  
• Country: Spain | • I: CareHand app for exercises and self-management and monitoring tools (n=14)  
• Control group: paper-based home exercises (n=22)  
• Duration: 3 mo  
• Survey dates: baseline, 1, 3, and 6 mo | • Drop out: 16%  
• I: 7% and C: 22%  
• Function: app based>paper based  
• Pain and disability for upper extremity: app based>paper based |
| Toussignant et al [35] | • Sample size: n=48  
• Average age: 66.0 y  
• Female: not reported  
• Disease: total knee arthroplasty  
• Country: Canada | • I: telerehabilitation by videoconference with a physiotherapist (n=24)  
• Control group: conventional PT (n=24)  
• Duration: 2 mo  
• Survey dates: baseline, 2, and 6 mo | • Drop out: 15%  
• I: 12% and C: 17%  
• Disability: telerehabilitation=conventional PT  
• Function: telerehabilitation>conventional PT  
• Functional activity, physical functioning, and physical pain: conventional PT>telerehabilitation |

IA: Intervention group.  
GPR: Global postural re-education.
Synthesis of Results by Best Evidence Synthesis

Tables 5, 6, and 7 show the results of the best-evidence synthesis with regard to the cluster of the localization of the musculoskeletal diseases, patient-reported outcomes, and medical treatment types, respectively.

Regarding the localization of the musculoskeletal diseases, there was strong evidence that digital physical health exercises had a positive impact on the musculoskeletal diseases located in the back. Although moderate evidence was obtained for diseases located in the shoulder and hip, evidence for fibromyalgia (the entire body) is limited. Conflicting evidence was found for diseases located in the knee and hand. For the patient-reported outcomes, there was strong evidence that digital physical health exercises had a positive impact on disability and quality of life. Conflicting evidence was found for pain and function. Regarding the medical treatment types, operative and conservative therapies both achieved conflicting evidence. Figure 2 shows the evidence found across the 3 defined clusters for studies included in the best-evidence synthesis.

Table 5. Best-evidence synthesis for the localization of the musculoskeletal diseases.

<table>
<thead>
<tr>
<th>Localization</th>
<th>Study</th>
<th>Musculoskeletal disease</th>
<th>Results</th>
<th>Study quality</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back</td>
<td>Abadiyan et al [20]</td>
<td>Chronic neck pain</td>
<td>+\textsuperscript{a}</td>
<td>High</td>
<td>Strong\textsuperscript{b}</td>
</tr>
<tr>
<td>Back</td>
<td>Chhabra et al [31]</td>
<td>Chronic low back pain</td>
<td>+</td>
<td>Acceptable</td>
<td>Strong\textsuperscript{b}</td>
</tr>
<tr>
<td>Back</td>
<td>Fatoye et al [24]</td>
<td>Chronic low back pain</td>
<td>=\textsuperscript{c}</td>
<td>High</td>
<td>Strong\textsuperscript{b}</td>
</tr>
<tr>
<td>Back</td>
<td>Ozden et al [29]</td>
<td>Chronic low back pain</td>
<td>+</td>
<td>High</td>
<td>Strong\textsuperscript{b}</td>
</tr>
<tr>
<td>Shoulder</td>
<td>Choi et al [23]</td>
<td>Frozen shoulder</td>
<td>=</td>
<td>High</td>
<td>Moderate</td>
</tr>
<tr>
<td>Hip</td>
<td>Nelson et al [28]</td>
<td>Total hip arthroplasty</td>
<td>=</td>
<td>High</td>
<td>Moderate</td>
</tr>
<tr>
<td>Full body</td>
<td>Hernando-Garrijo et al [33]</td>
<td>Fibromyalgia</td>
<td>=</td>
<td>Acceptable</td>
<td>Limited</td>
</tr>
<tr>
<td>Knee</td>
<td>Allen et al [21]</td>
<td>Knee osteoarthritis</td>
<td>=</td>
<td>High</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
<tr>
<td>Knee</td>
<td>Bennell et al [30]</td>
<td>Chronic knee pain</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
<tr>
<td>Knee</td>
<td>Fleischman et al [25]</td>
<td>Total knee arthroplasty</td>
<td>=</td>
<td>High</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
<tr>
<td>Knee</td>
<td>Hardt et al [32]</td>
<td>Total knee arthroplasty</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
<tr>
<td>Knee</td>
<td>Moffet et al [26]</td>
<td>Total knee arthroplasty</td>
<td>=</td>
<td>High</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
<tr>
<td>Knee</td>
<td>Nelligan et al [27]</td>
<td>Knee osteoarthritis</td>
<td>+</td>
<td>High</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
<tr>
<td>Knee</td>
<td>Tousignant et al [35]</td>
<td>Total knee arthroplasty</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
<tr>
<td>Hand</td>
<td>Blanquerro et al [22]</td>
<td>Carpal tunnel release</td>
<td>+</td>
<td>High</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
<tr>
<td>Hand</td>
<td>Rodríguez Sánchez–Laulhé et al [34]</td>
<td>Rheumatoid arthritis</td>
<td>=</td>
<td>Acceptable</td>
<td>Conflicting\textsuperscript{b}</td>
</tr>
</tbody>
</table>

\textsuperscript{a}50% of the outcomes were significantly better in the intervention group than in the control group.

\textsuperscript{b}The level of evidence was determined from all studies in the same localization.

\textsuperscript{c}No statistically significant difference between the intervention and control groups.
<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Study</th>
<th>Assessment tools</th>
<th>Results</th>
<th>Study quality</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disability</td>
<td>Abadiyan et al [20]</td>
<td>Neck Disability Index</td>
<td>+(^a)</td>
<td>High</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td></td>
<td>Blanquero et al [22]</td>
<td>Disabilities of Arm, Shoulder and Hand Questionnaire</td>
<td>+</td>
<td>High</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Disability</td>
<td>Chhabra et al [31]</td>
<td>Modified Oswestry Disability Index</td>
<td>+</td>
<td>Acceptable</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Disability</td>
<td>Fatoye et al [24]</td>
<td>Oswestry Disability Index</td>
<td>=(^c)</td>
<td>High</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Disability</td>
<td>Özden et al [29]</td>
<td>Oswestry Disability Index</td>
<td>+</td>
<td>High</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Disability</td>
<td>Rodríguez Sánchez-Laulhé et al [34]</td>
<td>Disabilities of Arm, Shoulder and Hand Questionnaire</td>
<td>=</td>
<td>Acceptable</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Quality of life</td>
<td>Abadiyan et al [20]</td>
<td>Short Form Health 36 Questionnaire</td>
<td>+</td>
<td>High</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Quality of life</td>
<td>Bennell et al [30]</td>
<td>Assessment of Quality of Life–2</td>
<td>+</td>
<td>Acceptable</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Quality of life</td>
<td>Nelligan et al [27]</td>
<td>Assessment of Quality of Life-6D</td>
<td>+</td>
<td>High</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Quality of life</td>
<td>Nelson et al [28]</td>
<td>Short Form Health 12 Questionnaire/European Quality of Life 5 Dimensions 5 Level Version</td>
<td>=</td>
<td>High</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Quality of life</td>
<td>Özden et al [29]</td>
<td>Short Form Health 36 Questionnaire</td>
<td>+</td>
<td>High</td>
<td>Strong(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Abadiyan et al [20]</td>
<td>Visual analog scale</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Allen et al [21]</td>
<td>WOMAC(^d)</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Bennell et al [30]</td>
<td>Numeric rating scale</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Blanquero et al [22]</td>
<td>Visual analog scale</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Chhabra et al [31]</td>
<td>Numeric rating scale, Current Symptom Score</td>
<td>=</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Choi et al [23]</td>
<td>Visual analog scale</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Fleischman et al [25]</td>
<td>KOOS(^e)</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Hardt et al [32]</td>
<td>Numeric rating scale</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Hernando-Garijo et al [33]</td>
<td>Visual analog scale</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Moffet et al [26]</td>
<td>WOMAC</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Nelligan et al [27]</td>
<td>Numeric rating scale</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Özden et al [29]</td>
<td>Visual analog scale</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Rodríguez Sánchez-Laulhé et al [34]</td>
<td>Visual analog scale</td>
<td>=</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Pain</td>
<td>Tousignant et al [35]</td>
<td>WOMAC</td>
<td>_(^f)</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Allen et al [21]</td>
<td>WOMAC/30-s chair stand test/Timed up and go test/2-min step test, single-leg stand</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Bennell et al [30]</td>
<td>WOMAC</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Choi et al [23]</td>
<td>Range of motion</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Fleischman et al [25]</td>
<td>KOOS</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Hardt et al [32]</td>
<td>Range of motion/Timed up and go test/10-m walk test/30-s chair stand test/Knee Society Score</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Hernando-Garijo et al [33]</td>
<td>Arm curl test, 6-min walk test</td>
<td>=</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Moffet et al [26]</td>
<td>KOOS/Stair test/6-min walk test</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Study</td>
<td>Assessment tools</td>
<td>Results</td>
<td>Study quality</td>
<td>Evidence</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>------------------</td>
<td>---------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td>Function</td>
<td>Nelligan et al [27]</td>
<td>WOMAC, KOOS</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Nelson et al [28]</td>
<td>Timed up and go test</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Özden et al [29]</td>
<td>Timed up and go test</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Rodríguez Sánchez-Laulhé et al [34]</td>
<td>Michigan Hand Outcome Questionnaire</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Function</td>
<td>Tousignant et al [35]</td>
<td>WOMAC/Timed up and go test/Functional Autonomy Measurement System</td>
<td>−</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
</tbody>
</table>

\(^a\)50% of the outcomes were significantly better in the intervention group than in the control group.

\(^b\)The level of evidence was determined from all studies in the same outcomes.

\(^c\)No statistically significant differences between the intervention and control groups.

\(^d\)WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

\(^e\)KOOS: Knee Injury and Osteoarthritis Outcome Score.

\(^f\)>50% of the outcomes were significantly better in the control group than in the intervention group.

Table 7. Best-evidence synthesis for the medical treatment types.

<table>
<thead>
<tr>
<th>Therapy</th>
<th>Study</th>
<th>Musculoskeletal disease</th>
<th>Results</th>
<th>Study quality</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operative</td>
<td>Blanquero et al [22]</td>
<td>Carpal tunnel release</td>
<td>+(^a)</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Operative</td>
<td>Fleischman et al [25]</td>
<td>Total knee arthroplasty</td>
<td>=(^c)</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Operative</td>
<td>Hardt et al [32]</td>
<td>Total knee arthroplasty</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Operative</td>
<td>Moffet et al [26]</td>
<td>Total knee arthroplasty</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Operative</td>
<td>Nelson et al [28]</td>
<td>Total hip arthroplasty</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Operative</td>
<td>Tousignant et al [35]</td>
<td>Total knee arthroplasty</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Abadiyan et al [20]</td>
<td>Chronic neck pain</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Allen et al [21]</td>
<td>Knee osteoarthritis</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Bennell et al [30]</td>
<td>Chronic knee pain</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Chhabra et al [31]</td>
<td>Chronic low back pain</td>
<td>+</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Choi et al [23]</td>
<td>Frozen shoulder</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Fatoye et al [24]</td>
<td>Chronic low back pain</td>
<td>=</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Hernando-Garijo et al [33]</td>
<td>Fibromyalgia</td>
<td>=</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Nelligan et al [27]</td>
<td>Knee osteoarthritis</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Özdén et al [29]</td>
<td>Chronic low back pain</td>
<td>+</td>
<td>High</td>
<td>Conflicting(^b)</td>
</tr>
<tr>
<td>Conservative</td>
<td>Rodríguez Sánchez-Laulhé et al [34]</td>
<td>Rheumatoid Arthritis</td>
<td>=</td>
<td>Acceptable</td>
<td>Conflicting(^b)</td>
</tr>
</tbody>
</table>

\(^a\)50% of the outcomes were significantly better in the intervention group than in the control group.

\(^b\)The level of evidence was determined from all studies in the same therapy.

\(^c\)No statistically significant difference between the intervention and control groups.
Discussion

Principal Findings

This systematic review aimed to evaluate the impact of digital physical health exercises on patients with musculoskeletal diseases concerning the localization of the musculoskeletal disease, patient-reported outcomes, and medical treatment types. In addition, a best-evidence synthesis was conducted to estimate the direction and strength of the existing evidence. The main findings were that (1) strong evidence was found for a positive impact on musculoskeletal diseases located in the back and on the patient-reported outcomes of disability and quality of life (Figure 2). Our findings are partly supported by a previous systematic review with a meta-analysis [13] showing moderate-quality evidence for the positive impact on the patient-reported outcome of disability. In contrast to the previous review [13] and to another systematic review [12], conflicting evidence for the patient-reported outcomes of pain and function was found. It should be noted that 1 study [12] found some clinical benefits for pain and function but did not conduct an evidence synthesis or a meta-analysis. In addition, the outcomes of pain and function represent health-related outcomes, and the outcomes of disability and quality of life are the resulting consequences. Therefore, pain acts as a protective mechanism and can lead to disability [50]. With appropriate exercises,
patients learn to compensate for their disabilities [20,30], whereas exercise alone can provoke pain [51]. As disabilities are part of the concept of health-related quality of life [52], these outcomes are mutually dependent, and identical strong evidence is plausible.

In addition, it should be mentioned that both previous systematic reviews included all types of digital health interventions, and we explicitly focused our systematic review on the impact of digital physical health exercises. Regarding this, our findings add that this type of intervention shows strong evidence to have an overall positive impact on the musculoskeletal diseases located in the back, independent of the investigated outcomes [20,24,29,31]. Back-related musculoskeletal diseases usually arise because of muscular causes and are often caused by a lack of physical activity [53]. Participants recruited in back pain–related studies are often middle-aged and have an office occupation [20]. The use of digital physical health exercises in such patients can be considered highly effective because of the increased physical activity targeting muscle strengthening and the teaching of exercise techniques [20,24,29,31]. Overall, the application of digital physical health exercise in patients with musculoskeletal diseases shows versatile positive effects, especially for musculoskeletal diseases located in the back and for the improvement of disabilities and quality of life. However, the type of digital health interventions seems to influence the effects on the specific patient-reported outcome, and more studies to investigate this relationship are needed.

The second main finding was that moderate evidence was found for a beneficial effect on musculoskeletal diseases of the shoulder and hip (Figure 2). As this systematic review is the first to evaluate the association between digital physical health exercises and different localizations of musculoskeletal diseases, no evidence levels from previous research is available for clarification. There is only 1 other systematic review on the effectiveness of digital health interventions for total hip arthroplasty [54]. The review found no significant improvements in the studied patient-reported outcomes. For the shoulder, another systematic review examined the effectiveness of telerehabilitation for musculoskeletal diseases compared with normal in-person physiotherapy [55] and found very low to low evidence. In this context, our findings suggest that digital physical health exercises may also be effective in treating musculoskeletal diseases of the shoulder and hip. However, it should be noted that only 1 study each was found for shoulder- and hip-specific musculoskeletal diseases, whereas several studies were found for back- or knee-specific musculoskeletal diseases (Table 5). Therefore, our results must be interpreted with caution, as a small number of high-quality studies may result in stronger evidence, according to the definitions of the best-evidence synthesis [18], than the presence of many lower-quality studies. Overall, the results demonstrated that digital physical health exercises had a positive effect on a variety of health-related outcomes, regardless of the localization of the musculoskeletal diseases. However, the number of studies investigating the relationship between the effectiveness of digital health interventions and the localization of musculoskeletal diseases is small, and more studies are needed, especially for localizations other than the knee and back.

An additional interesting finding is the conflicting evidence in the medical treatment types concerning operative and conservative approaches (Figure 2). Although the underlying reasons remain unknown, it can be stated that the operative treatment (ie, carpal tunnel release and total knee arthroplasty) has no impact on the overall stimulus-response mechanism of the digital intervention, requiring further studies for clarification.

**Limitations**

Although this systematic review increases knowledge of the positive impacts of digital physical health exercises on musculoskeletal diseases, there are a few limitations. Because of the heterogeneity of the included studies (eg, different numbers of patients, interventions, body regions, and control groups), a meta-analysis could not be performed. Instead, and as an established alternative approach, a best-evidence synthesis [18] was used. A strength of this approach is that it is possible to estimate an evidence level for various categories despite the large study heterogeneity. However, a limitation is that no quantitative analysis (eg, in terms of statistical significance) can be conducted [56]. An additional limitation of our review is that we did not register the study plan in PROSPERO a priori. The reason is that according to the PRISMA guidelines, registration is currently recommended but not mandatory [15]. Furthermore, all studies that included digital health interventions beyond active exercises were not included. Therefore, some studies could be lost, but the aspect of physical exercise as an established clinical treatment for musculoskeletal diseases could be focused on for the first time.

**Conclusions**

There is strong to moderate evidence for the beneficial impact of digital physical health exercises for musculoskeletal diseases located in the back, shoulder, and hip. There is limited or conflicting evidence for other localizations. In addition, strong evidence was found for the patient-reported outcomes of disability and quality of life, whereas conflicting evidence exists for other commonly investigated patient-reported outcomes such as pain and function. Thus, digital physical health exercises could have a positive effect on a variety of health-related outcomes of musculoskeletal diseases. To implement digital physical health exercises in evidence-based medicine for musculoskeletal diseases, more high-quality randomized controlled trials are needed to clarify the relationship between the impact of digital physical health exercises and clinically relevant factors such as localization, patient-reported outcomes, and medical treatment types.
Acknowledgments

The study was funded by the Open Access Publishing Fund of Leipzig University supported by the German Research Foundation within the Open Access Publication Funding program. The authors acknowledge the support from Leipzig University for Open Access Publishing.

Authors' Contributions

JN, FW, and MWH conceived the study and were involved in the conceptual process. JN performed the literature search and data analysis, and MWH validated these steps. JN wrote a raw version of the manuscript, and FW and MWH revised the raw version and wrote the final manuscript together with JN. CG critically revised the final manuscript and made additional changes. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA 2020 Checklist.

References

5. Musculoskeletal health status in Europe - summery report. v5.0. eumusc.net. URL: http://www.eumusc.net/workpackages_wp4.cfm [accessed 2023-07-03]


34. Rodríguez-Sánchez-Laulhé P, Luque-Romero LG, Barrero-García FJ, Biscarri-Carbonero Á, Blanquero J, Suero-Pineda A, et al. An exercise and educational self-management program delivered with a smartphone app (CareHand) in adults...
with rheumatoid arthritis of the hands: randomized controlled trial. JMIR Mhealth Uhealth 2022 Apr 07;10(4):e35462 [FREE Full text] [doi: 10.2196/35462] [Medline: 35389367]


Abbreviations

PICOS: Population Intervention Comparison Outcome Study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Abstract

Background: Successful post-knee replacement rehabilitation requires adequate access to health information, social support, and periodic monitoring by a health professional. Mobile health (mHealth) and computer-based technologies are used for rehabilitation and remote monitoring. The extent of technology use and its function in post-knee replacement rehabilitation care in low and middle-income settings is unknown.

Objective: To inform future mHealth intervention development, we conducted a scoping review to map the features and functionality of existing technologies and determine users’ perspectives on telerehabilitation and technology for self-management.

Methods: We followed the Joanna Briggs Institute methodology for scoping reviews. We searched the Embase, Medline, PsycINFO via OVID, and Cochrane Central Register of Controlled Trials databases for manuscripts published from 2001 onward. We included original research articles reporting the use of mobile or computer-based technologies by patients, health care providers, researchers, or family members. Studies were divided into the following 3 categories based on the purpose: validation studies, clinical evaluation, and end user feedback. We extracted general information on study design, technology features, proposed function, and perspectives of health care providers and patients. The protocol for this review is accessible in the Open Science Framework.

Results: Of the 5960 articles, 158 that reported from high-income settings contributed to the qualitative summary (64 studies on mHealth or telerehabilitation programs, 28 validation studies, 38 studies describing users’ perceptions). The highest numbers of studies were from Europe or the United Kingdom and North America regarding the use of a mobile app with or without wearables and reported mainly in the last decade. No studies were from low and middle-income settings. The primary functions of technology for remote rehabilitation were education to aid recovery and enable regular, appropriate exercises; monitoring progress of pain (n=19), activity (n=20), and exercise adherence (n=30); 1 or 2-way communication with health care professionals to facilitate the continuum of care (n=51); and goal setting (n=23). Assessment of range of motion (n=16) and gait analysis (n=10) were the commonly validated technologies developed to incorporate into a future rehabilitation program. Few studies (n=14) reported end user involvement during the development stage. We summarized the reasons for satisfaction and dissatisfaction among users across various technologies.

Conclusions: Several existing mobile and computer-based technologies facilitate post-knee replacement rehabilitation care for patients and health care providers. However, they are limited to high-income settings and may not be extrapolated to low-income settings. A systematic needs assessment of patients undergoing knee replacement and health care providers involved in rehabilitation,
To our knowledge, there are no existing scoping reviews that address our aims [19]. The information from this review will help us and other researchers make an informed decision on future mobile health (mHealth) interventions for monitoring post-knee arthroplasty rehabilitation care by physiotherapists and orthopedic surgeons and to promote self-management by individuals. This review will also help highlight existing gaps in the context of low and middle-income countries (LMICs).

Methods

We conducted this scoping review following the Joanna Briggs Institute (JBI) methodology for JBI Scoping Reviews [20] and consulted the PRISMA-ScR (Preferred Reporting Items for Systematic Review and Meta-Analyses Extension for Scoping Reviews) checklist for reporting [21]. The protocol was registered at the Open Science Framework [22].

Data Sources and Searches

To identify relevant studies, an electronic database literature search was conducted in the Embase, Medline, PsycINFO via OVID, and Cochrane Central Register of Controlled Trials (CENTRAL) databases using the following key terms: “Knee arthroplasty OR Knee replacement,” “mobile,” “web,” “remote sensor,” “computer,” “telerehabilitation,” and “m-health” (Tables S1 and S2 in Multimedia Appendix 1). The search was executed in October 2021 and updated in August 2023. The search was restricted to 2001 onward. There were no language restrictions during the search. We searched the reference list of included articles to identify potentially eligible studies.

Study Selection

Predefined inclusion criteria were articles reporting the use of mobile or computer apps or any other technologies such as sensor-based devices for delivering or monitoring rehabilitation either scheduled or following knee joint replacement. We also included proof-of-concept papers that described the development process of a mobile or technology-based solution for rehabilitation. The purpose of technology could be for a health care provider to monitor rehabilitation adherence, to aid patient-health care provider communication, to promote self-management, to act as reminders, or to act as a source of education or any other function that is aimed at rehabilitation care following knee replacement. The app or technology could be used by patients, health care providers, researchers, or a family member. Included studies could have been conducted in the community or home for any clinical setting in any geographic region. The studies were required to be original research articles, and we included experimental and observational studies using quantitative or qualitative research methods. Reviews (narrative or systematic reviews), non-English

Introduction

Knee arthroplasty is the gold standard treatment for end-stage osteoarthritis when conservative treatments fail to relieve symptoms [1]. Wound care and postarthroplasty physiotherapy are essential components of this treatment. Poor adherence to physiotherapy could delay the recovery and lead to suboptimal functional outcomes [2]. Beyond in-hospital clinical care and initiation of physical therapy before discharge, continued and reliable access to information, support from health care providers, awareness of the recovery pathway, easy access to rehabilitation centers, and periodic monitoring are influential factors for optimal recovery [3-6]. In addition to an uneventful surgery, postarthroplasty outcomes are associated with several patient-related factors such as their preoperative physical and mental state, comorbidities, social support, and socioeconomic status, emphasizing the need for personalized approaches [7]. Hence, monitoring of the rehabilitation phase is essential, whether at clinics, in rehabilitation units, or at home [8-10].

Technology-assisted remote monitoring methods are increasingly being advocated in high-income countries. There is low to moderate-quality evidence on the superiority of telerehabilitation compared with unsupported home-based rehabilitation and noninferiority compared with clinic-based monitoring with respect to range of motion (ROM), pain, function, quality of life, and cost-effectiveness at 3 months between clinic-based and home-based rehabilitation strategies using technology [11-17]. Hence, current evidence supports the adaptation of technology-based rehabilitation as feasible, as safe, and as good as clinic-based monitoring with an additional benefit of saving out-of-pocket expenditure. Technology-based approaches are diverse, varying from telehealth [17] to virtual reality techniques [13] aimed at improving adherence to physical therapy and facilitating remote monitoring [12] of patient progress during the post-acute rehabilitation phase [18].

Therefore, the aim of this scoping review was to summarize the extent, range, and nature of technology used for provision of rehabilitation or to monitor progress following knee arthroplasty. This scoping review aimed to address the following objectives:

1. To map the characteristic features and functionality of the technologies, guiding or theoretical framework for designing the technology, and evaluation methodologies of mobile technology–based apps for rehabilitation monitoring and self-management following knee arthroplasty
2. To understand the patient and physical therapist perspectives regarding the use of mobile technology–based apps for rehabilitation monitoring and self-management following knee arthroplasty

 KEYWORDS

knee arthroplasty; telerehabilitation; mHealth; rehabilitation; monitoring; self-management; knee; arthroplasty; social support; mHealth intervention; development; scoping review; knee replacement
articles, and articles without abstracts or full texts were excluded.

Data Extraction
Screening of manuscript titles and abstracts was conducted by 2 independent reviewers using the web app Rayyan [23]. Prior to screening, reviewers discussed inclusion and exclusion criteria to ensure consistency between individuals. Two reviewers assessed the eligibility of the full text, and disagreements were resolved by discussion. Systematic reviews were not included in the review but were used to obtain potentially relevant references. Multiple publications originating from a single technology were grouped and presented as 1 study.

For data charting purposes, the studies were divided into the following 3 categories: (1) studies that had no rehabilitation program but included an app or a technology to assess ROM or gait and were validation studies, (2) studies reporting the use of a mobile or computer app or a telehealth delivery platform for a rehabilitation program with or without sensor-based devices and wearable sensors, (3) studies that reported end users’ perceptions of the technology used for rehabilitation monitoring. Data on the general information for the studies, features of the technology, the proposed function, and perspectives of health care providers and patients were extracted and entered in Microsoft Excel. If only the protocol of a planned study was available, there was no information on clinical evaluation, or the study included <6 individuals, we did not extract data beyond the general information.

Results
Search Results
The database search, including the ad hoc search, yielded 5960 articles. Of these articles, 300 articles were considered potentially relevant. Of these, 158 articles were included for data extraction, 131 articles were excluded, and 11 articles were not available (Table S3 in Multimedia Appendix 1). Of the 158 articles, 91 articles (64 studies) reported the clinical evaluation of a technology-based rehabilitation program, 29 articles (28 studies) reported the validation or a proof of concept of technology intended to be used for rehabilitation, and 13 articles were protocols of evaluation studies. In addition, 25 articles reported end users’ perceptions on technology (Figure 1) as stand-alone articles or part of clinical evaluation studies (n=13), totaling 38 studies. The 13 studies that reported the perceptions of technology that were also included in rehabilitation program studies were removed from the final list of included full-text articles to avoid double counting.

Figure 1. Process of identifying and including studies according to PRISMA-ScR (Preferred Reporting Items for Systematic Review and Meta-Analyses Extension for Scoping Reviews).
Technology for Rehabilitation

Characteristics of the 105 Studies

Studies were reported from Europe and the United Kingdom (n=45) [11, 24-60, 62-66, 169, 170], North America (n=34) [6, 67-99], Australia and New Zealand (n=10) [100-109], and Asia (n=16) [110-125]. None of the studies were from LMICs. Reports of mobile-based technologies represented the highest number (54/105, 51.4%) [6, 25-27, 31, 32, 36, 37, 41-50, 52, 55, 56, 58-60, 62, 64, 67-69, 74, 75, 78, 80, 82, 83, 90-92, 95, 101, 103, 106-108, 112, 113, 117, 121-126, 169], followed by computer applications (31/105, 29.5%) [24, 29, 30, 33-35, 39, 53, 54, 57, 65, 66, 70, 73, 76, 79, 84-86, 89, 93, 97, 98, 100, 102, 104, 111, 114, 116, 120, 127], and tele/video/web conferencing (20/105, 19%) for rehabilitation monitoring [28, 38, 51, 63, 71, 72, 77, 81, 87, 88, 94, 96, 99, 105, 109, 110, 115, 118, 119, 170]. The highest use of mobile apps associated with or without a wearable was in Europe and the United Kingdom, followed by North America. Tele/video/web conferencing was used across regions, with the highest number in North America (Figure 2).

Figure 2. Number of studies published by region based on different technologies (n=105).

Validation Studies

There were 28 validation studies. Studies that validated stand-alone technologies included those to assess ROM (n=16) [24, 29, 43, 45, 48, 49, 52, 68, 73, 84, 100, 101, 111, 112, 116, 125] or gait or posture (n=10) [29, 30, 33, 53, 57, 74, 89, 97, 102, 124], and 2 studies involved technologies to monitor exercises [98, 114]. The technologies involved were either wearables (n=20) [24, 29, 30, 33, 45, 48, 49, 57, 68, 73, 84, 89, 97, 98, 100, 111, 112, 114, 116, 125], sensor-based devices (nonwearables; n=4) [53, 66, 102, 124], or inbuilt sensors available within a smartphone (n=4) [43, 52, 74, 101] (Table S4 in Multimedia Appendix 1).

In terms of study design, 9 were cross-sectional studies [33, 48, 52, 57, 84, 89, 97, 101, 116], 7 were cohort or longitudinal studies [45, 53, 68, 74, 100, 111, 125], 5 were pre-post studies [29, 30, 43, 73, 102], 1 was an uncontrolled trial [112], 1 was a randomized controlled trial (RCT) [66], and 5 were articles that described the proof of concept or development plan for the technologies [24, 49, 98, 114, 124]. The participant sample size ranged from 1 to 60. Most studies reported reliability between a standard or universal goniometer and smartphone app goniometry and the clinical evaluation of sensors to measure gait parameters (Table S4 in Multimedia Appendix 1). In 7 studies, gait was measured using sensors provided by a health care provider in a hospital setting [29, 33, 57, 74, 89, 97, 102], and 3 studies did not describe the measurement setting [30, 53, 124].

Clinical Evaluation Studies

There were 64 clinical evaluation studies. The technology consisted of a mobile or computer app with a wearable device (n=18) [6, 26, 31, 32, 39, 44, 46, 50, 54, 64, 67, 69, 90, 92, 95, 106, 108, 169], a mobile or computer app with a sensor-based device (n=13) [25, 34, 35, 40, 42, 65, 70, 76, 79, 85, 86, 93, 120], only a mobile app (n=14) [36, 37, 55, 56, 62, 75, 78, 80, 83, 107, 113, 117, 123, 128], or only telephone or videoconferencing (n=19) for remote monitoring [28, 38, 51, 63, 71, 72, 77, 81, 88, 94, 96, 99, 105, 109, 110, 115, 118, 119, 170]. Of the studies that used a mobile app, 9 studies were developed only for iOS [55, 67, 69, 71, 77, 92, 106, 107, 109], 1 was an Android app [42], 7 were for both Android and iOS devices [28, 36, 56, 88, 108, 115, 117], and 21 studies did not specify the platform (Multimedia Appendix 2). A web-based clinician portal for synchronous or asynchronous remote monitoring of patients was reported by 36 studies (Table 1). The number of published studies and the intervention arm sample size (ranging from 7 to 2292), especially for those that included wearable sensors and mobile apps, steadily increased over the last 2 decades (Figure 3).
<table>
<thead>
<tr>
<th>First author, year</th>
<th>Web portal</th>
<th>Devices</th>
<th>Sensor-based devices</th>
<th>Peer</th>
<th>App name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexander, 2023 [67]</td>
<td>✓</td>
<td>Apple Watch</td>
<td>—a</td>
<td>—</td>
<td>mymobility</td>
</tr>
<tr>
<td>An, 2021 [110]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Argent, 2019 [169]</td>
<td>—</td>
<td>IMUb</td>
<td>Avatar</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Bade, 2020 [166]</td>
<td>—</td>
<td>In-shoe sensors</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Bell, 2020 [90]</td>
<td>✓</td>
<td>InterACTION IMU</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Bini, 2017 [71]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Capture proof</td>
</tr>
<tr>
<td>Blasco, 2022 [28]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>WeChat app</td>
</tr>
<tr>
<td>Campbell, 2019 [72]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>StreaMD</td>
</tr>
<tr>
<td>Chughtai, 2018 [76]</td>
<td>✓</td>
<td>—</td>
<td>VERAc</td>
<td>—</td>
<td>VERA</td>
</tr>
<tr>
<td>Chughtai, 2019 [75]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>✓</td>
<td>PReHab</td>
</tr>
<tr>
<td>Colomina, 2021 [31]</td>
<td>✓</td>
<td>Fitbit Flex 2</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>De Berardinis, 2022 [26]</td>
<td>✓</td>
<td>Magnetic sensors with Velcro bands</td>
<td>—</td>
<td>—</td>
<td>kari</td>
</tr>
<tr>
<td>Doiron-Cadrin, 2020 [77]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Reacts Lite</td>
</tr>
<tr>
<td>Duong, 2023 [106]</td>
<td>✓</td>
<td>Fitbit, ActivPal, Goniometer Pro</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Eichler, 2019 [34]</td>
<td>✓</td>
<td>Kinect sensor</td>
<td>—</td>
<td>—</td>
<td>MainReha app</td>
</tr>
<tr>
<td>Fung, 2012 [79]</td>
<td>—</td>
<td>—</td>
<td>Wii sensor balance</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Gohir, 2021 [36]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>i-Beat app</td>
</tr>
<tr>
<td>Gray, 2022 [37]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Digital Joint School using GoWell health program</td>
</tr>
<tr>
<td>Gunduz, 2021 [38]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Hadamus, 2022 [40]</td>
<td>—</td>
<td>—</td>
<td>Kinetic camera</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Hardwick-Morris, 2022 [107]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Physitrack</td>
</tr>
<tr>
<td>Hong, 2022 [80]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Digital Musculoskeletal Surgical Care Program app</td>
</tr>
<tr>
<td>Huang, 2017 [113]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yishu</td>
</tr>
<tr>
<td>Janhunen, 2023 [42]</td>
<td>—</td>
<td>—</td>
<td>Kinect sensor with TV and tablet</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Juhl, 2016 [44]</td>
<td>✓</td>
<td>IMU</td>
<td>—</td>
<td>—</td>
<td>ICURA app</td>
</tr>
<tr>
<td>Klement, 2019 [81]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Knapp, 2021 [83]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Kramer, 2003 [99]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>First author, year</td>
<td>Web portal Monitoring</td>
<td>Devices Wearables</td>
<td>Sensor-based devices</td>
<td>Peer</td>
<td>App name</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>Kuether, 2019 [85]</td>
<td>✓</td>
<td>—</td>
<td>VERA</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lam, 2016 [86]</td>
<td>✓</td>
<td>IMU</td>
<td>—</td>
<td>ReHab system</td>
<td></td>
</tr>
<tr>
<td>Lebleu, 2023 [46]</td>
<td>✓</td>
<td>Activity tracker</td>
<td>Garmin vivofit 4</td>
<td>—</td>
<td>moveUP Therapy</td>
</tr>
<tr>
<td>LeBrun, 2022 [78]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>MyChart app</td>
</tr>
<tr>
<td>Li, 2023 [115]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Lu, 2021 [117]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>McDonall, 2022 [147]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Milliren, 2022 [88]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Ubicare Smart X</td>
</tr>
<tr>
<td>Nuevo, 2023 [50]</td>
<td>✓</td>
<td>Accelerometer, gyroscope, magnetometer (DyCare)</td>
<td>—</td>
<td>—</td>
<td>ReHub</td>
</tr>
<tr>
<td>Osterloh, 2023 [51]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>✓</td>
<td>YOLii</td>
</tr>
<tr>
<td>Park, 2017 [118]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Park, 2023 [119]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Piqueras, 2013 [54]</td>
<td>✓</td>
<td>(WAGYRO)</td>
<td>Avatar</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Pournajaf, 2022 [65]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Pronk, 2020 [55]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Pain coach app</td>
</tr>
<tr>
<td>Prvu Bettger, 2019 [70]</td>
<td>✓</td>
<td>—</td>
<td>VERA</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ramkumar, 2019 [92]</td>
<td>✓</td>
<td>Motion sensors</td>
<td>—</td>
<td>—</td>
<td>Focus ventures RPM</td>
</tr>
<tr>
<td>Russell, 2011 [105]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Scheper, 2019 [56]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Woundcare app</td>
</tr>
<tr>
<td>Su, 2015 [120]</td>
<td>—</td>
<td>Kinect sensor</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Summers, 2023 [93]</td>
<td>✓</td>
<td>—</td>
<td>Electro-mechanical device</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Szöts, 2016 [170]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Timmers, 2019 [62]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>The Patient Journey app</td>
</tr>
<tr>
<td>Torpil, 2022 [63]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Toussignant, 2011 [94]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Tripuraneni, 2021 [95]</td>
<td>✓</td>
<td>Smart watch</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>van Dijk-Huisman, 2020 [64]</td>
<td>✓</td>
<td>MOX activity monitor</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Visperas, 2021 [96]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wang, 2023 [121]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>✓</td>
<td>WeChat app</td>
</tr>
<tr>
<td>Zhang, 2021 [123]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>✓</td>
<td>WeChat app</td>
</tr>
</tbody>
</table>

*aNot applicable.*

*bIMU: inertial motion unit.*

*cVERA: Virtual Exercise Rehabilitation Assistant.*
Although most studies described the features and functionality of the technology to deliver the intervention, they lacked details about the technological aspects that could benefit future researchers. For example, 2 studies [31, 86] explicitly reported information on the software, programming language and tools used, or calibration procedures either along with the main study or cited the article that described the development phase. Information on conceptualization of the technology-based intervention was described in only 3 studies [37, 38, 120]. End users’ involvement was typically late during the development phase (ie, prototype stage) and involved refining the functionalities and features of technology [32, 51, 64, 65, 75-77, 80, 86, 88, 115, 118, 169] prior to deployment. Patient feedback on their needs at an early development phase was reported only by Blasco et al [28].

Clinical effectiveness was tested using an RCT design in 57.8% (37/64) of the studies [6, 25, 28, 34-36, 39, 42, 44, 45, 50, 51, 54, 55, 62, 63, 65, 67, 70-72, 77, 79, 90, 94-96, 99, 105, 106, 108-110, 117-119, 128, 170], and the rest of the studies were either retrospective comparative cohort studies (n=3) [26, 37, 78], uncontrolled cohort studies (n=9) [46, 56, 69, 76, 81, 83, 85, 92, 169], cross-sectional studies (n=1) [86], or non-RCTs (n=14) [31, 32, 38, 40, 64, 75, 80, 88, 93, 107, 113, 115, 120, 123].

We found 13 study protocols, of which 12 were RCTs published between 2013 and 2023 [11, 41, 47, 58-60, 82, 87, 91, 103, 104, 122], for which we could not find a published report and hence were not included in this summary. User experience was measured in trials using quantitative (n=9) [32, 34, 38, 50, 55, 94, 96, 105, 123], qualitative (n=2) [61, 128], and mixed methods (n=3) [39, 90, 109] approaches.

Application Functionality for Rehabilitation Programs

The key functionalities of the telerehabilitation technologies extracted from 64 studies are summarized under 4 themes, namely education and enablement, monitoring progress, communication, and goal setting (Table 2).
Table 2. Themes of the key functionalities of the telerehabilitation technologies.

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Exercise</th>
<th>Monitoring progress</th>
<th>Functions</th>
<th>Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repository, Diary</td>
<td>Tracker or reminder</td>
<td>Biofeedback</td>
<td>VR</td>
<td>Feedback to patient</td>
</tr>
<tr>
<td>Alexander, 2023 [67]</td>
<td>✓</td>
<td><em>c</em></td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>An, 2021 [110]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Argent, 2019 [169]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bade, 2020 [166]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bell, 2020 [90]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bini, 2017 [71]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Campbell, 2019 [72]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Chughtai, 2018 [76]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Chughtai, 2019 [75]</td>
<td>✓</td>
<td>✓</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>De Berardinis, 2022 [26]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Doiron-Cadrin, 2019 [77]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Duong, 2023 [106]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>—</td>
</tr>
</tbody>
</table>

https://mhealth.jmir.org/2024/1/e47843
<table>
<thead>
<tr>
<th>First author, year</th>
<th>Exercise</th>
<th>Monitoring progress</th>
<th>Functions</th>
<th>Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eichler, 2019 [34]</td>
<td>✓</td>
<td>SA, AP</td>
<td></td>
<td>Exercise 1-way, 2-way</td>
</tr>
<tr>
<td>Farr-Wharton, 2020 [108]</td>
<td>✓</td>
<td>AA, AP</td>
<td></td>
<td>DS</td>
</tr>
<tr>
<td>Fung, 2012 [79]</td>
<td>✓</td>
<td>SA</td>
<td></td>
<td>Lower extremity function</td>
</tr>
<tr>
<td>Gianola, 2020 [35]</td>
<td>✓</td>
<td>SA</td>
<td></td>
<td>Exercise</td>
</tr>
<tr>
<td>Gohir, 2021 [36]</td>
<td>✓</td>
<td>AA, AP</td>
<td></td>
<td>Exercise 1-way, 2-way</td>
</tr>
<tr>
<td>Gray, 2022 [37]</td>
<td>✓</td>
<td>SP</td>
<td></td>
<td>Exercise 2-way</td>
</tr>
<tr>
<td>Gunduz, 2021 [38]</td>
<td>✓</td>
<td>SA, SP</td>
<td></td>
<td>Exercise 2-way</td>
</tr>
<tr>
<td>Hadanus, 2022 [40]</td>
<td>✓</td>
<td>SA, SP</td>
<td></td>
<td>Exercise 2-way</td>
</tr>
<tr>
<td>Hardwick-Morris, 2022 [107]</td>
<td>✓</td>
<td>SP</td>
<td></td>
<td>Exercise 2-way</td>
</tr>
<tr>
<td>Hong, 2022 [80]</td>
<td>✓</td>
<td>SP</td>
<td></td>
<td>Recovery goals 2-way</td>
</tr>
<tr>
<td>Huang, 2017 [113]</td>
<td>✓</td>
<td>SA</td>
<td></td>
<td>Exercise</td>
</tr>
<tr>
<td>Janhunen, 2023 [42]</td>
<td>✓</td>
<td>SA</td>
<td></td>
<td>Exercise</td>
</tr>
<tr>
<td>Juhl, 2016 [44]</td>
<td>✓</td>
<td>SP</td>
<td></td>
<td>Exercise 2-way</td>
</tr>
<tr>
<td>Klement, 2019 [81]</td>
<td>✓</td>
<td>SP</td>
<td></td>
<td>Exercise 1-way, 2-way</td>
</tr>
<tr>
<td>Knapp, 2021 [83]</td>
<td>✓</td>
<td>SA</td>
<td></td>
<td>NU</td>
</tr>
<tr>
<td>First author, year</td>
<td>Exercise</td>
<td>Repository</td>
<td>Diary</td>
<td>Tracker or reminder</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>------------</td>
<td>-------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Kramer, 2003 [99]</td>
<td></td>
<td>—</td>
<td>—</td>
<td>✓</td>
</tr>
<tr>
<td>Kuether, 2019 [85]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lam, 2016 [86]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lebleu, 2023 [46]</td>
<td>✓ ✓</td>
<td>—</td>
<td>—</td>
<td>✓</td>
</tr>
<tr>
<td>LeBrun, 2022 [78]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Li, 2023 [115]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Lu, 2021 [117]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>McDonnell, 2022 [147]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Milliren, 2022 [88]</td>
<td></td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Nuevo, 2023 [50]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Osterloh, 2023 [51]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Park, 2017 [118]</td>
<td>—</td>
<td>✓</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Park, 2023 [119]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Piqueras, 2013 [54]</td>
<td>✓ ✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>First author, year</td>
<td>Exercise</td>
<td>Repository</td>
<td>Diary</td>
<td>Tracker or reminder</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>------------</td>
<td>-------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Pournajaf, 2022 [65]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>✓</td>
</tr>
<tr>
<td>Prvu Bettger, 2019 [70]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>✓</td>
</tr>
<tr>
<td>Ramkumar, 2019 [92]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Scheper, 2019 [56]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Su, 2015 [120]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>✓</td>
</tr>
<tr>
<td>Summers, 2023 [93]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Szöts, 2016 [170]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Torpil, 2022 [63]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Tripuraneni, 2021 [95]</td>
<td>✓</td>
<td>—</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Visperas, 2021 [96]</td>
<td>✓</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wang, 2023 [121]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>First author, year</td>
<td>Exercise</td>
<td>Monitoring progress</td>
<td>Functions</td>
<td>Communication Mode</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>---------------------</td>
<td>-----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Zhang, 2021 [123]</td>
<td>✓</td>
<td>— — — — — — — — —</td>
<td>2-way</td>
<td>Audio, text, video</td>
</tr>
</tbody>
</table>

a VR: virtual reality.  
b ROM: range of motion.  
c Not applicable.  
d SP: synchronous from physiotherapist.  
e F2F: face to face.  
f SA: synchronous from app.  
g AP: asynchronous from physiotherapist.  
h SC: scheduled call.  
i AA: asynchronous from app.  
j DS: danger signs.  
k NU: non-use.

**Education and Enablement**

An exercise repository in the form of videos, text, or infographics was one of the main features in the studies (n=53), of which only 20 studies described the list of exercises (Table S5 in Multimedia Appendix 1). Education for patients was part of the rehabilitation program in 17 studies. Table S6 in Multimedia Appendix 1 lists the topic areas covered in the education materials. Regarding exercise, 6 studies reported using an e-diary for maintaining an exercise log, 11 studies reported using reminders to perform exercises, and 13 studies reported using a tracker for exercise adherence (Multimedia Appendix 2). Feedback on the appropriateness of exercise performance was synchronous (biofeedback or virtual reality) from the app (n=19), directly from the health care provider via a video call with the patients (patient performing exercise live, measurement of ROM during video call, transmission of virtual avatar data to health care provider; n=14), or provided via both (n=6; Table 2). Feedback to the patient, which was either in the form of push notifications or a progress summary, was asynchronous from the app using automated programs in 2 studies. Asynchronous feedback from a health care provider in the form of instructions, messages, or an exercise regimen was reported in 13 studies. Feedback via both the app and a health care provider was provided in 3 studies (Table 2). Only 7 studies [6, 51, 75, 115, 123, 128] had an option for peer support for patients.

**Measuring Progress**

Measurement of patient-reported outcomes such as pain (n=19) was an inbuilt feature in the app. Changes in knee function and activity were monitored directly via wearables or captured using patient-reported outcome measures. These included ROM in 15 studies, knee function in 8 studies, physical activity in 20 studies, sedentary behavior in 5 studies, and sleep in 4 studies. Automatic alerts were provided to the health care provider for any danger signs such as knee pain, wound health, opioid consumption, function, ROM, number of steps, exercise adherence, and any negative response to questions after entering the postoperative follow-up in 9 studies; for non-use of the technology by patients in 4 studies; and for scheduled consultations in 18 studies (Table 2).

**Communication**

Mobile app–enabled 1-way communication included push messages, notifications, reminders, patients’ replies to inbuilt questions in the app, information sent to the patient by the health care team, and an SMS text messaging bot (n=10). Two-way communication, either via an app or in face-to-face visits, was reported in 41 studies. In addition, 11 studies reported a combination of both 1 and 2-way communication, and 1 study did not provide sufficient information about communication. Electronic communication was delivered in the form of text, audio or video messages, and direct communication (Table 2).

**Goal Setting**

Goal setting for exercises, activity, pain management, knee function, ROM, muscle strength, rehabilitation, and discharge as part of the rehabilitation program was reported in 23 studies. The goals were set by either the health care provider or the patient (Table 2).

**End Users’ Perceptions**

Of the 38 studies that reported user perspectives, 2 focused on the perspectives of health care providers, 27 focused on the perspectives of patients and caregivers, and 9 focused on the perspectives of both groups (health care providers and patients and caregivers). The approach for data collection was quantitative (n=23), qualitative (n=9), or mixed methods (n=6). The sample size ranged from 2 to 200 health care providers and from 5 to 2292 patients (Tables S7 and S8 in Multimedia Appendix 1).
Commonly used quantitative questionnaires to assess satisfaction were the System Usability Scale [129] and the net promoter score [130]. To ratify the experience with telerehabilitation, the Telemedicine Perception Questionnaire was used [131]. Acceptability and usability were assessed using the acceptance of information technology questionnaire [132] and the Telemedicine Usability Questionnaire [133]. Some studies used bespoke questionnaires to report user experience and satisfaction [32, 39, 61, 90, 94, 105, 109, 134-146]. Overall, health care providers perceived telerehabilitation and the use of technology such as biosensors as a way of improving efficiency in providing care [146], patient adherence to exercises [39, 136, 146], patient-physician communication [136], and case management [137, 146]. The main factors associated with user satisfaction with e-consultations were reliable technology, good voice or image synchronization, the refresh rate of images, sound quality, and operability of the peripherals [94, 96, 138, 139]. The key factors they perceived would influence use and uptake of technology were decreased workload (rather than increased) [140], reliability of measurements aided by technology [146], ability to measure functional outcomes objectively [141], clearer criteria when choosing appropriate patients to be enrolled in the program [140], self-efficacy in the use of technology [94, 138, 146], and ease of reporting and tracking of patient data [90]. Patients and health care providers felt e-learning modules, push notifications, and appropriate feedback from sensors and virtual reality improved self-management [138, 142-144] (Figure 4).

Figure 4. Perceptions of patients and health care providers about the technology used. ADL: activities of daily living; FAQs: frequently asked questions; KR: knee replacement; mHealth: mobile health; VR: virtual reality.

Patient satisfaction levels were reported when teleconsultation was provided via a computer, smartphone, or tablet [34, 39, 55, 56, 80, 92, 105, 121, 123, 134, 135, 145, 147-149]; telephone [61], videoconferencing [38, 77, 94, 105, 139, 141, 150-152], a web-based system [32, 50, 90, 96, 140, 153], and an mHealth-enabled integrated care model [46, 88, 138]. Patients were satisfied with telemonitoring due to improved access to services, continued support after discharge from hospital, ability for self-management, reduced need for clinic visits, reduction in cost and travel time, ability of health care providers to provide personalized care [32, 61, 94, 121, 136, 138, 140, 141, 145, 153-155], ease of use [34, 50, 55, 56, 92, 105, 135, 138, 147, 148], motivation to perform exercises [134, 135], sense of security with remote monitoring [134, 155], and empathetic communication by a health care provider [121, 135, 136, 145, 152, 155]. The reasons for dissatisfaction were lack of an in-person examination, shorter appointment times, delay in receiving reports (eg, x-ray), and an inability to transfer pictures from one technology to another [140, 145, 149, 153]. Patients provided suggestions for the app functionalities to improve the ease of use such as minimal clicks, an instructional video for app navigation, and restriction of commercial advertisements [149]. Home modifications [149], emotional well-being, information related to activities of daily living in simple text,
dietary advice, frequently asked questions, and use of traditional medicine for postoperative pain management were a few of the suggestions for app content [121].

Patients were generally satisfied with the telerehabilitation program and were ready to recommend it to others [39, 80, 85, 96, 121, 135, 151]. The use of technology for rehabilitation was influenced by computer literacy [141, 150]. However, interruption of virtual physiotherapy sessions due to poor internet issues [139] was not commonly reported (Figure 4).

Discussion

Principal Findings

This scoping review summarized the extent, user perceptions, range, and nature of technologies used to support rehabilitation following knee arthroplasty. All studies reported in this review were from upper and middle-to-upper–income countries, with a steep increase in studies in the last decade. The technologies focused on enabling patients to remember prescribed exercises as well as be able to perform them appropriately by providing synchronous and asynchronous feedback via biosensors or virtual reality. Motivation and support during recovery via technology-enabled 1-way or 2-way communication gave patients access to health care providers. Self-management and monitoring of progress were dependent on active input using e-diaries by patients or passive input through wearables. In the context where these technologies were evaluated, end users were satisfied and found remote monitoring to be acceptable for routine use.

The last decade has seen an exponential increase in the number of arthroplasties worldwide [156]; however, a corresponding increase in technological solutions to facilitate remote monitoring is nonexistent in resource-limited settings such as LMICs where the need for monitoring and a continuum of care may be higher due to lower literacy levels and lack of access to rehabilitation clinics. Research on this topic that can inform clinical practice is nonexistent in the LMIC context. Despite a high penetration of the smartphone market [157] in LMICs, a higher initial investment to develop the technology, especially in the health care sector [158], or a lack of publication of such efforts could be reasons. In LMICs, there is an increasing trend of lower limb joint replacement procedures [156]. High out-of-pocket expenditures incurred due to home visits by physiotherapists or clinic visits by patients [159] dictate the need for a cost-effective and feasible technology-based strategy to fit the context while using lessons learned from available research.

There is unequivocal evidence that there is a need for physical and psychological support from professionals during the recovery period for pain management, adherence to exercises, and modifications to therapy planning based on one’s progress [3, 160, 161]. The apps were either focused on a single function (such as communication or knowledge transfer) or were multifunctional. They were generally received well by end users; however, the usability and acceptability of these applications or remote monitoring modalities cannot be extrapolated to low health literacy and tech literacy settings. The challenges we expect with using remote monitoring in the LMIC context could be inequitable smartphone access or tech literacy, internet speed, affordability of wearables, the burden to the health system if these needs are provided free of cost, and the need for educational content in multiple languages in countries with a non-native English-speaking, multilingual population such as in India [162].

Implications for Future Research

mHealth interventions have the potential to expand the reach and effectiveness of health support by facilitating behavior change. However, to ensure these “digital behavior health interventions” effectively engage users and are effective, both microengagement (the mHealth interface) and macroengagement (evidence-based behavior change techniques) are essential [163, 164]. However, we found only a handful of studies that reported user involvement during the development stage [28, 32, 51, 58, 64, 65, 75-77, 80, 86, 88, 115, 118, 169]. Studies rarely provided an adequate explanation of the theoretical behavioral framework behind the technology-based interventions [165].

Since the context and technologies are so varied, any new applications that are developed, especially in the LMIC context, should undertake formative research with end users to understand their needs, understand their preferences, and study the local digital regulatory requirements before investing time and effort. Feasibility and pilot testing by a multidisciplinary team should be crucial steps before a full-scale evaluation [69, 166], and embedding end users’ involvement and documenting their experiences at every stage are vital to refining future interventions [164]. Further, the rehabilitation protocols should map the application features with the desired function [167, 168], and this should be confirmed by means of a process evaluation embedded within the clinical evaluation to inform the mechanism of the impact in a real-life setting [147].

Limitations

This review needs to be interpreted in light of the following limitations. This scoping review focused only on technology interventions for post-knee replacement rehabilitation and hence cannot be extrapolated to other orthopedic procedures. We did not include articles for which the full text was not available. Further, incomplete reporting on the features and functions of the technology is possible and may have affected our qualitative summary and conclusion.

We did not perform a consultation phase as per the guidelines [20], and the research question was formulated upon discussion between the researchers of the scoping review team, physiotherapists, and clinicians. We limited our search from 2001 onward; however, since knee arthroplasty and mHealth came into practice in the last 2 decades, this restriction in the search may not have an implication for our review findings.

Conclusion

Several technologies have been identified to promote adherence, increase self-efficacy, enhance self-management, and support remote monitoring. However, all the available technologies have been developed and used in developed countries. The need for remote monitoring is compelling in resource-limited contexts where these technologies were evaluated, end users were satisfied and found remote monitoring to be acceptable for routine use.

Principal Findings

This scoping review summarized the extent, user perceptions, range, and nature of technologies used to support rehabilitation following knee arthroplasty. All studies reported in this review were from upper and middle-to-upper–income countries, with a steep increase in studies in the last decade. The technologies focused on enabling patients to remember prescribed exercises as well as be able to perform them appropriately by providing synchronous and asynchronous feedback via biosensors or virtual reality. Motivation and support during recovery via technology-enabled 1-way or 2-way communication gave patients access to health care providers. Self-management and monitoring of progress were dependent on active input using e-diaries by patients or passive input through wearables. In the context where these technologies were evaluated, end users were satisfied and found remote monitoring to be acceptable for routine use.

The last decade has seen an exponential increase in the number of arthroplasties worldwide [156]; however, a corresponding increase in technological solutions to facilitate remote monitoring is nonexistent in resource-limited settings such as LMICs where the need for monitoring and a continuum of care may be higher due to lower literacy levels and lack of access to rehabilitation clinics. Research on this topic that can inform clinical practice is nonexistent in the LMIC context. Despite a high penetration of the smartphone market [157] in LMICs, a higher initial investment to develop the technology, especially in the health care sector [158], or a lack of publication of such efforts could be reasons. In LMICs, there is an increasing trend of lower limb joint replacement procedures [156]. High out-of-pocket expenditures incurred due to home visits by physiotherapists or clinic visits by patients [159] dictate the need for a cost-effective and feasible technology-based strategy to fit the context while using lessons learned from available research.

There is unequivocal evidence that there is a need for physical and psychological support from professionals during the recovery period for pain management, adherence to exercises, and modifications to therapy planning based on one’s progress [3, 160, 161]. The apps were either focused on a single function (such as communication or knowledge transfer) or were multifunctional. They were generally received well by end users; however, the usability and acceptability of these applications or remote monitoring modalities cannot be extrapolated to low health literacy and tech literacy settings. The challenges we expect with using remote monitoring in the LMIC context could be inequitable smartphone access or tech literacy, internet speed, affordability of wearables, the burden to the health system if these needs are provided free of cost, and the need for educational content in multiple languages in countries with a non-native English-speaking, multilingual population such as in India [162].

Implications for Future Research

mHealth interventions have the potential to expand the reach and effectiveness of health support by facilitating behavior change. However, to ensure these “digital behavior health interventions” effectively engage users and are effective, both microengagement (the mHealth interface) and macroengagement (evidence-based behavior change techniques) are essential [163, 164]. However, we found only a handful of studies that reported user involvement during the development stage [28, 32, 51, 58, 64, 65, 75-77, 80, 86, 88, 115, 118, 169]. Studies rarely provided an adequate explanation of the theoretical behavioral framework behind the technology-based interventions [165].

Since the context and technologies are so varied, any new applications that are developed, especially in the LMIC context, should undertake formative research with end users to understand their needs, understand their preferences, and study the local digital regulatory requirements before investing time and effort. Feasibility and pilot testing by a multidisciplinary team should be crucial steps before a full-scale evaluation [69, 166], and embedding end users’ involvement and documenting their experiences at every stage are vital to refining future interventions [164]. Further, the rehabilitation protocols should map the application features with the desired function [167, 168], and this should be confirmed by means of a process evaluation embedded within the clinical evaluation to inform the mechanism of the impact in a real-life setting [147].

Limitations

This review needs to be interpreted in light of the following limitations. This scoping review focused only on technology interventions for post-knee replacement rehabilitation and hence cannot be extrapolated to other orthopedic procedures. We did not include articles for which the full text was not available. Further, incomplete reporting on the features and functions of the technology is possible and may have affected our qualitative summary and conclusion.

We did not perform a consultation phase as per the guidelines [20], and the research question was formulated upon discussion between the researchers of the scoping review team, physiotherapists, and clinicians. We limited our search from 2001 onward; however, since knee arthroplasty and mHealth came into practice in the last 2 decades, this restriction in the search may not have an implication for our review findings.

Conclusion

Several technologies have been identified to promote adherence, increase self-efficacy, enhance self-management, and support remote monitoring. However, all the available technologies have been developed and used in developed countries. The need for remote monitoring is compelling in resource-limited contexts where these technologies were evaluated, end users were satisfied and found remote monitoring to be acceptable for routine use.
countries where knee arthroplasty is on the rise. However, irrespective of the context, it is important to involve a multidisciplinary team and include users’ perspectives during the development stage.

**What Was Already Known About the Topic**

Computer and mobile technologies to support rehabilitation following knee arthroplasty are in wide use. Telerehabilitation and remote monitoring are as effective and safe as clinic-based rehabilitation programs. They reduce out-of-pocket expenditure or health cost expenditure by reducing the time to discharge following surgery and the number of clinic visits after discharge.

**What This Study Adds**

This study provides a map of the types of technology and the functionality of mobile and computer-based multifunction applications. We summarized end users’ perceptions and reasons for satisfaction or dissatisfaction with available technology. The findings reflect the lack of research and readily available technologies for LMICs.

**Acknowledgments**

The authors would like to thank distinguished Professor Gordon Guyatt, in the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, for providing expert comments. We also thank Dr. Shyamashree Biswas, Research Intern at the George institute, for assisting with the quality check of the extracted data.

This scoping review is a part of a fellowship funded by DBT/Wellcome Trust India Alliance (grant number: IA/CPhI/20/1/505224).

**Authors’ Contributions**

ND conceptualized the protocol and conducted the search. ND, SP, PS, and ShP screened and extracted the data. The first draft was written by SP. RM and AK interpreted the study findings and gave significant feedback to the early drafts. RMad provided expert comments and suggestions and edited the later version of manuscript. ND takes responsibility for the data. All authors read and agreed to the final version of the manuscript.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Supplementary information on search strategy, included and excluded studies.

[DOCX File, 546 KB - mhealth_v12i1e47843_app1.docx ]

Multimedia Appendix 2

Raw data extraction file for rehabilitation program studies.

[XLSX File (Microsoft Excel File), 39 KB - mhealth_v12i1e47843_app2.xlsx ]

Multimedia Appendix 3

PRISMA-ScR checklist.

[PDF File (Adobe PDF File), 105 KB - mhealth_v12i1e47843_app3.pdf ]

**References**


23. Rayyan. URL: https://www.rayyan.ai/ [accessed 2023-12-11]


130. Net Promoter System. URL: https://www.netpromotersystem.com/about/measuring-your-net-promoter-score/ [accessed 2023-12-12]


McDonald J, Redley B, Livingston P, Hutchinson A, de Steiger R, Botti M. A nurse-led multimedia intervention to increase patient participation in recovery after knee arthroplasty: hybrid type II implementation study. JMIR Hum Factors 2022 May 19;9(2):e36959 [FREE Full text] [doi: 10.2196/36959] [Medline: 35588363]


Hip and knee replacement. OECD iLibrary. URL: http://tinyurl.com/vx339863 [accessed 2023-12-12]


Abbreviations

CENTRAL: Cochrane Central Register of Controlled Trials
JBI: Joanna Briggs Institute
LMIC: low and middle-income countries
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Review and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial
ROM: range of motion

©Sabhya Pritwani, Purnima Shrivastava, Shruti Pandey, Ajit Kumar, Rajesh Malhotra, Ralph Maddison, Niveditha Devasenapathy. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
The Goldilocks Dilemma on Balancing User Response and Reflection in mHealth Interventions: Observational Study

Lyndsay A Nelson1,2, PhD; Andrew J Spieker3, PhD; Lauren M LeStourgeon1,2, MPH; Robert A Greevy Jr3, PhD; Samuel Molli1,2, BS; McKenzie K Roddy1,2, PhD; Lindsay S Mayberry1,2,4, MS, PhD

Abstract

Background: Mobile health (mHealth) has the potential to radically improve health behaviors and quality of life; however, there are still key gaps in understanding how to optimize mHealth engagement. Most engagement research reports only on system use without consideration of whether the user is reflecting on the content cognitively. Although interactions with mHealth are critical, cognitive investment may also be important for meaningful behavior change. Notably, content that is designed to request too much reflection could result in users’ disengagement. Understanding how to strike the balance between response burden and reflection burden has critical implications for achieving effective engagement to impact intended outcomes.

Objective: In this observational study, we sought to understand the interplay between response burden and reflection burden and how they impact mHealth engagement. Specifically, we explored how varying the response and reflection burdens of mHealth content would impact users’ text message response rates in an mHealth intervention.

Methods: We recruited support persons of people with diabetes for a randomized controlled trial that evaluated an mHealth intervention for diabetes management. Support person participants assigned to the intervention (n=148) completed a survey and received text messages for 9 months. During the 2-year randomized controlled trial, we sent 4 versions of a weekly, two-way text message that varied in both reflection burden (level of cognitive reflection requested relative to that of other messages) and response burden (level of information requested for the response relative to that of other messages). We quantified engagement by using participant-level response rates. We compared the odds of responding to each text and used Poisson regression to estimate associations between participant characteristics and response rates.

Results: The texts requesting the most reflection had the lowest response rates regardless of response burden (high reflection and low response burdens: median 10%, IQR 0%-40%; high reflection and high response burdens: median 23%, IQR 0%-51%). The response rate was highest for the text requesting the least reflection (low reflection and low response burdens: median 90%, IQR 61%-100%) yet still relatively high for the text requesting medium reflection (medium reflection and low response burdens: median 75%, IQR 38%-96%). Lower odds of responding were associated with higher reflection burden (P<.001). Younger participants and participants who had a lower socioeconomic status had lower response rates to texts with more reflection burden, relative to those of their counterparts (all P values were <.05).

Conclusions: As reflection burden increased, engagement decreased, and we found more disparities in engagement across participants’ characteristics. Content encouraging moderate levels of reflection may be ideal for achieving both cognitive investment and system use. Our findings provide insights into mHealth design and the optimization of both engagement and effectiveness.

(JMIR Mhealth Uhealth 2024;12:e47632) doi:10.2196/47632

KEYWORDS

engagement; mobile phone; text messaging; messaging; SMS; diabetes; diabetic; mobile health; mHealth; technology; user response; users; quality of life; engagement; mHealth management; management; socioeconomic; effectiveness; support person; support worker; support persons; text message; text messages; reflection; behavior change
Introduction

Background

Mobile health (mHealth) is transforming health delivery as a highly convenient and effective approach for supporting individuals with chronic conditions [1-3]. Delivered via phones, tablets, and wearables, mHealth provides education, motivation, monitoring, and other forms of support to improve health behaviors. SMS text messaging is one form of mHealth that is uniquely poised to benefit everyone, including people who are older, are disadvantaged, and are from traditionally minoritized racial or ethnic backgrounds [4-6]. A critical factor influencing mHealth effectiveness is users’ engagement or interaction with the technology, which is typically measured via system use [7-9]. Across the mHealth literature, engagement tends to be highly variable [10,11], which has spurred a whole body of research that aims to understand predictors of engagement, including user characteristics and intervention features (eg, intervention duration and frequency of sending content) [10-14]. However, very little research has attended to the type of mHealth content that users are expected to engage with [15] and, more specifically, how the content may be requesting more or less cognitive reflection.

The primary goal in having users engage with mHealth content is health behavior change. With respect to mHealth interventions, there is a hypofocus on wanting the user to interact with the technology (eg, responding to a text message), with less consideration of whether the user is reflecting on the content cognitively (eg, reflecting on past behavior and planning future behavior) [8]. Although interaction with the technology is a critical measure, there is a growing consensus that cognitive investment is also important for meaningful behavior change in many types of mHealth interventions [16-18]. Notably, content may be designed in a way that represents a low response burden, thereby easily eliciting a response (ie, producing high engagement), but such content may not evoke the necessary cognitive reflection required to change behavior [18]. Alternatively, content that is designed to encourage deeper reflection may overwhelm users, which risks them disengaging completely. Understanding how to strike the balance between response and reflection has critical implications for effective engagement (ie, engagement needed to impact outcomes) [19].

To our knowledge, no studies have explored how these variations would impact users’ engagement with text messages, as measured via response rates. We also explored support persons’ characteristics that were associated with response rates for each type of text and described the different responses to each text.

Methods

Study Design and Eligibility

This study was conducted as part of the FAMS 2.0 RCT. The trial design, intervention details, and outcomes for persons with diabetes and support persons were published [22-24]. For the trial, dyads comprising a person with diabetes and their support person were randomized to FAMS or a control condition. We recruited persons with diabetes who were receiving care for type 2 diabetes at Vanderbilt University Medical Center primary care clinics. Enrolling persons with diabetes were asked to invite a support person to participate with them and receive text messages; however, support person invitation and enrollment were not required. We defined a support person as any family member or friend with whom the person with diabetes would feel comfortable talking about diabetes management and health goals. Eligible support persons were aged ≥18 years, could speak and read English, and had a mobile phone separate from that of the person with diabetes. The only exclusion criterion was the inability to receive and respond to a text after training. For this study, we analyzed data from support persons in dyads that were randomly assigned to the intervention group (FAMS).

Ethical Considerations

The Vanderbilt University Institutional Review Board approved all study procedures (institutional review board number: 200398; approved April 8, 2020), and the trial was registered on ClinicalTrials.gov (trial number: NCT04347291).

Procedure

While enrolling persons with diabetes into the trial, research assistants collected contact information for a potential support person. A research assistant then contacted potential support persons to verify interest and eligibility, obtain verbal informed consent, and ask support persons to complete a baseline survey. Surveys were completed by phone with a research assistant, on the web via an emailed link, or via a mailed paper copy, per participants’ preferences. All survey data were stored in REDCap (Research Electronic Data Capture; Vanderbilt University), a secure web-based platform that supports data capture for research studies [25,26]. In addition to collecting data on sociodemographic characteristics, surveys asked support persons to choose the time of day they wanted to receive text messages among support persons. Because support persons’ engagement with the text messages was an optional component of the intervention, we determined that this was an opportunity to explore how changing the content of these texts might impact response rates without compromising our ability to evaluate FAMS’ effects. Over the course of the RCT [22], we used a pragmatic approach to vary both the reflection burden and the response burden of the two-way text messages sent to support persons assigned to the intervention. In this observational study, our primary goal was to explore how these variations would impact users’ engagement with text messages, as measured via response rates. We also explored support persons’ characteristics that were associated with response rates for each type of text and described the different responses to each text.

Objective

Our team previously developed an mHealth intervention (delivered via text messages and phone calls) called Family/Friend Activation to Motivate Self-care (FAMS) [20,21]. FAMS is a diabetes self-management intervention that targets persons with type 2 diabetes and provides the option for persons with diabetes to invite a support person to also receive text messages. We recently evaluated FAMS in a randomized controlled trial (RCT) [22], and during routine monitoring in the first few weeks of the trial, we observed a low response rate
messages. Relevant survey responses were transferred from REDCap to our technology partner, PerfectServe, using an automated application programming interface. PerfectServe used participant information to tailor, schedule, and send text messages to support persons for 9 months. Support persons could earn a total of US $120 for completing all study surveys (through 15 mo for the larger RCT). There was no compensation for receiving or responding to text messages.

The Intervention

Persons with diabetes received daily text message support and monthly coaching sessions, during which they set behavioral diabetes self-management goals (as detailed by Mayberry et al [22]). Support persons received text messages that were designed to increase dialogue about and facilitate their involvement in the diabetes self-management of the persons with diabetes; a one-way message was sent 3 to 4 times per week, and a two-way message (also known as an interactive text message) was sent once per week. One-way messages were either a general text message about providing diabetes self-management support or a text message tailored to the identified diabetes goals of the persons with diabetes. Two-way messages asked support persons about how they supported the health of the persons with diabetes. Support persons who replied to the two-way text received an automated response, thanking them for their answer.

Although an individual support person’s intervention experience lasted 9 months (36 wk), intervention delivery for the trial lasted 2 years. Over the course of those 2 years, we varied both the response burden (the level of information requested for the response relative to that of other messages) and the reflection burden (the level of cognitive reflection requested relative to that of other messages) of the weekly two-way text messages, which were sent to support persons in 6 fixed periods (ie, waves). The waves coincided with the weeks of the trial; they did not coincide with the weeks of each individual support person’s intervention experience. Figure 1 includes the content for each version of the text message, the weeks of the trial when each text was sent (ie, calendar time), and the respective waves. We started the trial (wave 1) by sending a text message that was high in both reflection burden and response burden (high/high). In wave 2, we tested a text that was low in both reflection burden and response burden (low/low), and then in wave 3, we tested a text that involved medium reflection burden and low response burden (medium/low). In wave 4, we retested the high/high message to help determine if the point at which the text was sent during the trial impacted engagement. In wave 5, we sought to delineate the relative impacts of reflection burden and response burden; therefore, we tested a text message that was high in reflection burden and low in response burden (high/low). Finally, we closed out the trial by retesting the low/low text message (wave 6). The decisions about what messages to test were made iteratively based on response rates to the prior message, with the goal of learning how much reflection we could request while still achieving a relatively high response rate.
Of note, each support person only received the versions of the two-way text message that were sent during their 36-week trial participation, with most (120/148, 81.1%) receiving 2 or 3 different versions and no participants receiving the same message in 2 separate waves. Because this analysis used data from support persons only, we refer to them as participants henceforth.

**Measures**

**Sociodemographic and Relationship Characteristics**

We collected self-reported data on age, gender, race, ethnicity, socioeconomic status (measured based on education [ie, years in school] and annual household income), and health literacy (assessed via the Brief Health Literacy Screen [27]). In addition, we asked whether participants were cohabitating with the persons with diabetes and the frequency with which they provided diabetes-specific helpful involvement to the persons with diabetes at baseline, as assessed via the Family and Friend Involvement in Adults’ Diabetes (FIAD) helpful subscale, support person version [28].

**Engagement**

We operationalized engagement by using response rate (ie, two-way messages responded to divided by the two-way messages sent, for each participant).
Analyses

Statistical Analysis Overview

All statistical analyses were performed by using R version 4.2.1 (R Foundation for Statistical Computing). We described participant characteristics via means and SDs or via frequencies and percentages, as appropriate. Except for when examining temporality, message waves that included the same version of the two-way text message were grouped together. Because this study was exploratory, we did not perform sample size calculations.

Overall Engagement by Text Message Version

For each version of the text message, we determined the proportion of two-way text messages sent to support persons that received a response by study week (ie, calendar time). We also generated summary statistics (means, medians, and first and third quartiles) for response rates at the participant level; reporting both mean and median provides more detailed information on the distribution of data. If participants withdrew during their intervention experience, we calculated their response rates based on the data available prior to their withdrawal. To account for repeated measures within participants, we used generalized estimating equations with a working independence correlation structure and a logistic link function to compare the odds of responding to two-way text messages across the four versions.

Participant Characteristics and Engagement

We used Poisson regression to estimate associations (as incidence rate ratios) between participant characteristics and text message response rates for each version of the text message. We included the number of two-way messages sent to a participant as an offset term in order to account for variation in the number of messages sent to each participant in a given wave; therefore, the exponentiated coefficients from the Poisson regression model compared response rates on a per-message basis. Participant characteristics included age, race and ethnicity (non-Hispanic White vs minoritized race or ethnicity), gender, education (years), annual household income ($\geq 50,000 per year), health literacy (Brief Health Literacy Screen), whether the persons with diabetes and support persons were cohabitating, and self-reported baseline helpful involvement (FIAD). Further, we multiplied participants’ age by 10 to allow for easier interpretation of the results. Especially in regression models, it can be difficult to interpret the association between age and an outcome when the change in the outcome is based on a single-year change in age (ie, the coefficients end up being too small). Scaling the age variable in this way allowed us to interpret the findings in a more meaningful way, that is, we compared groups that differed in age by 1 decade rather than 1 year.

For this analysis, we excluded 5 participants who were missing all baseline data. However, missing covariate values were otherwise addressed via multiple imputation by chained equations (M=500 iterations).

Types of Responses

We characterized the responses to each version of the two-way text message. For the low/low, medium/low, and high/low texts, we reported the frequency of responses based on what the respective text requested (eg, “Yes,” “No,” “1,” “2,” “3,” “4,” or “5”). For the high/high texts, 2 team members reviewed responses and categorized each as being either high effort or low effort. High-effort responses included comments on what went well that week, comments on what could go better next week, or both, and they referred to a diabetes self-management behavior such as diet, exercise, stress management, or communication (eg, “[He] and I got out several times this week walking after work. Our biggest problem is watching portion size when we are eating. Always continue to work on that.”). A low-effort response consisted of only a brief phrase that did not reference a diabetes self-management behavior (eg, “Things went well” and “We were on vacation this week”) or did mention a behavior but was unclear as to what went well or what could go better next week (eg, “Walking”).

Results

Participant Characteristics

In the trial, of the 150 support person participants who were enrolled and randomized to receive the FAMS intervention, 2 withdrew before the intervention started. The remaining 148 were included in the analyses (Table 1). The mean age was 50.3 (SD 14.7) years; 28.4% (42/148) of participants were men, and 33.1% (49/148) reported a minoritized racial or ethnic background. The mean length of education was 14.9 (SD 2.5) years, and 31.1% (46/148) of participants had annual household incomes of <$US 50,000. Over half (84/148, 56.8%) were spouses or partners of the persons with diabetes, and 70.3% (104/148) were cohabitating with the persons with diabetes. Further, 9 participants withdrew at some point during the intervention; the analyses below reflect their engagement during the time they participated.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong> (y), mean (SD)</td>
<td>50.3 (14.7)</td>
</tr>
<tr>
<td><strong>Gender</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>42 (28.4)</td>
</tr>
<tr>
<td>Women</td>
<td>101 (68.2)</td>
</tr>
<tr>
<td><strong>Race and ethnicity</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>92 (62.2)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>29 (19.6)</td>
</tr>
<tr>
<td>Other non-Hispanic races</td>
<td>12 (8.1)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8 (5.4)</td>
</tr>
<tr>
<td>Missing</td>
<td>7 (4.7)</td>
</tr>
<tr>
<td><strong>Socioeconomic status</strong></td>
<td></td>
</tr>
<tr>
<td>Education (y), mean (SD)</td>
<td>14.9 (2.5)</td>
</tr>
<tr>
<td><strong>Annual household income (US $), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;35,000</td>
<td>25 (16.9)</td>
</tr>
<tr>
<td>35,000-49,999</td>
<td>21 (14.2)</td>
</tr>
<tr>
<td>50,000-74,999</td>
<td>22 (14.9)</td>
</tr>
<tr>
<td>75,000-99,999</td>
<td>24 (16.2)</td>
</tr>
<tr>
<td>≥100,000</td>
<td>40 (27)</td>
</tr>
<tr>
<td>Missing or unknown</td>
<td>16 (10.8)</td>
</tr>
<tr>
<td>Health literacy (BHLS), mean (SD)</td>
<td>13.7 (1.5)</td>
</tr>
<tr>
<td><strong>Relationship variables</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Relationship type</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>Spouse or partner</td>
<td>84 (56.8)</td>
</tr>
<tr>
<td>Parent</td>
<td>15 (10.1)</td>
</tr>
<tr>
<td>Son or daughter</td>
<td>22 (14.9)</td>
</tr>
<tr>
<td>Grandchild</td>
<td>7 (4.7)</td>
</tr>
<tr>
<td>Friend</td>
<td>11 (7.4)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (2.7)</td>
</tr>
<tr>
<td>Missing</td>
<td>5 (3.3)</td>
</tr>
<tr>
<td><strong>Helpful involvement (FIAD), mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>2.7 (0.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Cohabitating with person with diabetes</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>104 (70.3)</td>
<td></td>
</tr>
</tbody>
</table>

10 participants did not report their date of birth (ie, age).
5 participants did not provide data on gender.
8 participants did not report years of education.
BHLS: Brief Health Literacy Screen.
5 participants did not have data for the BHLS measure.
FIAD: Family and Friend Involvement in Adults' Diabetes.
5 participants did not have data for the FIAD helpful subscale.
8 participants did not have data about cohabitating with the persons with diabetes.

**Overall Engagement by Text Message Type**

Figure 2 presents the proportion of two-way text messages that received a response within each week of the trial (ie, by calendar time). Notably, text message response rates for waves 1 and 4 (both were high/high message waves) were comparable, as were those for waves 2 and 6 (both were low/low message waves). Table 2 includes descriptive statistics for the overall and
participant-level response rates for each version of the two-way text. The median response rates for the high/high, medium/low, and low/low messages were 23% (IQR 0%-51%), 75% (IQR 38%-96%), and 90% (IQR 61%-100%), respectively. When we kept reflection burden high but lowered response burden in the high/low message, the median response rate (10%, IQR 0%-40%) was closest to that for the high/high message, suggesting that reflection burden was responsible for the lower response rates seen with the high/high message.

**Figure 2.** Text message response rates by week across each wave. Response data were excluded for the first 4 weeks and the last 5 weeks of the trial when <5 individuals were receiving the intervention.
Response rates for each version of the two-way text message.

<table>
<thead>
<tr>
<th>Reflection burden and response burden</th>
<th>Participants included in analysis, n</th>
<th>Participant-specific response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>%, mean (SD)</td>
</tr>
<tr>
<td>High and high</td>
<td>127</td>
<td>33 (33)</td>
</tr>
<tr>
<td>Low and low</td>
<td>123</td>
<td>74 (32)</td>
</tr>
<tr>
<td>Medium and low</td>
<td>125</td>
<td>65 (35)</td>
</tr>
<tr>
<td>High and low</td>
<td>55</td>
<td>26 (32)</td>
</tr>
</tbody>
</table>

We also compared the odds of responding to the four versions of the text message (Table 3). When compared to the high/high message, the odds of responding to the low/low message was 53% (95% CI 42%-65%) higher, the odds of responding to the medium/low message was 40% (95% CI 31%-49%) higher, and the odds of responding to the high/low message was 7.5% (95% CI 0.7%-14%) lower. All other pairwise comparisons (Table 3) indicated decreasing odds of responding at increasing levels of reflection burden.

Comparison of text message response rates by text message version. Included are odds ratios (ORs) and 95% CIs, along with P values.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative to high reflection burden and high response burden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low reflection burden and low response burden</td>
<td>1.53 (1.42-1.65)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Medium reflection burden and low response burden</td>
<td>1.40 (1.31-1.49)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High reflection burden and low response burden</td>
<td>0.93 (0.86-0.99)</td>
<td>.03</td>
</tr>
</tbody>
</table>

Relative to low reflection burden and low response burden

| Medium reflection burden and low response burden | 0.92 (0.86-0.98) | .01     |
| High reflection burden and low response burden  | 0.60 (0.55-0.66) | <.001   |

Relative to medium reflection burden and low response burden

| High reflection burden and low response burden  | 0.66 (0.61-0.72) | <.001   |

Participant Characteristics Associated With Odds of Responding

Table 4 presents estimated incident rate ratios, along with 95% CIs and P values, from multivariate Poisson regression models that were used to identify participant characteristics predictive of response rate. Younger participants, participants who were not cohabitating with the persons with diabetes, and participants who had a lower socioeconomic status had lower response rates to both the high/high message and the medium/low message compared to those of older participants, participants who were cohabitating, or participants who had a higher socioeconomic status, respectively. The only characteristic associated with response rates for the low/low message was gender, such that men had lower response rates. Further, the only characteristic associated with response rates for the high/low message was age, such that younger age was associated with lower response rates. Across message versions, younger participants had lower response rates to any message with more burden than the low/low message, and participants who were not cohabitating with the persons with diabetes had lower response rates to the higher-burden messages than those of participants who were cohabitating. Race, ethnicity, health literacy, and baseline helpful involvement provided to the person with diabetes did not show patterns indicating the prediction of response rates to any message version.
Despite the potential of mHealth to enhance self-management support and quality of life, there are still key gaps in understanding how to optimize mHealth engagement [16,17,19]. Most engagement research reports only on system use without consideration of the cognitive reflection done in the process of engaging with the content [16,29]. Ideally, we want to encourage reflection that results in meaningful behavior change, but it is unclear how much we can request, with respect to reflection, before users disengage. We varied the reflection and response burdens of two-way text messages to examine how these variations impacted users’ engagement, as assessed via response rates. We found, generally, that as the reflection burden of the message increased, participants’ engagement decreased. Importantly, when the same version of the text was sent at different points in the trial, participants’ engagement was consistent, suggesting that the message itself was key for response rates. The response rates for the high/low message were similar to those for the high/low message, and this supports reflection burden (vs response burden) being the primary driver of lower engagement. We also found evidence that as the reflection burden of the message increased, there were more disparities in engagement across participant characteristics. This finding helps inform who we may lose with content requiring more from users and thus how to ensure mHealth interventions do not widen existing health disparities.

Research focused on promoting mHealth engagement has proliferated in recent years, with the primary goal of increasing system use [30]. Across such studies, we see evidence that content that is more personalized; uses simpler, nontechnical language; and is empowering results in higher engagement [10,19,31-35]. Less research has compared specific types of language; and is empowering results in higher engagement [30]. Across such studies, we see evidence that content requiring more from users and thus how to ensure mHealth interventions do not widen existing health disparities.

### Discussion

#### Principal Results

Despite the potential of mHealth to enhance self-management support and quality of life, there are still key gaps in understanding how to optimize mHealth engagement [16,17,19]. Most engagement research reports only on system use without consideration of the cognitive reflection done in the process of engaging with the content [16,29]. Ideally, we want to encourage reflection that results in meaningful behavior change, but it is unclear how much we can request, with respect to reflection, before users disengage. We varied the reflection and response burdens of two-way text messages to examine how these variations impacted users’ engagement, as assessed via response rates. We found, generally, that as the reflection burden of the message increased, participants’ engagement decreased. Importantly, when the same version of the text was sent at different points in the trial, participants’ engagement was consistent, suggesting that the message itself was key for response rates. The response rates for the high/low message were similar to those for the high/low message, and this supports reflection burden (vs response burden) being the primary driver of lower engagement. We also found evidence that as the reflection burden of the message increased, there were more disparities in engagement across participant characteristics. This finding helps inform who we may lose with content requiring more from users and thus how to ensure mHealth interventions do not widen existing health disparities.

#### Types of Responses

In this section, we report on engagement at the text message level (vs the participant level). For the high/high messages, 1429 texts were sent, and 445 responses were received. Further, 13 responses were excluded from the analysis because the content was uninterpretable or was not relevant to the two-way text prompt. The reviewers categorized each response into the content that is more personalized; uses simpler, nontechnical language; and is empowering results in higher engagement [30]. Across such studies, we see evidence that content requiring more from users and thus how to ensure mHealth interventions do not widen existing health disparities.

Research focused on promoting mHealth engagement has proliferated in recent years, with the primary goal of increasing system use [30]. Across such studies, we see evidence that content that is more personalized; uses simpler, nontechnical language; and is empowering results in higher engagement [10,19,31-35]. Less research has compared specific types of

### Table . Participant characteristics predicting text message response rates for each version of the text. Presented are estimated incident rate ratios (IRRs) and 95% CIs, along with \( P \) values.\(^a\)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>High reflection/high response</th>
<th>Low reflection/low response</th>
<th>Medium reflection/low response</th>
<th>High reflection/low response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR (95% CI)</td>
<td>( P ) value</td>
<td>IRR (95% CI)</td>
<td>( P ) value</td>
</tr>
<tr>
<td>Age (y x 10)</td>
<td>1.14 (1.06-1.23)</td>
<td>.004(^b)</td>
<td>0.99 (0.95-1.03)</td>
<td>.67</td>
</tr>
<tr>
<td>Race and ethnicity</td>
<td>1.07 (0.85-1.35)</td>
<td>.57</td>
<td>0.96 (0.84-1.09)</td>
<td>.51</td>
</tr>
<tr>
<td>Gender (men)</td>
<td>0.62 (0.49-0.78)</td>
<td>&lt;.001(^b)</td>
<td>0.78 (0.68-0.89)</td>
<td>&lt;.001(^b)</td>
</tr>
<tr>
<td>Education</td>
<td>0.98 (0.94-1.02)</td>
<td>.34</td>
<td>1.00 (0.98-1.03)</td>
<td>.77</td>
</tr>
<tr>
<td>Income</td>
<td>0.71 (0.57-0.89)</td>
<td>.003(^b)</td>
<td>1.05 (0.91-1.21)</td>
<td>.51</td>
</tr>
<tr>
<td>BHLS(^c)</td>
<td>1.04 (0.97-1.12)</td>
<td>.30</td>
<td>1.04 (1.00-1.09)</td>
<td>.06</td>
</tr>
<tr>
<td>Cohabitating</td>
<td>1.37 (1.09-1.73)</td>
<td>.007(^b)</td>
<td>1.03 (0.90-1.19)</td>
<td>.64</td>
</tr>
<tr>
<td>FIAD(^d)</td>
<td>1.05 (0.94-1.17)</td>
<td>.41</td>
<td>1.00 (0.93-1.07)</td>
<td>.99</td>
</tr>
</tbody>
</table>

\(^a\)A total of 5 support persons without baseline characteristics were excluded from this analysis: 2 were excluded from the models for the high/high, low/low, and medium/low messages; 1 was excluded from the models for the high/high and medium/low messages; 1 was excluded from the models for the high/high and medium/low messages; and 1 was excluded from the models for the low/low and medium/low messages.

\(^b\)\( P \) < .05.

\(^c\)BHLS: Brief Health Literacy Screen.

\(^d\)FIAD: Family and Friend Involvement in Adults’ Diabetes.
content and has rarely tested different types within the same study. An exception is a recent study by Klimis et al [15], wherein they used machine learning to demonstrate that text messages with informative (providing health facts or education) and instructional (providing tips or recommendations) message intents were associated with increased engagement, while notification messages that addressed noneducational matters (eg, welcome and exit messages) were associated with reduced engagement [15]. Our study targeted a two-way message and varied the levels of reflection and response burdens in that message. By adjusting the reflection level specifically, we gained unique insight into how engagement with each message variation may ultimately influence behavior change [17,18].

The main way in which our study differs from others in this area of research is that we looked beyond system use as the sole dimension of mHealth engagement. Our goal was not necessarily to see which message resulted in the highest response rate but rather to determine how much reflection we could request from users and achieve a level of interaction that suggested that they were still interested in the content.

Although our results show generally that engagement decreases with more reflection, the nuances in our findings allow us to provide unique recommendations around mHealth design. For instance, it may be best to alternate through content with different levels of reflection burden. Although users were more likely to respond to content that was lower in reflection burden, nearly all (350/426, 82.2%) of the responses that we received to the high-processing messages included a high-effort level of reflection. The act of asking people to reflect stimulates internal thoughts that are difficult to measure without a response [36] but may still occur among some persons who do not respond. Alternating content may help promote periodic responding and reflecting throughout an intervention experience. Another option involves using an adaptive intervention to tailor the content based on each person’s responsiveness. That is, everyone could start receiving content with a high reflection burden, but if a person’s response rate starts to drop, they could then switch to content with a moderate reflection burden. Finally, especially in situations where there is limited flexibility with the mHealth functionality, researchers may consider sending the medium/low and high/low text messages to all participants, given that the content encouraged a moderate level of reflection (more than the low/low message) yet still yielded a high response rate.

Limitations

Our study has several limitations to acknowledge. For instance, our results are based on an SMS text messaging intervention, which is a specific form of mHealth. It is possible that users would have responded differently if the content was delivered via an app or wearable technology. Importantly, compared to apps and other internet-dependent technologies, SMS text messaging is both lower in cost and more easily accessed, and it tends to have higher rates of engagement [4,37]. In addition, this study recruited people with diabetes and their support persons from a specific region in Middle Tennessee. We acknowledge that the findings may not be generalizable to other types of individuals who are living in other locations. Relatedly, the content asked about how the support persons supported the health of the persons with diabetes, and engagement may differ when asking about a user’s own health; however, the marked differences in engagement across message types support broader applications.

Another limitation of our work is that we restricted our assessment of engagement to a behavioral measure (ie, responding to the text) and did not have a way to assess participants’ cognitive investment or experience with each version of the text. Based on our analysis of participants’ responses to the high/high message, it appeared that responders were cognitively engaged, but we were not able to compare cognitive engagement across the other messages. In addition, the sample size for the high/low message analysis (n=55) was considerably smaller compared to those for the other message analyses, which was due to testing the high/low message during only 1 wave near the end of the trial when fewer participants were enrolled. Relatedly, we did not have the time or a sufficient number of participants toward the end of the trial to test the medium/low and high/low text messages in a second wave, and we do not know for certain whether engagement with these texts could be impacted by temporality; however, as engagement with the high/high and low/low texts remained similar across multiple waves, it is unlikely. The ordering of the text messages was variable across participants, and due to the observational nature of this study, we cannot determine the extent that ordering may have impacted results; however, the average response rates and trends across message versions and waves provide general insights on how these variations may impact engagement. Finally, we did not assess the impact of engagement on outcomes, as this fell outside the scope of our study; however, other studies in digital health have examined this association [38,39].

Conclusions

In order for individuals to benefit from mHealth and achieve desired effects on outcomes, engagement with the mHealth tool is needed. Our results help elucidate how truly complex the nature of engagement is. Although our past work and that of others have demonstrated the importance of behaviorally interacting with mHealth interventions [9,31,40,41], this measure represents one piece of a larger puzzle. Engagement may be best conceptualized as including both a behavioral dimension and a cognitive dimension. Balancing these dimensions may be what is ultimately needed to achieve effective engagement for impacting intended outcomes. Our study contributes to a growing body of research that encourages a more nuanced approach to studying engagement that goes beyond measuring system use. We hope that our findings help advance the field of mHealth and inform intervention design, with the goal of optimizing both engagement and effectiveness.
represent the official views of the NIH. MKR is funded by the National Center for Advancing Translational Sciences (NIH; grant KL2TR002245).

Authors’ Contributions
LAN, AJS, LSM, RAG Jr, and MKR significantly contributed to the conception of the study and planned analyses. LAN wrote the manuscript. AJS analyzed the quantitative data, and RAG Jr oversaw data analyses. LML and SM analyzed responses to the high/high message. LSM was the principal investigator of the Family/Friend Activation to Motivate Self-care (FAMS) study. All authors were involved in data interpretation and manuscript revision and approved the final version submitted for publication.

Conflicts of Interest
None declared.

References

17. Yeager CM, Benight CC. If we build it, will they come? issues of engagement with digital health interventions for trauma recovery. Mhealth 2018 Sep 11;4:37. [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]


Abbreviations
FAMS: Family/Friend Activation to Motivate Self-care
FIAD: Family and Friend Involvement in Adults' Diabetes
mHealth: mobile health
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
Investigating Citizens’ Acceptance of Contact Tracing Apps: Quantitative Study of the Role of Trust and Privacy

Grace Fox¹, PhD; Lisa van der Werff¹, PhD; Pierangelo Rosati², PhD; Theo Lynn¹, PhD

¹Irish Institute of Digital Business, Dublin City University, Dublin, Ireland
²J.E. Cairnes School of Business & Economics, University of Galway, Galway, Ireland

Corresponding Author:
Theo Lynn, PhD
Irish Institute of Digital Business
Dublin City University
Collins Ave, Dublin9
Dublin
Ireland
Phone: 353 1 700 6873
Email: theo.lynn@dcu.ie

Abstract

Background: The COVID-19 pandemic accelerated the need to understand citizen acceptance of health surveillance technologies such as contact tracing (CT) apps. Indeed, the success of these apps required widespread public acceptance and the alleviation of concerns about privacy, surveillance, and trust.

Objective: This study aims to examine the factors that foster a sense of trust and a perception of privacy in CT apps. Our study also investigates how trust and perceived privacy influence citizens’ willingness to adopt, disclose personal data, and continue to use these apps.

Methods: Drawing on privacy calculus and procedural fairness theories, we developed a model of the antecedents and behavioral intentions related to trust and privacy perceptions. We used structural equation modeling to test our hypotheses on a data set collected at 2 time points (before and after the launch of a national CT app). The sample consisted of 405 Irish residents.

Results: Trust in CT apps was positively influenced by propensity to trust technology (β=.074; P=.006), perceived need for surveillance (β=.119; P<.001), and perceptions of government motives (β=.671; P<.001) and negatively influenced by perceived invasion (β=−.224; P<.001). Perceived privacy was positively influenced by trust (β=.466; P<.001) and perceived control (β=.451; P<.001) and negatively influenced by perceived invasion (β=−.165; P<.001). Prelaunch intentions toward adoption were influenced by trust (β=.590; P<.001) and perceived privacy (β=.247; P<.001). Prelaunch intentions to disclose personal data to the app were also influenced by trust (β=.215; P<.001) and perceived privacy (β=.208; P<.001) as well as adoption intentions before the launch (β=.550; P<.001). However, postlaunch intentions to use the app were directly influenced by prelaunch intentions (β=.530; P<.001), but trust and perceived privacy only had an indirect influence. Finally, with regard to intentions to disclose after the launch, use intentions after the launch (β=.665; P<.001) and trust (β=.215; P<.001) had a direct influence, but perceived privacy only had an indirect influence. The proposed model explained 74.4% of variance in trust, 91% of variance in perceived privacy, 66.6% of variance in prelaunch adoption intentions, 45.9% of variance in postlaunch use intentions, and 83.9% and 79.4% of variance in willingness to disclose before the launch and after the launch, respectively.

Conclusions: Positive perceptions of trust and privacy can be fostered through clear communication regarding the need and motives for CT apps, the level of control citizens maintain, and measures to limit invasive data practice. By engendering these positive beliefs before launch and reinforcing them after launch, citizens may be more likely to accept and use CT apps. These insights are important for the launch of future apps and technologies that require mass acceptance and information disclosure.

(JMIR Mhealth Uhealth 2024;12:e48700) doi:10.2196/48700

KEYWORDS
privacy; trust; public health surveillance; contact tracing; mobile apps; adoption; information disclosure
**Introduction**

**Background**

The outbreak of COVID-19 and the ensuing global pandemic resulted in many governments undertaking digital government transformation [1] through the introduction of public health surveillance technologies including contact tracing (CT) apps [2,3]. As a result, and unlike previous health emergencies, governments had access to an unprecedented volume, variety, and velocity of location and health data [4]. The use of such data for epidemiological surveillance can aid in decision support, accelerate case identification, interrupt community transmission, and enable public health communication [5]. Notwithstanding these benefits, the pace at which these apps have been implemented and the level of surveillance they enable have raised ethical concerns [6] and fears around privacy and public trust [7].

The success of CT apps is dependent on uptake by large populations [8], and privacy-related concerns have been positioned as a critical barrier facing government-introduced CT apps [9]. Government-introduced CT apps differ from surveillance technologies and mobile health (mHealth) apps, as they combine both location-based data and electronic personal health information (ePHI). Both these contexts, in themselves, raise significant privacy concerns, particularly with respect to potential secondary use and government intrusion [10]. Thus, the COVID-19 pandemic presents a unique empirical context to explore citizens’ perceptions of health surveillance using mobile apps that capture both location-based data and ePHI. Government-introduced CT apps constitute a new public health context. Although there is an established literature based on traditional CT, digital CT is an innovation that can only be fully explored during pandemics, and thus research opportunities are limited. Given the novel context for individuals and public health agencies, where the former engage with new or unfamiliar trust referents, it is particularly critical to explore how individuals marry competing beliefs about surveillance, trust, and government motives for introducing these technologies and how these beliefs influence their behavioral responses. Extant literature has demonstrated the importance of trust in the government in influencing CT adoption [11,12] and perceptions of CT technologies [13] and shown that privacy concerns represent a barrier to adoption [9].

**Prior Work**

This study builds upon important extant research focusing on the acceptance of CT apps to delve further into the role of privacy and trust and addresses 3 gaps in the literature. First, existing studies support the importance of trust in driving the acceptance of CT apps, but the approach to measuring trust and the trust referent under examination varies. For example, studies have found that high trust in the national government, the health care system, and science positively impacted willingness to use CT apps in Switzerland [14]. A US-based study found that trust in COVID-19 information positively influenced citizens’ comfort with and acceptance of CT [8]. In this study, our emphasis is on technology-related trust, as opposed to trust in an individual or organization. Specifically, we examined trust perceptions regarding a specific technology, namely a CT app. In addition, we investigated the perceptual factors that shape trust, namely perceptions of government motivations, need for surveillance, and propensity to trust technology (PTTT).

Second, studies have investigated the influence of privacy on CT adoption, with many studies finding that privacy concerns reduce intentions toward adopt CT apps [11,14]. Several studies conclude that privacy represents a barrier to the success of CT apps, with respondents in several countries citing privacy concerns as a reason for not installing apps [9] or conversely, those with low privacy concerns are more likely to use CT apps [12]. However, the influence of privacy concerns on CT adoption intentions was weak in another study [15]. Although privacy concerns are the most common proxy for measuring privacy across many contexts, the negative connotation and failure to directly capture privacy suggest the need for more precise operationalization [16]. Thus, we focus on perceived privacy defined as “an individual’s self-assessed state in which external agents have limited access to information about him or her” [16]. In this study, perceived privacy refers to a citizen’s belief in the level of privacy afforded by the CT app. Perceived privacy influenced intentions toward CT apps in a recent Brazilian study, thus supporting its use [17]. To further our understanding of perceived privacy, this study investigated the role of trust, perceived control, and perceived intrusion in shaping citizens’ perceptions of privacy.

Third, many CT studies are cross-sectional in nature, with the exception of a small number of longitudinal studies [12]. In addition, the literature focuses largely on whether citizens adopt CT apps or engage in behaviors recommended by CT apps such as staying at home [14,18]. This study broadens our understanding of citizens’ acceptance of CT apps by examining 2 variables related to acceptance, namely intention to adopt or continue using the app and willingness to disclose personal information. These acceptance variables are measured before and after the app launch, thereby deepening our understanding of how privacy and trust influence intentions toward and use of CT apps.

We argue that understanding the determinants of success of CT apps is critical not only for future digital CT but also for other contexts that require rapid digital technology adoption by the population [19]. This paper proceeds with an overview of the hypothesized relationships and our research context. Our methodology, data analysis, and results are presented in the following sections. The Discussion section outlines the implications of this study. The paper concludes with the limitations and avenues for future research.

**Theory**

**Overview**

Privacy Calculus Theory (PCT) posits that before engaging in a behavior such as adopting a new technology or disclosing personal information, individuals will conduct a cognitive comparison of the costs and benefits associated with this behavior [20]. Individuals are likely to engage in the behavior for as long as the benefits outweigh the costs [20]. Thus, PCT has direct comparisons with the concept of calculus-based trust,
which underpins trust decisions when engaging with new or unfamiliar trust referents [21]. PCT has been operationalized in many contexts using a variety of belief-based variables that can be grouped into confidence beliefs and risk or privacy beliefs. An extension of PCT that holds considerable promise for understanding privacy and trust in the context of CT apps is the inclusion of the procedural fairness theory. Procedural fairness refers to an individual’s perception that a particular activity is conducted fairly [22]. In the context of information privacy, fairness refers to the perception that personal data are collected and used fairly. Culnan and Armstrong [22] proposed that perceptions of procedural fairness can help citizens to “strike a balance between the competing forces of privacy and information use.” Individuals’ perceptions of the fairness of an organization’s data collection and use practices can influence their decision-making related to technology use and information disclosure [23]. In this study, we investigated the drivers of trust and privacy through the procedural fairness lens, which suggests the importance of factors related to the legitimacy of data collection (ie, the need for government surveillance and perceptions of the government’s motive for the app), the costs to the citizen (ie, perceived intrusion), and the level of autonomy and input citizens are afforded (ie, perceived control). Furthermore, the wider literature on fairness and trust suggests that individual differences in citizens’ PTTT are likely to play an important role alongside procedural fairness perceptions in trust [24], particularly given the unfamiliar referent of the CT app. In addition, the theory of reasoned action (TRA) allows us to consider the influence of these perceptions on behavioral outcomes. The TRA argues that individuals’ behaviors are determined by their beliefs, attitudes, and intentions [25]. We propose that individuals will express positive intentions toward downloading the app and disclosing information if they believe that the app demonstrates fairness.

Hypotheses

Trust is an important factor in the success of CT apps as demonstrated in recent studies [14]. Indeed, trust allows individuals to overcome concerns about uncertainty and fosters a willingness to engage in trust-related behaviors, such as disclosing information and engaging with technology [26]. In this study, our emphasis is on technology-related trust, which refers to individuals’ beliefs that the technology in question will perform as expected [27]. Trust perceptions refer to the extent to which the CT app will consistently deliver the proposed services and act in citizens’ best interest.

A significant body of theoretical work suggests that variables related to trust propensity are important drivers of trust perceptions [28], particularly in new and unfamiliar trust referents [29]. PTTT refers to a general tendency that is not specific to one trustee or situation but focuses on individuals’ willingness to depend on technology across different contexts and technologies [27]. We propose that, in the current context, PTTT will positively influence trust in CT apps.

- **Hypothesis 1a:** PTTT will have a positive association with trust in the app.

Surveillance programs are often introduced following large-scale events such as terrorist attacks [30]. Given the public health emergency caused by the COVID-19 pandemic, the importance of surveillance technologies, such as CT apps, is clear. The link between surveillance and trust has long attracted discussion with Dutton et al [31], highlighting the existence of trust tension between the government’s need to collect surveillance data and citizens’ concerns about the excessive use of this information. They assert that developing trust is imperative for resolving this tension. As CT apps require the surveillance of large groups of people [32], citizens must understand the need for government surveillance in the general sense to build trust in a CT app. Need for government surveillance refers to individuals’ perceptions that the government requires authority to access personal information using web-based means [33].

- **Hypothesis 1b:** need for surveillance will have a positive association with trust in the app.

Procedural fairness theory suggests that if individuals believe that the government’s motivations to introduce the app are rooted in good intentions, such as reducing virus transmission, they will express higher trust in the app to perform consistently and with their best interests in mind. Indeed, a perception of benevolent motives is at the heart of theories regarding trustworthiness [28], and perceptions of trustworthiness at the government level are likely to trickle down to influence trust in related referents [34]. Accordingly, we posit that government motive will engender trust in the app.

- **Hypothesis 1c:** government motive will have a positive association with trust in the app.

Government surveillance technologies can garner negative opinions, which may lead individuals to alter their behaviors. Perceived intrusion is described as a “harmful incursion into the personal information space” [35]. This relates to procedural justice and individuals’ perceptions of whether data are collected fairly in the CT app. Government surveillance technologies can be viewed as intrusive, but not all intrusions are considered harmful [35]. For example, the CT app may be viewed as intrusive, but as data are collected to reduce virus transmission, some individuals may not view this as harmful. Thus, only if individuals believe that the intrusiveness of the app is harmful to them, then their trust in the app is likely to be reduced.

- **Hypothesis 1d:** perceived intrusion will have a negative association with trust in the app.

Studies have asserted that privacy concerns represent a barrier to the success of CT apps [14,36]. However, there is a lack of research examining if citizens are likely to accept CT apps when they perceive that they provide some level of privacy. As perceived privacy refers to a perception that access to personal information by external agents is limited, the relevance of perceived intrusion as a privacy cost resulting from the use of an app is apparent. Indeed, the potential of CT apps to violate citizens’ privacy has been raised [36]. Thus, if individuals believe that the app is intrusive in a harmful manner in their informational space, they are less likely to believe that the app affords them privacy.

- **Hypothesis 2a:** perceived intrusion will have a negative association with perceived privacy.
Perceived control is described as individuals’ perceptions of their ability to control their personal information [37]. If individuals perceive that they maintain control over their information when transacting with a technology, they are more likely to feel comfortable that the technology will not act in a harmful manner [37], strengthening their perception of privacy in that context. A recent study supports the positive association between perceived control and perceived privacy of CT apps in Brazil [17]. We proposed that if individuals believe that they maintain control in the app, they will express higher levels of perceived privacy.

- **Hypothesis 2b**: perceived control will have a positive association with perceived privacy.

Finally, we argue that from a theoretical perspective, trust in the app will act as an uncertainty-reducing mechanism [38,39] and a heuristic that allows citizens to form privacy perceptions. Specifically, if citizens believe that the app will perform consistently and with their best interests in mind when using personal information, they will believe that the app provides some degree of privacy. This uncertainty reduction provides a foundation for facilitating other judgments of a technological artifact. Thus, trust in the app may influence perceptions of privacy.

- **Hypothesis 2c**: trust in the app will have a positive association with perceived privacy.

The success of CT apps is largely dependent on a critical mass of people downloading [40] and disclosing personal information. Thus, we draw on the TRA to consider 2 context-critical dependent variables: adoption intention and willingness to disclose personal information. Before the app launch, these variables were behavioral intentions. Adoption intention is described as an individual’s internal subjective judgment of the probability that they will perform the behavior in question [25]. The willingness to disclose information is based on an individual’s willingness to provide personal information when using the app [33]. Trust has repeatedly been identified as a driver of behavioral intentions across a range of contexts, including the acceptance of e-government technologies [41] and surveillance [30]. Indeed, behavioral operationalizations of trust often use behavioral intentions related to disclosure and reliance [42]. Empirical evidence suggests that perceived privacy influences intentions toward the use and data disclosure in general CT apps [17]. Although our study focuses on the privacy perception of a government-led CT app as opposed to general CT apps, we argue for similar effects. Finally, if individuals express high intentions toward downloading the app, we argue that they will be more willing to disclose personal information, as it is crucial to the app’s functionality.

- **H3a-b**: trust in the app before the launch (a) and perceived privacy (b) will have a positive association with adoption intentions before the launch.
- **H4a-c**: trust in the app before the launch (a), perceived privacy (b), and adoption intentions before the launch (c) will have a positive association with disclosure intentions before the launch.

There have been calls for research to understand the perceptions of a technology before and after the launch [43]. In the context of CT apps, it is important to explore how both perceptions of privacy and trust influence individuals’ behavioral intentions after launch. Thus, we examined both intentions after the launch. As some individuals may have already downloaded the app, adoption intentions are represented by future use intentions, which encompasses intentions to continue use among app users and intentions to adopt in the future among nonusers. TRA asserts that intentions will lead to behavior [44]. In other words, individuals’ intentions to download the app before the launch will be positively related to their use intentions after the launch. We draw on the TRA to posit effects similar to those hypothesized for before the launch. We argue that trust perceptions regarding the app and perceived privacy will positively impact use intentions after the launch.

- **H5a-b**: trust in the app (a) and perceived privacy (b) will have a positive association with intentions to use after the launch.
- **H5c**: adoption intentions before the launch will have a positive association with intentions to use after the launch.

The influence of trust and privacy on the willingness to disclose a specific surveillance technology after the launch has not been explored. Again, we leverage the TRA intention-behavior link and our hypotheses before the launch and argue that trust and perceived privacy will positively impact willingness to disclose personal information after the launch. Finally, as was the case before the launch, we argue that if individuals express high intentions toward using the app after the launch, they will be more willing to disclose personal information.

- **H6a-b**: trust in the app (a) and perceived privacy (b) will have a positive association with disclosure intentions after the launch.
- **H6c**: intentions to use after the launch will have a positive association with disclosure intentions after the launch.

The hypotheses are depicted in Figure 1.
Methods

Study Context

On July 6, 2020, the COVID Tracker app was launched in Ireland. COVID Tracker has 3 main features. CT uses Bluetooth and anonymous ID numbers to log phones within close contact for more than 15 minutes. It downloads the anonymous ID numbers of people who have tested positive and provides an alert if the user has been in close contact with those ID numbers. Check-in allows users to check for symptoms and seek health advice. The updates provide an overview of the daily COVID-19 figures. Within 48 hours of launch, 1 million people had downloaded the app [45]. By November 2021, the COVID Tracker app had over 1.7 million active users, representing 67% to 76% of the total possible Irish users [46].

Instrument Development

We used existing scales when developing our instrument with minor wording amendments to adapt the items to the context. We provide the full list of items in Table S1 in Multimedia Appendix 1. The survey at T1 included general constructs related to PTTT, the need for government surveillance, and situationally framed constructs related to the proposed app, namely government motive for introducing the app, trust in the app, perceived surveillance in the app, perceived control in the app, and perceived privacy in the app. The dependent variables included intention to download the app on launch and willingness to disclose personal information to use the app. At T2, the emphasis was on future behavioral intentions, namely the use of the app and information disclosure. At T2, we asked participants if they had downloaded the app, and app users’ intentions to continue to use the app were examined and nonusers’ intentions to download the app in the future were examined. These intentions were combined as use intentions for analysis. Willingness to disclose personal information to the app was examined across both users and nonusers of the app. In addition, gender, age, and education were used as the control variables. Respondents were asked if they had any health condition that left them particularly susceptible to COVID-19. This was also a control variable. Both surveys were piloted and tested among a small panel of survey design experiments, and several wording amendments were made. Respondents were asked to answer demographic and health questions first, followed by general perceptual constructs and control variables, the order of which was randomized. In the third section, participants were presented with a neutrally framed description of the proposed national CT app at T1, and a description of the launched app was presented at T2. The final section examined perceptions of the app, behavioral intentions, and willingness to disclose personal data, the order of which was randomized.
Ethical Considerations

Ethics approval was obtained from the university’s research ethics committee before the launch of the survey (DCUREC/2020/096).

Recruitment

Qualtrics (Qualtrics International Inc) was used to host and administer the survey using their panel services. An attention check was included to screen for unengaged responses. A total of 1109 complete responses were received at T1 and were recontacted at T2. After 2 follow-up invitations, 405 responses were received at T2, achieving a response rate of 36.5%. Responses at T1 and T2 were, on average, 77 days apart. Incomplete responses and responses failing the attention check were removed using Qualtrics. The sample characteristics are illustrated in Table 1, along with the population characteristics as per the latest census at the time of data collection. Overall, the sample demographics were similar to the population characteristics of Ireland, as reported in the last census, and included respondents from the 26 counties within the country. Of the 405 respondents in T2, 202 had downloaded the app and 203 had not yet downloaded the app.

Table 1. Sample and population characteristics (N=405).

<table>
<thead>
<tr>
<th>Sample, n (%)</th>
<th>Population (%)&lt;sup&gt;a,b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>180 (44.4)</td>
</tr>
<tr>
<td>Woman</td>
<td>225 (55.6)</td>
</tr>
<tr>
<td>Rather not say</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>13 (3.2)</td>
</tr>
<tr>
<td>25-44</td>
<td>124 (30.6)</td>
</tr>
<tr>
<td>45-64</td>
<td>173 (42.7)</td>
</tr>
<tr>
<td>≥65</td>
<td>95 (23.5)</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>186 (45.9)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>26 (6.4)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>36 (8.9)</td>
</tr>
<tr>
<td>Student</td>
<td>11 (2.7)</td>
</tr>
<tr>
<td>Unavailable for work</td>
<td>42 (10.4)</td>
</tr>
<tr>
<td>Retired</td>
<td>104 (25.7)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Secondary school</td>
<td>157 (38.8)</td>
</tr>
<tr>
<td>Trade</td>
<td>5 (1.2)</td>
</tr>
<tr>
<td>Diploma</td>
<td>32 (7.9)</td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>133 (32.8)</td>
</tr>
<tr>
<td>Other qualification</td>
<td>64 (15.8)</td>
</tr>
<tr>
<td>Doctorate degree</td>
<td>14 (3.5)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Population figures are based on data provided by the Irish Central Statistics Office in the latest population census at the time of data collection (ie, 2016).

<sup>b</sup>Employment and education figures include all people aged ≥15 years living in Ireland in 2016, whereas our sample only includes people aged ≥18 years.

<sup>c</sup>N/A: not applicable.

Results

Reliability and Validity Testing

Data analysis was performed using IBM AMOS (version 25.0). The proposed model comprising 11 constructs was examined using Confirmatory Factor Analysis with further detail provided in Table S2 in Multimedia Appendix 1. In total, 3 items were dropped from the PTTT because of their low loadings. The model indicated a good fit: $c_{min}/df=1.805$, comparative fit index=0.980, root mean square error of approximation=0.045, and standardized root mean squared residual=0.034. A test of equal specific bias was conducted to examine potential common method bias among the data [47]. This test demonstrated an unevenly distributed bias; thus, the specific bias construct was...
Hypotheses Testing
The causal model was tested using Structural Equation Modeling (SEM) in AMOS. The model indicated a good fit \( \chi^2_{min}/df \approx 2.835 \), comparative fit index=0.985, root mean square error of approximation=0.067, and standardized root mean squared residual=0.021. H1a-d focused on the antecedents of trust in the app. H1a proposed a positive relationship between PTTT (Propensity to trust technology) and trust. Hypothesis 1b posited that the perceived need for government surveillance would positively influence trust. This was also supported (hypothesis 1b: \( \beta = 0.119; P < 0.001 \)). H1c posited that government motive would be positively related to trust. The data supported the hypothesis (hypothesis 1c: \( \beta = 0.671; P < 0.001 \)). The negative relationship between perceived intrusion and trust was supported (hypothesis 1d: \( \beta = -0.224; P < 0.001 \)). The following set of hypotheses examined the antecedents of perceived privacy. Hypothesis 2a proposed a negative association between perceived intrusion and perceived privacy. These data supported hypothesis 2a (\( \beta = -0.165; P < 0.001 \)). We hypothesized that perceived control and trust would be positively related to perceived privacy. Both relationships were supported (hypothesis 2b: \( \beta = 0.451; P < 0.001 \); H2c: \( \beta = 0.466; P < 0.001 \)).

In terms of T1 behavioral intentions, it was posited that perceived privacy and trust would positively influence the intention to adopt. Both relationships were supported (hypothesis 3a: \( \beta = 0.247; P < 0.001 \); hypothesis 3b: \( \beta = 0.590; P < 0.001 \)). H4a-c proposed that trust, perceived privacy, and adoption intentions would positively influence willingness to disclose personal information. The data revealed that trust (\( \beta = 0.215; P < 0.001 \)), perceived privacy (\( \beta = 0.208; P < 0.001 \)), and adoption intentions (\( \beta = 0.550; P < 0.001 \)) each positively influenced willingness to disclose. In terms of T2 behavioral intentions, hypothesis 5a to 5c proposed that trust, perceived privacy, and adoption intentions (T1) would all influence use intentions (T2). Both trust (\( \beta = 0.124; P = 0.15 \)) and perceived privacy (\( \beta = 0.042; P = 0.60 \)) had a positive but nonsignificant influence on intentions. T1 adoption intentions significantly influenced use intentions.
supporting hypothesis 5c ($\beta =.530; P<.001$). Finally, trust, perceived privacy, and use intentions at T2 were proposed to positively influence the willingness to disclose at T2. Perceived privacy ($\beta =.042; P=.40$) had a nonsignificant influence, whereas trust and use intentions had significant relationships supporting hypothesis 6a and hypothesis 6c (hypothesis 6a: $\beta =.250; P<.001$; H6c: $\beta =.655; P<.001$). In terms of control variables, COVID-19 vulnerable illness had a significant negative effect on individuals’ willingness to disclose at T1 ($\beta =-.043; P=.009$), and education had a positive effect on T2 use intentions ($\beta =.075; P=.04$).

The model explains 74.4% of variance in trust, 91% of variance in perceived privacy, 66.6% of variance in T1 adoption intentions, 45.9% of variance in T2 adoption intentions, and 83.9% and 79.4% of variance in willingness to disclose at T1 and T2. Bootstrapping using 2000 samples and a confidence level of 90% was conducted in AMOS to explore the indirect effects. The findings revealed that perceived privacy had a significant influence on T2 adoption intentions ($\beta =.131; P=.001$) and willingness to disclose at T2 ($\beta =.127; P=.04$). Similarly, trust had a significant influence on intention to download ($\beta =.394; P<.001$) and willingness to disclose at T2 ($\beta =.386; P<.001$). Further detail is provided in Table S3 in Multimedia Appendix 1.

**Discussion**

**Principal Findings**

This study focuses on understanding how citizens’ beliefs shape their perceptions of privacy and trust to influence their acceptance of a CT app for COVID-19. Our study found that trust in the app was positively influenced by the PTTT, perceived government motive, and perceived need for government surveillance, whereas perceived intrusion had a negative influence. Perceived privacy was positively shaped by perceptions of control and trust and negatively shaped by perceived invasion. The study examined citizens’ acceptance of CT app at 2 time intervals. Before launch, the intention to adopt the app was positively influenced by trust and perceived privacy, and willingness to disclose personal information to the app was influenced by adoption intentions, trust, and perceived privacy. However, postlaunch use intentions were influenced only by prelaunch adoption intentions, whereas willingness to disclose personal information was influenced by trust and postlaunch use intentions but not by perceived privacy. Although the insignificant results may suggest that perceived privacy is only important before launch, and the influence of trust on use intentions diminishes over time, post hoc bootstrapping analysis revealed that both perceived privacy and trust had significant indirect relationships with use intentions and willingness to disclose information at T2. This suggests that both perceptions play a role in influencing behavioral intentions before and after the launch.

**Contributions**

Studies have shown that trust and privacy are important factors in the success of health surveillance technologies such as CT apps [11]. Our study leverages the procedural fairness theory to understand how citizens’ perceptions of trust and privacy emerge in the context of a CT app. This context is interesting, as the technology in question was introduced by the national government and backed by several organizations with the app’s potential benefits extending to the public at large. Therefore, it is important to look beyond the role of a single organization in driving perceptions of fairness to consider a broader set of antecedents that drive perceptions of trust and privacy in this context. Indeed, as research has shown the importance of trust and privacy in the success of mHealth and health technologies introduced by health care organizations and indeed national health systems [51], our study contributes to the broader health technology literature.

The first contribution of our study is the deeper understanding of how trust is formed in this context. Lack of trust in the government has been identified as a barrier to CT app adoption [18]. Thus, it is important to provide governments and public health organizations with insights into how trust in CT can be fostered [8]. Our findings bolster assertions regarding the important role of fairness perceptions and suggest that citizens’ trust perceptions regarding the app are formed based on their beliefs about the legitimacy of data collection, perceived autonomy, and perceived costs. Legitimacy is represented by citizens’ perceptions of the need for government surveillance and perceptions of the government’s motive for introducing the app, autonomy is captured by perceptions of control over one’s information in the app, and perceived costs to the individual relate to perceptions of personal intrusion.

The second contribution of our study is the investigation of how perceptions of privacy are formed. Many studies have highlighted the negative impact of privacy concerns on CT adoption [11,52-54]. However, we argue that privacy can be seen as a factor driving adoption if citizens believe that apps can afford them with some level of privacy. Our findings demonstrate that citizens’ privacy perceptions are shaped by trust in the app, which encompasses legitimacy perceptions and their perception of control offered by the app, and are negatively influenced by perceptions of intrusion. By highlighting the importance of fairness perceptions and elucidating the role of several perceptual factors at the governmental level (need for surveillance and government motives) and the app level (perceived intrusion and perceived control), which have been sparingly studied to date, our study advances our understanding of how privacy perceptions are developed in this context.

Understanding the factors driving CT app acceptance is paramount for future outbreaks [8]. The third contribution relates to understanding how citizens form intentions toward CT apps at different stages of the implementation process. Our study provides support for the influence of perceived privacy on individuals’ intentions to adopt an app and willingness to disclose personal information before launch and support for an indirect influence over time on future use and willingness to disclose data. This finding supports many studies that found that privacy concerns negatively impacted adoption intentions toward mHealth [51]. In the context of a national CT app, if individuals perceive that the app offers a sufficient state of privacy, they will express positive intentions toward adoption and information disclosure before and after the app launch. Put
simply, perceived privacy can have a sustained positive influence on behavioral intentions.

Trust has been widely studied within the privacy and other domains to understand individuals’ intentions to disclose information [42]. Trust in the app was found to influence individuals’ adoption and disclosure intentions before launch, indirectly influencing use intentions after launch and directly influencing disclosure intentions after launch. These findings suggest that the influence of initial trust perceptions prevails over time and may operate as a heuristic for interacting with the app on an ongoing basis. The stability of trust perceptions and their ongoing influence are a relatively nascent topic, although some theorists have suggested the possibility of trust as a heuristic [55]. Our research provides empirical evidence for this phenomenon and offers further support for claims that initial trust perceptions might be relatively robust and long acting [56].

**Implications for Practice**

The findings of this study have several practical implications. First, the trust tension between public good and the intrusiveness of surveillance technologies has led researchers to emphasize the importance of effective trust-building strategies when introducing surveillance programs [31]. Indeed, citizens in the United States and Germany have expressed concerns regarding possible surveillance stemming from CT apps [9]. Our study shows that citizens’ perceptions of trust and privacy can be influenced by fairness perceptions based on their beliefs regarding the need for surveillance and the government’s motives for introducing surveillance technologies; the perceived control they are offered over their personal information and negatively influenced by their perceptions of the intrusiveness of these technologies. Thus, governments should focus on transparency in their public health surveillance efforts, including the involvement of data protection authorities and civil liberties advocates throughout the project life cycle, potentially through a privacy advisory committee [57]. This transparency should be extended to communication with citizens on the need and purpose of a technology while stressing the control they have over their personal information. Our research suggests that early communications that shape first impressions are particularly important. Such practices not only comply with data protection laws, such as the EU (European Union) General Data Protection Regulation, but also foster a sense of trust and ultimately influence the use of technology.

Second, our findings highlight the positive influence of privacy perceptions on adoption and disclosure. Thus, we argue that privacy should not be viewed as a barrier to new technologies, such as mHealth or CT apps, but rather as an important consideration throughout the design, implementation, and postlaunch stages. Designers should ensure compliance with the regulatory requirements for consent and control. Governments and other organizations charged with introducing new technologies should ensure that they clearly communicate their compliance with regulations and the considerations of individuals’ personal data. Given that CT apps provide data on the location, copresence, and potentially ePHI of not only the focal person but also others that they have been in contact with, the principles of both necessity and proportionality would appear to be key. As per Lenca and Vayena [4], data collection must (1) be proportional to the seriousness of the public health threat, (2) be limited to what is necessary to achieve a specific public health objective, and (3) be scientifically justified. Policy makers and public health decision makers need to consider what communication and control mechanisms can be introduced to (1) build trust with the public and (2) repair trust, if necessary. This includes declaring what data will be collected and used while the app is live and by whom, confirming that data have been deleted, when no longer relevant (as is the case with COVID-19 data) or once the app is no longer required.

Third, in the context of technologies that require mass acceptance and willingness to share personal data, the focus cannot be placed solely on the number of downloads but must account for actual use and disclosure behaviors. Individuals’ intentions to download CT apps influence their willingness to disclose personal information both before and after launch. Once they have downloaded the app, it is critical that decision makers encourage use and that the widespread use of the app is linked, through public communication, to successful intervention strategies so that the benefits to the individual and society are reinforced.

**Limitations and Directions for Future Research**

This study has several limitations. First, other factors may influence privacy and trust perceptions or moderate the relationships between trust and privacy and adoption. Although it is not possible to consider all potential antecedents and intervening variables, it would be interesting to explore the role of other prominent perceptions, such as perceived sensitivity, as these apps require users to disclose identifying information, health information, and location information, all of which are arguably sensitive. Second, although our study considers 2 important technology use outcomes, before and after the launch, this approach has limitations. First, the collection of data from the same respondents at multiple time points inevitably led to a drop in responses. Although we sent repeated invitations during the second phase of data collection, the final sample that completed both surveys was smaller. Although this is commonplace within this approach and our sample characteristics are similar to the broader population of Ireland, we acknowledge that a large sample would be ideal and stress the importance of considering the sample size when drawing inferences from our study. Second, our 2 time points did not allow us to take full advantage of the potential to model longitudinal change trajectories over time. Further work is needed to incorporate time more fully into our understanding of how privacy and trust influence adoption and use behaviors.

Third, our study relies on individuals’ self-reported adoption and disclosure intentions. This approach is commonplace in the privacy and technology adoption literature streams, and it would not have been possible to study actual behaviors. However, we must acknowledge that intentions are not always matched with behaviors and that information disclosure is not always accurate or true. In other settings, it may be more feasible to collect more objective behavioral data, and we would encourage researchers to do so, particularly in settings where widespread adoption is...
required for success. In addition, studies may go beyond our focus to understand disclosure behaviors at a deeper level and examine privacy-protective behaviors, such as withholding information or falsifying information. These protective behaviors are potentially dangerous in contexts such as CT apps because of the reliance on accurate data to track virus transmission.

Finally, our study explores a public health surveillance context where the focal person volunteers to participate and therefore has notice of the surveillance, control of their data and gives explicit consent. There are several conditions under which public health surveillance, including name reporting, may be undertaken without notice or explicit patient consent with well-established justifications in public health ethics, science, and law [58]. Even in the context of COVID-19, digital CT has not always been voluntary. In China, there is evidence of digital CT without notice or consent [59]. Furthermore, even when the focal person has notice and gives consent, contacts of the focal person have not given explicit consent. Although the primary focal person is subject to direct active surveillance, the secondary focal person is subject to passive indirect surveillance. In addition to the ethical issues that such practices raise, particularly where there is coordination and data exchange between private firms and the government [60], testing the theoretical framework developed in this study in this new context may provide a fruitful avenue of research. Similarly, aggregated anonymous spatiotemporal data sourced from commercial providers have been used as proxies for human movement and social interaction and as indicators of the effectiveness of social distancing interventions [61]. Although these data are currently anonymous, governments have already mandated access to identifiable data on the basis that the public interest overrides privacy rights [62]. This context may provide interesting insights and further extend our understanding of the limits of consumer acceptance of governmental health surveillance.

Conclusions

The COVID-19 pandemic was the first time governments implemented large-scale digital CT. Its success as a public health intervention depended on rapid technology adoption by a significant proportion of the population. Here, surveillance is active, and the target of government surveillance through COVID-19 digital CT apps is an active participant in sharing data with the government on their personal health status, their location, and often their social network. The opportunity to study such an empirical context is not only rare but also the time frame for research is limited. Understanding the formation of individuals’ perceptions of trust and privacy in this context and how these perceptions influence their acceptance of digital CT apps is critical not only for informing the design of future digital CT initiatives but also for other situations that require rapid digital technology adoption by a significant proportion of society. If governments wish to leverage the power of digital technologies to control future public health threats, we recommend 3 principles to guide the design of both their surveillance initiatives and communications with the public—necessity, transparency, and proportionality—before and after the launch.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey items, validity, and reliability testing.

[PDF File (Adobe PDF File), 172 KB - mhealth_v12i1e48700_app1.pdf]

References


Abbreviations

AVE: average variance extracted
CT: contact tracing
ePHI: electronic personal health information
EU: European Union
mHealth: mobile health
PCT: Privacy Calculus Theory
PTTT: propensity to trust technology
TRA: theory of reasoned action

Edited by L Buis; submitted 10.05.23; peer-reviewed by Z Zrubka, M Shoji; comments to author 18.07.23; revised version received 20.10.23; accepted 06.12.23; published 18.01.24.

Please cite as:
Fox G, van der Werff L, Rosati P, Lynn T
Investigating Citizens’ Acceptance of Contact Tracing Apps: Quantitative Study of the Role of Trust and Privacy
JMIR Mhealth uhealth 2024;12:e48700
URL: https://mhealth.jmir.org/2024/1/e48700
doi:10.2196/48700
PMID:38085914

©Grace Fox, Lisa van der Werff, Pierangelo Rosati, Theo Lynn. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Abstract

Background: Complete documentation of critical care events in the accident and emergency department (AED) is essential. Due to the fast-paced and complex nature of resuscitation cases, missing data is a common issue during emergency situations.

Objective: This study aimed to evaluate the impact of a tablet-based resuscitation record on documentation completeness during medical resuscitations and nurses’ perceptions of the use of the tablet app.

Methods: A mixed methods approach was adopted. To collect quantitative data, randomized retrospective reviews of paper-based resuscitation records before implementation of the tablet (Pre-App Paper; n=176), paper-based resuscitation records after implementation of the tablet (Post-App Paper; n=176), and electronic tablet-based resuscitation records (Post-App Electronic; n=176) using a documentation completeness checklist were conducted. The checklist was validated by 4 experts in the emergency medicine field. The content validity index (CVI) was calculated using the scale CVI (S-CVI). The universal agreement S-CVI was 0.822, and the average S-CVI was 0.939. The checklist consisted of the following 5 domains: basic information, vital signs, procedures, investigations, and medications. To collect qualitative data, nurses’ perceptions of the app for electronic resuscitation documentation were obtained using individual interviews. Reporting of the qualitative data was guided by Consolidated Criteria for Reporting Qualitative Studies (COREQ) to enhance rigor.

Results: A significantly higher documentation rate in all 5 domains (ie, basic information, vital signs, procedures, investigations, and medications) was present with Post-App Electronic than with Post-App Paper, but there were no significant differences in the 5 domains between Pre-App Paper and Post-App Paper. The qualitative analysis resulted in main categories of “advantages of tablet-based documentation of resuscitation records,” “challenges with tablet-based documentation of resuscitation records,” and “areas for improvement of tablet-based resuscitation records.”

Conclusions: This study demonstrated that higher documentation completion rates are achieved with electronic tablet-based resuscitation records than with traditional paper records. During the transition period, the nurse documenters faced general problems with resuscitation documentation such as multitasking and unique challenges such as software updates and a need to familiarize themselves with the app’s layout. Automation should be considered during future app development to improve documentation and redistribute more time for patient care. Nurses should continue to provide feedback on the app’s usability and functionality during app refinement to ensure a successful transition and future development of electronic documentation records.

(JMIR Mhealth Uhealth 2024;12:e46744)  doi:10.2196/46744
Introduction

Background

The completeness of documentation of critical care events in the accident and emergency department (AED) is essential for (1) the continuity of patient care, (2) medicolegal issues [1], (3) improving accessibility to critical information needed for research [2], and (4) serving as evidence for quality outcome measures [3]. Traditionally, documentation is performed on paper. Due to the fast-paced and complex nature of resuscitation cases, missing data is a common issue during emergency situations. As much as 60% of essential data fields in prehospital paper records can be incomplete [4]. A study in a trauma center also found incompleteness in 18% of the mandatory elements for trauma resuscitation [5].

Over the last 2 decades, there has been a global trend of switching to electronic medical records (EMRs). It was estimated that about 46% of AEDs in the United States used EMRs in 2010 [6]. This percentage is expected to increase in the future. Fully functional EMRs have been shown to improve efficiency in AEDs [7]. Despite the growth in usage, very few studies have explored the impact of EMRs in AED settings. Furthermore, the perceptions toward EMRs are mixed. A study found that nurses and physicians generally had a negative perception toward EMRs in the AED. EMRs are considered to be ineffective, redundant, and prone to error [8]. In contrast, another study found that nurses perceived that their productivity increased and care was better coordinated after implementing EMRs [9].

Among the limited studies in the area, a retrospective review of trauma resuscitations in AED settings showed that EMRs can improve documentation completeness [5,10]. However, there is a lack of studies on medical resuscitations, which are more common than trauma resuscitations [11]. Medical resuscitations are performed on triage category I and II patients with life-threatening conditions such as myocardial infarction, sepsis, and stroke [12]. Medical resuscitations differ from trauma resuscitations in that they do not follow a single protocol. The differences in management protocols can make the process of documenting medical resuscitations different from that of trauma resuscitations.

The EMR system used by the Hospital Authority of Hong Kong is called the Clinical Management System (CMS). It is an integrated platform that allows clinical users to manage the following daily clinical activities [13]: (1) obtain clinical data including consultation notes, laboratory, and imaging results; (2) document clinical activities; and (3) provide clinical decision support.

In AEDs in Hong Kong, medical records are still part paper and part electronic, with the patient’s clinical notes being documented on paper. This type of mixed documentation has been shown to hinder effective communication and utilization of information in either record [14].

Recently, there has been a trend of switching to electronic documentation in AEDs. In 2020, 3 of 18 AEDs in Hong Kong had switched to an EMR system called the eAED. It was expected that, by 2023, about two-thirds of AEDs would have switched to the eAED. The eAED is meant to replace paper charts previously used to document a patient’s progress [15]. Despite the gradual adoption of EMRs in AEDs in Hong Kong, the use of electronic documentation during medical resuscitations has not occurred owing to the time-critical, fast-paced nature and lack of a suitable application.

However, with advances in computer processing power, a tablet-based system could fill the gap. Documentation efficiency and data precision have improved when a tablet-based app was used, in comparison with a desktop EMR, during a simulation [16]. In Hong Kong, a tablet-based system called “eResus” is being developed by the Hong Kong Hospital Authority for medical and trauma resuscitation documentation. With the implementation of the eAED and eResus, documentation in AEDs would become fully electronic.

Aim and Objectives

This study aimed to evaluate the impact of a tablet-based app on documentation completeness during medical resuscitations. The research questions were the following:

- What are the differences between paper and electronic tablet-based records on the levels of documentation completeness?
- What are the perceptions of emergency room nurses regarding documentation completeness when using eResus?

Hypothesis

This study hypothesized that the completeness of resuscitation documentation using electronic tablet-based records would be higher than that that using paper records.

Methods

Design

To answer research question 1, a randomized retrospective review of paper and electronic resuscitation medical records (N=528) was conducted using a documentation completeness checklist. The study was implemented during the transition from paper to electronic documentation, when only triage category II cases would be documented using the tablet-based app called eResus. Therefore, triage category II records were collected before (from November 2020 to December 2020) and after (from February 2021 to March 2021) implementation of the tablet-based eResus app. Paper records were collected before (Pre-App Paper) and 1 month after (Post-App Paper) implementation of eResus, while electronic records were collected 1 month after (Post-App Electronic) implementation of eResus. We randomly selected 176 records each for the

KEYWORDS

tablet computer; nursing documentation; paper resuscitation record; electronic resuscitation record; medical resuscitation; electronic medical record; documentation; resuscitation; electronic health record; nurses’ perception; traditional paper record; nurse

To answer research question 2, emergency nurses’ perceptions of the advantages, challenges, and areas for improvement of the electronic app for resuscitation documentation were obtained in individual interviews conducted in mid-April 2021, 3 months after the implementation of eResus. Reporting of the qualitative findings was guided by the Consolidated Criteria for Reporting Qualitative Studies (COREQ) [17], as delineated in the following sections, to enhance rigor.

Ethical Considerations

Ethical clearance (NTWC/REC/20098) from the study hospital and the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University (HSEARS20200826001) was obtained before the commencement of the study.

Data

Quantitative Data

A documentation completeness checklist was established based on the literature and a review of the department’s current medical resuscitation event documentation. The checklist consisted of 5 essential domains (ie, basic information, vital signs, procedures, investigations, and medications) of medical resuscitation as illustrated in Multimedia Appendix 1. Face and content validity of the checklist were determined by 4 experts in the emergency medicine field [18]. Experts were invited based on the following criteria: (1) worked in an AED and (2) published at least one article related to the accident and emergency field. The expert panel consisted of 1 associate consultant, 1 medical officer, and 2 advanced practice nurses (1 of which was a Fellow in Emergency Nursing). The content validity index (CVI) was calculated using the scale CVI (S-CVI). The S-CVI is calculated based on the number of items in the scale rated by the expert as “quite relevant” or “highly relevant” [19]. The S-CVI was further analyzed by universal agreement (UA) among experts (S-CVI/UA) and the average (S-CVI/Ave). The checklist’s S-CVI/UA was 0.822, and the S-CVI/Ave was 0.939.

The medical resuscitation documents were reviewed against the validated checklist by a researcher (CSY), and intrarater reliability was determined to ensure consistency. Intrarater agreement was calculated using the Cohen kappa [19]. We evaluated 5 cases at week 0 and week 2. The agreement between the 2 records was considered acceptable at a $\kappa$ of 0.884 (95% CI 0.671-1.105; $P<0.001$).

For each resuscitation documentation review, the researcher provided a dichotomous response of “Yes or No” for each item and identified the level of completeness based on the checklist. The patient diagnosis, length of medical resuscitation, initial triage category, and demographics including age and gender were also collected as part of the basic information. However, patient names and identification numbers were not collected. Data were kept anonymous by assigning codes only identifiable to the researcher.

Qualitative Data

Emergency room nurses were guided to discuss their thoughts on the eResus app’s features for documentation completeness through individual interviews with an onsite nurse who was one of the researchers (CSY). They understood the aim of the study, and their experience with the app was explored. Participants’ demographic data including age, gender, years of experience after graduation, and years of experience in their current specialty were collected for subsequent data analysis.

Sample Size

For the quantitative data, the Chi-square test was used to compare the differences in documentation completeness between the 3 groups. Based on the findings from a level 1 pediatric trauma center in 2015 [10], with an $\alpha$ of .05 and power of 0.80, a minimum sample size of 153 medical records per record set was required. To ensure an adequate sample size, 176 patient records were included in each of the paper and electronic record sets, resulting in a total sample size of 528 records (ie, 176 records each for the Pre-App Paper, Post-App Paper, and Post-App Electronic record sets).

For qualitative data, data saturation is the criterion to determine the sample size. Data are considered saturated when no new theoretical insights are gained from new data [20]. For this study, data saturation was achieved after 10 individual interviews, and 2 more interviews were conducted to confirm the data saturation.

Recruitment

The study was conducted in the AED of 1 hospital in Hong Kong. It is one the major local trauma centers providing 24-hour accident and emergency services and serves more than 190,000 patients per year, with over 300 resuscitation cases per month [21]. The tablet app was scheduled to be implemented in June 2020 but was postponed due to COVID-19. The app was eventually implemented in January 2021.

Quantitative Data Collection Method

In this study, we reviewed 2 types of resuscitation documents, namely paper and tablet-based resuscitation records. Completed resuscitation documents in paper format were attached to the patient’s CMS record by optical scanning as per usual practice. These records were stored in the CMS.

Training prior to the implementation of the electronic resuscitation record could lead to bias toward improved completeness of electronic documents [10]. Therefore, paper documentation records were collected before and 1 month after implementation of eResus to address this issue. First, baseline paper resuscitation records were collected prior to implementation of the eResus app (Pre-App Paper). After the implementation of the eResus app with training, there was a washout period of 1 month. After 1 month, the paper (Post-App Paper) and tablet-based (Post-App Electronic) resuscitation records were retrieved for analysis. Both paper and tablet resuscitation records involving trauma team activation or triage category 1 cases were excluded because the app did not cover these 2 types of cases at this stage.
For the purpose of this study, 3 lists of 3 months of case lists including eligible medical records from the Pre-App Paper, Post-App Paper, and Post-App Electronic record sets were retrospectively generated from the CMS and assigned a serial number. From each group, 176 records were randomly selected using a random number generator.

**Qualitative Data Collection Method**

In terms of qualitative data collection, purposive sampling was applied. The researcher conducted individual, voice-recorded interviews with each emergency nurse 3 months after the eResus implementation. The nurses were provided an explanation of the study, and written consent was obtained. The inclusion criteria included nurses (1) working in the AED of the hospital, (2) with experience using the eResus app, (3) who spoke Cantonese and were able to read English, and (4) working in their current position for more than 3 months.

Invitation emails were sent to colleagues. Eligible colleagues who replied to the email or expressed interest were invited to be interviewed according to their years of experience. Individual interviews were conducted in a quiet room or via Zoom. Each interview lasted about 1 hour or stopped when the interviewee felt that their viewpoint had been fully expressed. The interview guide is shown in Textbox 1.

**Textbox 1. Interview guide.**

<table>
<thead>
<tr>
<th>Opening question:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can you tell me your experience with using the eResus app until now?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Guiding questions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What are the main advantages and challenges with achieving high documentation completeness when using eResus in the resuscitation room?</td>
</tr>
<tr>
<td>2. How do you think eResus can be improved to help you achieve better documentation completeness?</td>
</tr>
</tbody>
</table>

**Data Analysis Methods**

For the quantitative data analysis, SPSS version 25 (IBM Corp) was used. Descriptive statistics such as means, standard deviations, frequencies, and percentages were used to present the study variables. Normality was tested using the Kolmogorov-Smirnov test, and the data were found to be not normally distributed. The Mann Whitney U test was used to compare mean ranks for age and clinical characteristics between the Pre-App Paper and Post-App Paper record sets as well as between the Post-App Paper and Post-App Electronic record sets to ensure the clinical characteristics of the 3 groups were comparable. Subsequently, Chi-square tests were used to compare the differences in proportions, such as the percentage of completeness between the Pre-App Paper and Post-App Paper record sets to determine any historical bias or effect from training and then between the Post-App Paper and Post-App Electronic record sets. Results with a P value <.05 were considered significant.

Each resuscitation record was manually reviewed against the study checklist for data element completeness. Each record was reviewed individually. Any incomplete data element was entered as an incomplete domain for the respective domain of the 5 domains, namely basic information, vital signs, procedures, investigations, and medications. For example, for records of the administration of 2 medications that use the same route, the record was treated as 2 separate data entries. If 1 of the data items (such as 1 missing medication name) was incomplete, the medication domain for that case was entered as incomplete. The number of entries for each domain of the resuscitation documentation was analyzed, delineating sections that were recorded at higher or lower frequencies.

For the qualitative data analysis, content analysis was performed [22]. First, the interview was audio-recorded and transcribed verbatim into Chinese. NVivo Pro 12 was used for data analysis. The researcher read through the transcript multiple times to become immersed in the data. Participants’ experiences with the eResus app, challenges, and possible solutions were extracted and summarized into meaning units. Third, each meaning unit was condensed and labelled with codes. Fourth, subcategories were identified by comparing the similarities and differences between different codes. Finally, the latent meanings of the subcategories were sorted into themes.

**Results**

**Quantitative Results**

Tables 1 and 2 present the characteristics of the 5 domains of resuscitation documentation. Comparisons were made between the Pre-App Paper and Post-App Paper record sets. There were no significant differences in characteristics or documentation completion between the Pre-App Paper and Post-App Paper record sets (Table 3).
Table 1. Comparisons using the Mann-Whitney U test among the 3 resuscitation record sets in patient age; length of resuscitation; and total numbers of vital sign entries, procedures, investigations, and medications in the resuscitation documentation for patients requiring medical resuscitation (N=528).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic information</td>
<td>Range</td>
<td>Mean (SD)</td>
<td>Range</td>
<td>Mean (SD)</td>
<td>U</td>
</tr>
<tr>
<td>Age (years)</td>
<td>4-98</td>
<td>61.6 (20.6)</td>
<td>3-101</td>
<td>61.6 (22.0)</td>
<td>15,121</td>
</tr>
<tr>
<td>Length of resuscitation (minutes)</td>
<td>3-216</td>
<td>40.6 (30.2)</td>
<td>5-175</td>
<td>41.7 (28.6)</td>
<td>14,673</td>
</tr>
<tr>
<td>Total number of vital sign entries</td>
<td>1-58</td>
<td>10.9 (8.0)</td>
<td>2-47</td>
<td>11.1 (7.1)</td>
<td>15,089</td>
</tr>
<tr>
<td>Total number of procedures</td>
<td>0-7</td>
<td>1.83 (1.34)</td>
<td>0-8</td>
<td>1.76 (1.26)</td>
<td>15,050</td>
</tr>
<tr>
<td>Total number of investigations</td>
<td>0-11</td>
<td>5.72 (1.89)</td>
<td>0-11</td>
<td>5.49 (1.92)</td>
<td>15,091</td>
</tr>
<tr>
<td>Total number of medications</td>
<td>0-12</td>
<td>1.65 (2.07)</td>
<td>0-15</td>
<td>1.70 (2.31)</td>
<td>15,085</td>
</tr>
</tbody>
</table>

Table 2. Gender differences among the 3 groups of resuscitation records for patients requiring medical resuscitation (N=528), as assessed using the Chi-square test.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Pre-App Paper record set (n=176), n (%)</th>
<th>Post-App Paper record set (n=176), n (%)</th>
<th>Post-App Electronic record set (n=176), n (%)</th>
<th>Difference among the groups</th>
<th>( \chi^2 ) (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>84 (47.7)</td>
<td>84 (47.7)</td>
<td>86 (48.9)</td>
<td>0.06 (2)</td>
<td>.97</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Differences in completion of the 5 domains of documentation between paper and electronic resuscitation records (N=528).

<table>
<thead>
<tr>
<th>5 domains</th>
<th>Pre-App Paper record set (n=176), n (%)</th>
<th>Post-App Paper record set (n=176), n (%)</th>
<th>Post-App Electronic record set (n=176), n (%)</th>
<th>Differences among groups</th>
<th>( \chi^2 ) (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic information</td>
<td>113 (64.2)</td>
<td>105 (59.7)</td>
<td>128 (72.7)</td>
<td>6.86 (2)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Vital sign</td>
<td>116 (65.9)</td>
<td>108 (61.4)</td>
<td>158 (89.8)</td>
<td>40.97 (2)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Procedures</td>
<td>123 (69.9)</td>
<td>127 (72.2)</td>
<td>176 (100)</td>
<td>63.50 (2)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Investigations</td>
<td>101 (57.4)</td>
<td>93 (52.8)</td>
<td>128 (72.7)</td>
<td>16.06 (2)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Medications</td>
<td>158 (89.8)</td>
<td>163 (92.6)</td>
<td>175 (99.4)</td>
<td>15.24 (2)</td>
<td>&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>

For the post-app comparison, there were no significant differences in the characteristics, except the number of vital sign entries, between the Post-App Paper and Post-App Electronic record sets (Tables 1 and 2). To answer research question 1, there was a significantly higher completion rate for all 5 domains in the Post-App Electronic record set than in Post-App Paper record set (Table 3).

Qualitative Results

Categories

The objective of the qualitative study was to explore nurses’ perceptions of the use of eResus for documentation completeness. The main categories identified were “advantages of tablet-based documentation of resuscitation records,” “challenges with tablet-based documentation of resuscitation records,” and “areas for improvement of tablet-based resuscitation records” (Textbox 2).
Textbox 2. Summary of the categories and subcategories.

<table>
<thead>
<tr>
<th>Advantages of tablet-based documentation of resuscitation records</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Structural guidance for documentation</td>
</tr>
<tr>
<td>• Easy to review and edit documentation</td>
</tr>
<tr>
<td>• Comparable mobility to paper and superior to desktop</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Challenges with tablet-based documentation of resuscitation records</th>
</tr>
</thead>
<tbody>
<tr>
<td>• System loading speed and stability</td>
</tr>
<tr>
<td>• Familiarization with the app layout</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Areas for improvement of tablet-based resuscitation records</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Need for speedy documentation and automated documentation</td>
</tr>
</tbody>
</table>

Data saturation was achieved after conducting individual interviews with 12 nurses. The mean age of the participants was 26.9 (SD 2.68) years, and 9 participants were female. Participants’ mean length of work experience was 4.46 (SD 2.20) years, with a range of 2.5 years to 8.5 years. Their mean length of work experience in AED was 3.3 (SD 1.87) years, with a range of 1.5 years to 8.5 years.

### Advantages of Tablet-Based Documentation of Resuscitation Records

#### Structural Guidance for Documentation

The electronic app included an extensive database that encompassed the essential aspects of resuscitation documentation. The participants appreciated the app’s preset data fields that prompted users to input essential data during documentation.

*(During documentation of blood glucose,) the interface displayed all the data field such as time, result, performer. You definitely cannot forget to input.* [D168-169]

*(After urinary catheter insertion) I may forget to write urinary output..., but eResus would prompt you if you did not enter.* [H 170-173]

The application made sure that you have 2 colleagues to countercheck the medication and documented their name before administration. [C164-165]

The built-in logic set by emergency physicians and nurses provided clinical management support and guidance to users during documentation. Certain data fields were auto filled, saving more time for nursing care.

*I found it convenient because the application would lead you how to input data in a step-by-step fashion.* [A49]

*After inputting the systolic blood pressure, it would automatically divert you to the diastolic blood pressure.* [C70-71]

*When asystole rhythm was chosen, the data field on blood pressure, pulse etc. would be prohibited from inputting...We no longer have to write “undetectable” over and over again.* [E47-52]

Furthermore, the electronic app reduced the need for verbal order prescriptions and allowed structured electronic prescriptions, which are less prone to error during documentation and administration.

*Verbal order was prone to miscommunication, distraction, and error in administration.* [C146-147]

*In the past, I would have to remember or write down physician’s verbal order... but now the drug name, dosage, infusion speed etc. would all be on the screen.* [K 41-47]

#### Easy to Review and Edit Documentation

Medical resuscitation documentation has to be done contemporaneously during resuscitation. Electronic documentation can ensure legibility compared with handwriting, and users were able to review specific aspects of the documentation for completeness using in-app features.

*Colleagues’ handwriting could be illegible; maybe everyone was in a hurry. And colleagues could misspell words, which could affect handover to ward colleagues.* [E22-24]

*Someone may accidentally splash alcohol onto the paper chart, and the word would become illegible.* [H209-210]

The application has a filter function which allows you to choose vital signs, allowing you to review vital sign inputs and trends or procedures, allowing you to review whether you have forgotten to document something. [K123-125]

Fragmented information was conveyed to the documenter from various sources, in a random sequence. Not all users can correctly recall the exact sequence of medical resuscitation as they document. Electronic documentation allowed the users time to edit the sequence rather than having to rewrite the whole resuscitation event on a new paper resuscitation record.

*If the handwriting was too ugly and the time sequence is too out of place, such as the medication administration time did not align to the corresponding row, then I would cross out the whole paper chart and rewrite it. But now, eResus can easily amend it.* [I 104-106]
Comparable Mobility to Paper and Superior to Desktop

Patients requiring medical resuscitation would often need to be transferred to another department for investigation or intervention. A tablet app can provide the mobility needed to document in various locations.

Let’s say the patient has to be escorted to computed tomography (CT). I would take the table to the CT suite (to continue the document). When we returned to the resuscitation bay, I could use the Bluetooth keyboard to continue the document. It’s better than desktop. [L 169-171]

You can bring it (to CT) like paper...don’t even need to bring pen, just use your fingers. [J 179-180]

Challenges With Tablet-Based Documentation of Resuscitation Records

System Loading Speed and Stability

Participants embraced the transition to electronic documentation. However, participants reported technical challenges due to the internet connection or app coding issues when using the electronic app that could compromise documentation completeness. The fast pace of medical resuscitations and contemporaneous nature of the documentation exacerbated the problem.

It has some technical problems...there was a time when it kept crashing and could not input data. [A58-60]

Sometimes, switching between different tab pages is rather slow. [F60]

The patient was...in asystole, we were conducting chest compression, and administering medication, but the application was still loading. [I 116-117]

Familiarization With the App Layout

Navigating through the various tab bars, interface, and data fields of the app was different from the paper resuscitation record that presented all the data fields on the same page. Users were required to tap multiple times to access the desired data fields on the tablet, which was more time-consuming. All participants received training prior to using the electronic app that could compromise documentation completeness. The fast pace of medical resuscitations and contemporaneous nature of the documentation exacerbated the problem.

I have to tab this and that before I could input data...if it was handwritten, it would be much quicker. [H 53-57]

This application has different tabs and options, which require a bit of thinking...it is like using a phone. [C33-43]

When you first encounter the application, you would need to spend time to learn the layout. But after you have become familiar with it, you would find the documentation process very smooth. [D147-149]

Areas for Improvement of Tablet-Based Resuscitation Records

All participants reported that speed of documentation was an important aspect in resuscitation documentation. During medical resuscitations, the case nurse was required to perform patient care and document contemporaneously. These resuscitation events were highly demanding and required speedy documentation:

Sometime, the documentation with eResus could take up lots of time. There were cases when we needed to document lots of medication right at the beginning. The application may not be able to document events in real time. [A63-64]

When you have many items pending documentation, you would be naturally prone to incomplete documentation. [B76-77]

The multitasking nature of the nursing practice posed competing demands between managing patient care and documentation, which required the nurses to compromise. Nurses decreased the frequency of taking vital signs. One participant said:

When handling less critical cases...I would take vital signs every 5 minutes (instead of 3) so that I can be more at ease when managing both the patient and documentation. [D113-117]

Users appreciated the auto retrieval of data from the Hospital Authority’s network and the auto fill of relevant fields. Relevant data previously inputted into the app were prepopulated either automatically or after the user’s approval:

It would auto-capture allergy status from CMS, an electronic health record system used in Hong Kong). [G 102-103]

It can retrieve the previous (vital sign) data. Then, I can tab it and paste it onto the data field...such as Glasgow Coma Scale score, etc. [H 44-46]

Faster documentation speed can improve documentation completeness and overall resuscitation quality. Participants believed that speed and improved care could be gained by automation:

If a multimonitor could automatically record vital signs and transfer data into the application, the user would spend less time inputting data and more time looking after the patient...or checking the (resuscitation) record for incompleteness. [D254-262]

(Automated vital sign recording) would be useful...but some factors could affect the reading’s accuracy; it should allow health care workers to verify the readings prior to documenting. [K 217-220]

Documentation speed can also be gained by flattening the user interface (UI). Participants found that, although organizing data fields into different categories and layers was logical, it made navigating through the layers inevitably slow. A more direct, intuitive UI is needed to improve the speed of data entry:

(The commonly used) items should be accessible with one tap. [G194]

https://mhealth.jmir.org/2024/1/e46744
The commonly performed investigations...that has many data fields should be more easily accessible.
[216-220]

Discussion

Principal Findings

This is the first study, particularly in Asia, to compare the completion rates of documentation between paper and tablet-based resuscitation records in the emergency room. Our results indicate that electronic documentation is promising, with a higher completion rate than with the paper format.

Our study results support the hypothesis that tablet-based documentation of resuscitation records results in a higher documentation completion rate than paper formats in all 5 domains. Previous studies in AED settings have been conducted to compare the completion rates of key data elements for trauma resuscitation records between paper and electronic formats for adult [23] and pediatric [10] trauma cases. Both studies found areas for improvement and degradation in the key data elements. However, it is difficult to directly compare the studies since the outcome measures were different and none of the studies in AED settings used tablet-based devices. Nevertheless, our study results were consistent with those of previous studies that supported that, with electronic resuscitation records, documentation completion rates were higher, particularly for basic information such as case start time and disposition but not for serial vital signs [10,23]. Interestingly, although no difference was found in the completeness of documenting vital signs and interventions in these previous studies, our study showed improvement in the vital signs, procedures, investigations, and medication domains. This may be due to the differences in the inherent design of the EMRs and the use of a tablet-based device instead of desktops, as explained by interview participants. In addition, contrary to the concern of bias for improved documentation completeness caused by training [10], our study showed no statistically significant differences between the Pre-App Paper and Post-App Paper record sets. This indicates that missing data with the paper format could be consistent since the documentation format has not changed (such as using the same paper form).

Our qualitative results further explain the reasons why the tablet-based device could improve documentation completeness. The structural design of the tablet-based resuscitation record provided guidance that contributed to the completeness. This guidance provides support to the documenter via various clinical support features such as structured prescriptions, preset data fields, and preset documentation logic. This structural guidance was developed by consulting local emergency physicians and nurses working in the AED. The guidance mimicked the normal workflow and thus supported the documentation process. Similar results were found in a previous study in which nurses had higher confidence using the EMR when they perceived that their suggestions were used to customize the system [24]. This also implies that the tablet-based device will be considered useful if it is country, institution, and department-specific.

However, similar to other studies, our qualitative findings supported that nurses have to multitask during work, which has been shown to compromise documentation completeness [8]. Furthermore, documentation in an EMR was perceived to be more time-consuming and complex [25]. Our study participants also experienced similar concerns with slow app loading speeds and needing to navigate through various tabbed pages, which increases the complexity of documentation.

Our participants further suggested that future development of the app should include automation features that would spare the documenter from manually inputting individual data into the app. Automating data input can reduce the documenters’ need to tab multiple times before finding the desired data field. This would be particularly useful for vital signs, which were the most frequently documented in this study. The UI should be flattened to facilitate input of other common data fields, which supports the concept that data fields that are more frequently recorded should be located in readily accessible spots [26]. With the automation of vital signs and an improved UI, the documenter would be able to spend more time on patient care and document in real time.

Implications for Emergency Room Nurses

Most emergency room nurses believed that the transition from paper to electronic charting can improve the quality of resuscitation documentation and patient safety. This study clearly demonstrates the potential of electronic charting to achieve that. During the transition period, the nurse documenters faced general problems with resuscitation documentation such as multitasking and unique challenges such as software updates and a subsequent need to become familiar with the app’s layout. Therefore, systematic, periodic needs assessments of nurse documenters using tablet-based devices, followed by corresponding training, should be conducted. Emergency room nurses should also be actively involved in the development and a subsequent need to become familiar with the app’s UI. Automation functions should be considered during the development of future apps to improve documentation and redistribute more time for patient care.

Limitations

This study has some limitations. It was not able to demonstrate the effect on documentation accuracy, rates of medication errors, the quality of patient care, or the process of clinical decision-making. Furthermore, since the sample was obtained from 1 AED only, the study findings may not be generalizable to other AEDs or other acute ward settings where the staffing and workflow may be different. In addition, this study excluded trauma cases and cardiopulmonary resuscitation cases; thus, its findings cannot be generalized to all resuscitation room situations in the AED.

Conclusions

This study demonstrates that a statistically higher completion rate in 5 domains essential to resuscitation documentation was achieved with a tablet-based device than with the traditional paper resuscitation documentation. Refinement of the device should be ongoing and include consultation with the users. Further studies can expand the scope to involve all medical resuscitation cases across AEDs.
Acknowledgments
We acknowledge all the emergency room nurses who participated in the interviews in this study. This study did not receive any funding.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Documentation completeness checklist.
[PDF File (Adobe PDF File), 343 KB - mhealth_v12i1e46744_app1.pdf]

References


Abbreviations

- **AED**: accident and emergency department
- **CMS**: Clinical Management System
- **COREQ**: Consolidated Criteria for Reporting Qualitative Studies
- **CT**: computed tomography
- **CVI**: content validity index
- **EMR**: electronic medical record
- **S-CVI**: scale content validity index
- **UA**: universal agreement
- **UI**: user interface

©Kin Cheung, Chak Sum Yip. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

Evidence of How Physicians and Their Patients Adopt mHealth Apps in Germany: Exploratory Qualitative Study

Tanja Schroeder¹,², BA, MA; Maximilian Haug², BEng, MSc; Andrew Georgiou¹, BA, DipArts, MSc, PhD; Karla Seaman¹, BPharm, MSCE, PhD; Heiko Gewald², BSc, MBA, EMBSc, PhD

¹Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia
²Institute for Digital Innovation, Faculty of Information Management, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany

Corresponding Author:
Tanja Schroeder, BA, MA
Centre for Health Systems and Safety Research
Australian Institute of Health Innovation
Macquarie University
Level 6, 75 Talavera Road
Sydney, 2109
Australia
Phone: 61 2 9850 ext 6281
Email: tanja.schroeder@mq.edu.au

Abstract

Background: The enactment of the “Act to Improve Healthcare Provision through Digitalisation and Innovation” (Digital Healthcare Act; Digitale-Versorgung-Gesetz [DVG]) in Germany has introduced a paradigm shift in medical practice, allowing physicians to prescribe mobile health (mHealth) apps alongside traditional medications. This transformation imposes a dual responsibility on physicians to acquaint themselves with qualifying apps and align them with patient diagnoses, while requiring patients to adhere to the prescribed app use, similar to pharmaceutical adherence. This transition, particularly challenging for older generations who are less skilled with technology, underscores a significant evolution in Germany’s medical landscape.

Objective: This study aims to investigate physicians’ responses to this novel treatment option, their strategies for adapting to this form of prescription, and the willingness of patients to adhere to prescribed mHealth apps.

Methods: Using an exploratory qualitative study design, we conducted semistructured interviews with 28 physicians and 30 potential patients aged 50 years and older from August 2020 to June 2021.

Results: The findings reveal several factors influencing the adoption of mHealth apps, prompting a nuanced understanding of adoption research. Notably, both physicians and patients demonstrated a lack of information regarding mHealth apps and their positive health impacts, contributing to a deficiency in trust. Physicians’ self-perceived digital competence and their evaluation of patients’ digital proficiency emerge as pivotal factors influencing the prescription of mHealth apps.

Conclusions: Our study provides comprehensive insights into the prescription process and the fundamental factors shaping the adoption of mHealth apps in Germany. The identified information gaps on both the physicians’ and patients’ sides contribute to a trust deficit and hindered digital competence. This research advances the understanding of adoption dynamics regarding digital health technologies and highlights crucial considerations for the successful integration of digital health apps into medical practice.

JMIR Mhealth Uhealth 2024 | vol. 12 | e48345 | p.218
https://mhealth.jmir.org/2024/1/e48345

KEYWORDS
mobile health apps; DiGA; adoption; prescription; mHealth; aging and individual differences

Introduction

Mobile devices have enormous potential to enhance the way patients receive medical care and health education [1]. Mobile health (mHealth) is a dynamic and expanding area of health care with short innovation cycles [1,2]. mHealth is closely related to telemedicine and eHealth. The fundamental difference lies in how the related services are delivered to the patient. According to the World Health Organization (WHO), mHealth expands the spectrum of functionalities to include a mobile component. Thus, mHealth is delivered through any mobile
device, from simple mobile phones, smartphones, and tablets to wearable devices used in health care settings [2].

To accommodate the growing number of mHealth apps (ie, software products for smartphones designed to support good health), Germany passed the “Act to Improve Healthcare Provision through Digitalisation and Innovation” (Digital Healthcare Act; Digitale-Versorgung-Gesetz [DVG]) in December 2019 and the Digital Health Applications Ordinance (Digitale-Gesundheitsanwendungen-Verordnung [DiGAV]) in October 2020 [3,4]. This ordinance enables physicians to prescribe mHealth apps (in German referred to as Digitale Gesundheitsanwendungen [DiGAs; digital health applications]) to their patients in the same way as any other medicine. A DiGA supports the recognition, monitoring, treatment, and alleviation of diseases, injuries, or disabilities [5]. To qualify as a DiGA, an mHealth app has to successfully pass the evaluation of the Federal Institute for Drugs and Medical Devices (Bundesinstitut für Arzneimittel und Medizinprodukte; BfArM). Only then will an app be included in the list of reimbursable digital health apps, the so-called DiGA directory. In September 2023, this directory comprised 48 admitted DiGAs [6]. DiGAs cover a wide array of the medical spectra. Popular examples are “Endo-App” (to treat endometriosis), “Kaia COPD” (for chronic obstructive pulmonary diseases), or “sinCephalea” (for the treatment of migraine).

The introduction of the “Act to Improve Healthcare Provision through Digitalisation and Innovation” represents a significant and innovative change to the German health care system [7]. The new legal framework elevates DiGAs to the ranks of medical devices [5]. Therefore, if physicians want to prescribe a DiGA, they are required to thoroughly inform themselves about which apps can help their patients, how they are to be used, and how a positive influence is expected to manifest. Patients are required and need to be able to adhere to the prescription; for example, they have to use the DiGA as stated by the physician. Thus, the physician needs to assess (implicitly or explicitly) whether the patient is likely to use the DiGA as prescribed. This does not only include adherence to, for example, training intervals (for orthopedic DiGAs) but also the general technology savviness of the patient, for example, whether the patient is able to download the app, install it on the smartphone, maintain updates, and so on. This question is specifically challenging when it comes to older users, who are often regarded as being not technology savvy.

Consequently, this innovation raises several questions for technology adoption research. Typically, adoption research concerns the individual user’s decision whether to use technology, either mandatory [8,9] or voluntary [10]. Now, a concerned third party (the physician) decides on behalf of the user whether the patient is expected to be willing and able to use an app on their smartphone voluntarily. Therefore, the physician’s assessment now includes not only the medical relevance of the DiGA but also whether the user might be able and willing to use it as prescribed. Although the latter assessment seems easy for younger people, the case is much more difficult for older patients. Potential doctor misperception raises concerns about a possible digital divide and ageism by doctors. A recent report by the WHO and the United Nations (UN) raises awareness of this issue and urges action to combat ageism, as it negatively impacts well-being and can lead to premature death and higher health care costs [11].

In the health care context, the resulting research questions are specifically relevant, as the influence of age on the adoption of mHealth apps has not yet received sufficient attention in the scientific discourse [12]. Therefore, this research addresses the complementary research questions:

1. What factors enable or hinder physicians to prescribe DiGAs?
2. What factors enable or hinder older users’ adoption of DiGAs?

Methods

Study Design

We developed an exploratory qualitative study design to answer the research questions and gain insights through semistructured interviews with representatives of the 2 relevant stakeholder groups: physicians and patients. The study adhered to the standards for reporting on qualitative research [13].

Study Setting

The data for our research were gathered in Germany.

Participants

The first qualitative study was conducted from August 2020 to June 2021. We interviewed 28 physicians (demographics are provided in Table 1) to assess how DiGAs can improve the health of their patients. Physicians were chosen as research objects because of their unique role as prescribers of DiGAs. The interviewees had different backgrounds in terms of IT affinity and previous experience with mHealth apps in general.
In the second study, we interviewed patients (ie, the potential users of a DiGA). To reflect the specific issues of aging patients, we chose interview partners over 50 years of age. Evidence shows that from this age onward, chronic diseases increase significantly [14]. Therefore, this age group is likely to represent a large part of the target group for the prescription of DiGAs. Furthermore, studies suggest that there are still age disparities in attitudes toward technology and that the aging population is often less comfortable using technology [15].

We conducted 30 interviews to determine the factors that influence potential patients to adopt DiGAs or not. Demographics of the interviewees are given in Table 2.
Table 2. Demographics of the patients.

<table>
<thead>
<tr>
<th>Code</th>
<th>Sex</th>
<th>Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Male</td>
<td>68</td>
</tr>
<tr>
<td>P2</td>
<td>Female</td>
<td>60</td>
</tr>
<tr>
<td>P3</td>
<td>Female</td>
<td>57</td>
</tr>
<tr>
<td>P4</td>
<td>Female</td>
<td>76</td>
</tr>
<tr>
<td>P5</td>
<td>Female</td>
<td>56</td>
</tr>
<tr>
<td>P6</td>
<td>Female</td>
<td>65</td>
</tr>
<tr>
<td>P7</td>
<td>Male</td>
<td>69</td>
</tr>
<tr>
<td>P8</td>
<td>Female</td>
<td>64</td>
</tr>
<tr>
<td>P9</td>
<td>Male</td>
<td>68</td>
</tr>
<tr>
<td>P10</td>
<td>Female</td>
<td>66</td>
</tr>
<tr>
<td>P11</td>
<td>Male</td>
<td>65</td>
</tr>
<tr>
<td>P12</td>
<td>Male</td>
<td>67</td>
</tr>
<tr>
<td>P13</td>
<td>Female</td>
<td>57</td>
</tr>
<tr>
<td>P14</td>
<td>Female</td>
<td>72</td>
</tr>
<tr>
<td>P15</td>
<td>Female</td>
<td>67</td>
</tr>
<tr>
<td>P16</td>
<td>Male</td>
<td>53</td>
</tr>
<tr>
<td>P17</td>
<td>Female</td>
<td>61</td>
</tr>
<tr>
<td>P18</td>
<td>Female</td>
<td>69</td>
</tr>
<tr>
<td>P19</td>
<td>Female</td>
<td>61</td>
</tr>
<tr>
<td>P20</td>
<td>Male</td>
<td>63</td>
</tr>
<tr>
<td>P21</td>
<td>Male</td>
<td>65</td>
</tr>
<tr>
<td>P22</td>
<td>Female</td>
<td>68</td>
</tr>
<tr>
<td>P23</td>
<td>Female</td>
<td>67</td>
</tr>
<tr>
<td>P24</td>
<td>Female</td>
<td>61</td>
</tr>
<tr>
<td>P25</td>
<td>Female</td>
<td>67</td>
</tr>
<tr>
<td>P26</td>
<td>Female</td>
<td>59</td>
</tr>
<tr>
<td>P27</td>
<td>Male</td>
<td>64</td>
</tr>
<tr>
<td>P28</td>
<td>Female</td>
<td>69</td>
</tr>
<tr>
<td>P29</td>
<td>Female</td>
<td>64</td>
</tr>
<tr>
<td>P30</td>
<td>Female</td>
<td>68</td>
</tr>
</tbody>
</table>

Recruitment

We recruited physicians by telephone from a community-based physician network in Germany and via social media. For the interviews with patients, we promoted our study to doctors and approached medical centers and clinics. To identify suitable interview partners (ie, potential patients), we presented the research project to local sports, communication, and civic groups; promoted the study on various social media platforms; and spoke to the local press. In this way, we motivated suitable candidates to contact us. The participation of all study participants, both physicians and patients, was voluntary.

Data Collection

All interviews were semistructured and led by a list of questions and general topics that the interviewers were supposed to address. The semistructured interview guidelines are provided in Multimedia Appendix 1. As a structure and topic basis for creating the questions, the constructs from well-known technology adoption research models (unified theory of acceptance and use of technology [UTAUT] [9]) and models for the analysis of health behavior (health belief model [HBM] [16]) were considered. The questions were primarily open and allowed the interviewees to explore their experiences and views. Supported by a systematic and comprehensive interview process, the interviewers had high degrees of freedom to conduct the interview in order to gain deeper insights. The questions were adjusted correspondingly for the following interviews to gain...
more profound knowledge for each interview. Interviews lasted around 20-45 minutes and were conducted face-to-face or over the phone by 1 researcher (TS). The interviews were conducted in German and were recorded, transcribed, and translated into English for further analysis [17].

**Data Analysis**

For the coding process, we used the NVivo10 software (QSR International). The research was conducted using an interpretative phenomenological analysis (IPA) [18]. IPA aims to examine the world perspective of the participants and, if possible, take an “insider perspective” [18] of the phenomenon under investigation. At the same time, IPA recognizes that research is a dynamic process. In parallel with the data collection, we scanned and coded the data from the first round of analysis. One researcher (TS) conducted a thematic analysis to identify patterns and themes. After analyzing the first 3 interview transcripts, 2 researchers (MH and TS) developed a coding framework using an inductive approach that allowed for identification of predominant themes. All emerging themes were cross-checked and discussed within the whole research team and developed iteratively to ensure definition and reliability. In the process, commonalities and differences between the respective perspectives were identified. This led to 9 different key themes concerning benefits and barriers, influencing factors on the intention to adopt a DiGA, and the outcome expectation from different perspectives as the main areas. Figure 1 lists the corresponding coding scheme according to Gioia et al [19]. The aim was to understand the specific properties of these areas and the influence of these factors in the context of the introduction of DiGAs.

---

**Figure 1.** Coding scheme using the methodology by Gioia et al. DiGA: Digitale Gesundheitsanwendungen (digital health application).
Ethical Considerations
This study did not require ethical approval according to the guideline of the applicable Ethics Committee of the Bavarian Universities [20], as no risks or harm was brought forward to the participants. All participants received an information and consent form explaining the requirements for participation. This included the opportunity to have the form explained to them if needed. All interviewees gave written or verbal consent before the interview started. Data collection, storage, and analysis were conducted in adherence to European Union General Data Protection Regulation (EU-GDPR). None of the participants were compensated.

Results
Overview
Our data demonstrate that only 4 of the interviewed physicians already prescribed DiGAs and none of the patients had used a DiGA before (Table S1 in Multimedia Appendix 1). Nevertheless, most physicians have experienced mHealth apps themselves. These were mostly used for medical information in their professional routine or privately for personal fitness and nutrition goals.

On the patient side, 20 participants had already had experience with mHealth apps or were actively using them (Table S2 in Multimedia Appendix 1). In this context, only fitness and nutrition apps were mentioned. These apps have a preventive character to help people to consciously lead a healthier and more active life. Most of the apps mentioned were those used in conjunction with smartwatches to measure activity, such as pedometers.

The DiGA Adoption Process of Physicians and Patients
The analysis indicates that for DiGAs in particular, additional steps are added to the traditional prescribing process. Because of the special role of the doctor as a gatekeeper, the doctor is the first to decide whether the patient is suitable for a DiGA. First, the familiar steps such as the patient’s trust in the doctor and the treating doctor’s determination of the patient’s medical condition represent the keystone of the process. Following these steps, when prescribing a DiGA, the next steps are the doctor’s consent to prescribe a DiGA under certain conditions and the assessment of the patient’s ability to use a DiGA. Thereafter, the doctor prescribes a DiGA, and the patient is given the opportunity to adopt it. The process ends with the expectation that the patient will continuously use the prescribed DiGA and report the results to the treating physician. This process is visualized in Figure 2, with the doctor and the patient being influenced by different factors that determine whether the DiGA is prescribed (doctor’s perspective) and accepted (patient’s perspective).

Factors Enabling or Hindering Physicians to Prescribe DiGAs
Overview
Our results indicated that DiGAs are not yet widely known and used at this early stage by the physicians we interviewed. They indicated that the level of information is still insufficient and that very few DiGAs are prescribed compared with the prescription of drugs. DiGAs represent a completely new and innovative approach, so entry challenges are not considered unusual. Nevertheless, these challenges are initially barriers to prescribing DiGAs according to our participants. We were able to identify the following factors that influence physicians to prescribe a DiGA or not.

Lack of Information
Another reason often mentioned in connection with rejection was a lack of information. Physicians often do not know which DiGAs are available and where they could get the necessary information. As a result, physicians are often reluctant and skeptical about prescribing DiGAs (“No, I have not been educated on what DiGAs are available, how to prescribe, and how to tell if a DiGA is effective. I would not know where to look for this information” [E2, E3, E8, E12, E16, E20, E21, E25, and E28]; “I have little time to search for all the information I need to prescribe a DiGA in good conscience” [E1, E4, E8, E15, E16, E20, and E21]). They tend to be negative out of concern for malpractice and the resulting liability risk. Likewise, the physicians described the requirements for evidence of a DiGA (requirements for security, functionality, data protection, information security and quality, and positive effects on care) in the assessment procedure of a DiGA as insufficient. Along with this concern is the fact that for provisionally included DiGAs, evidence-based studies on the benefit of the

Lack of IT Resources
The lack of IT resources was mentioned by the interviewed doctors as a barrier to prescribing DiGAs within the framework of the DVG. Numerous efforts have been made toward digitalization. These include networking with various players in the German health system, such as doctors, hospitals, pharmacies, and health insurance companies. So far, only limited resources are available for the implementation of these plans. The doctors interviewed describe that, for example, the provision of services for an electronic health record (EHR) could help to digitalize various processes and information. They see the benefit of a DiGA for their work as low as long as the patient files and the exchange between other actors in the health care system are not fully digitized. The introduction of an EHR in Germany has failed so far because of technical challenges.
app are not yet available, and the apps are therefore only provisionally included in the directory. A comprehensive explanation of the BfArM authorization process of a DiGA can counteract these problems. However, participants reported receiving little information from insurers, DiGA manufacturers, or the BfArM. Physician respondents stated that proactive communication from DiGA providers was limited and possible involvement in the development process was unknown (“These apps have often been developed without the support of physicians and are now being ‘fast-tracked’ to market—that is not exactly building trust” [E25]). This statement reflects a high level of mistrust. In contrast, one participant stated that he worked with an app development company to develop an app in his field and described that this approach was extremely helpful to reduce the information deficit, mistrust, and gain more confidence. This shows that involvement is possible, but that an exchange of information is necessary.

**Lack of Certification**

In medical circles, a lack of certification has been widely mentioned. Half of all interviewed physicians agree that DiGAs, similar to other medical products (eg, drugs and medical aids such as wheelchairs or bandages), should be provided with a known official certification to ensure more trust in the product (“I think a well-known and recognized certification in medicine could help to build confidence in DiGAs” [E2, E13, E15, E17, E26, and E28]). Certification would make it easy to recognize that the DiGAs are safe as a medical device and are also medically or technically suitable within the scope of the intended purpose stated by the manufacturer. A certification equivalent to that for medicines could be conceivable in this context.

**Lack of Involvement**

Another aspect that our interviewees criticized was the lack of involvement after the prescription of a DiGA. They described that they are little involved in the procedure after prescribing a DiGA and in many cases currently only take on the role of a “prescriber” (“I would like to be involved in the whole process from the manufacturer to the evaluation of the data with the patient” [E5]). This shows that the participating physicians apparently make a great distinction between a traditional medicine in the form of medication administration and the use of digital components. After prescribing a traditional medicine, the physician only gets feedback when the patient comes back after some time and tells them how the medicine works. But with digital options such as DiGAs, the requirement is now higher: participants demanded an adjustment of the involvement in the postprescription process. A preview of future digitization plans shows that the involvement of physicians in the digital feedback process will be considered.

**Lack of Financial Incentives**

Participants further stated that counseling for a DiGA is much more time-consuming than for medicines, but the monetary incentive is not there. As a result, we identified a lack of financial incentive. Financial pressure weighs on the physicians in this regard, which is not compensated for by health insurance companies. As a result, the incentive for prescribing ($2 [US $2.16] per prescription) and treatment support (eg, successful monitoring) is currently considered too low (“2 euros for prescribing or 7 euros for treatment support are in no way an incentive to prescribe a DiGA” [E2]).

Finally, the study also found differences in physicians’ skills, knowledge, and attitudes toward digital technology. We define this factor as digital literacy, which has 2 specific characteristics. Some of the surveyed physicians stated that they have the impression that a physician is either digitally interested and very open-minded or completely ignorant of new innovations, so that even educational conversations might fail (“Either you find it good as a physician and have dealt with it once or recommended it to your patients. Or you ignore it at first” [E6]). Along with this result, the physicians interviewed expressed the fear that they might lack knowledge, for example, when it comes to patients’ technical questions or that they would have to become a kind of “technical support” in the event of problems in this area or in the evaluation of DiGA analyses. Others, however, considered the introduction of DiGAs to be an advantage for their work and less of a hindrance or a problem.

**Factors Enabling or Hindering Older Users’ Adoption of DiGAs**

**Overview**

As indicated above, a process of influencing factors leads to the adoption of DiGAs by patients. Thus, if the physician determines a medical condition, is willing to prescribe a DiGA, and considers the patient’s ability to use a DiGA to be positive, the physician will eventually prescribe a DiGA so that the patient will have the opportunity to adopt it.

**Lack of Information**

Similar to the physicians, we also identified a lack of information on the patient side. The DiGA concept is rather rarely known by the patients, and experiences were only described in 3 of the 30 interviewed participants. Nevertheless, from the patient perspective, DiGAs are recognized as an innovative and profitable treatment option that can be carried out independently of the time and place of the doctor’s visit (“I want to have the flexibility to do my therapy when it suits me” [P1, P6, P9, P12, P18, P26, P29, and P30]). Patients described DiGAs as a helpful “bridge” and a refresher or repetition of therapy content, especially for patients waiting for an appointment with a specialist (eg, psychotherapist). Some patients pointed to a long history of illness and low chances of success of conventional therapies and considered DiGAs as another treatment option. The interviews thus confirmed that the population’s willingness for DiGAs is high. However, many patients lack further information (“My physician or my health insurance company haven’t informed me about it yet—how should I know?” [P2, P3, P12, P16, P19, P22, and P29]).

The factors influencing the adoption of DiGAs are discussed in more detail below. In this context, we are oriented toward the most well-known models in adoption research: the UTAUT [9], a model from information systems that measures the acceptance of a technology by users to gain access to individual user behavior, and the HBM [16], a theoretical model from health psychology that analyzes and predicts health-related behavior.
Self-Efficacy
Self-efficacy refers to a sense of control over one’s environment and behavior. Participants considered self-efficacy important but also taken for granted. A distinction was made between technology self-efficacy about the DiGA and health self-efficacy with a health aspect. We define technology self-efficacy as the patient’s belief in their competence to use the DiGA. Health self-efficacy here describes the patient being confident in managing their health. Participants expressed confidence in their technology self-efficacy. For all participants, everyday use of smartphones and various apps was normal and regular. Some indicated that they still prefer paper calendars or dislike apps for health prevention but saw digital documentation and treatment support as an advantage in case of possible illness (“I’m more used to pen and paper, but if you get used to it, it certainly has its advantages” [P1, P4, P12, P14, P21, and P29]). In terms of health self-efficacy, participants were mostly self-confident. No participant stated that they did not want to deal with illnesses or would rather not know how healthy or ill their body was. Most participants seemed to have a high level of health literacy and wanted to actively deal with their health (“Yes, I would rather be the person who would then say, I would like to know this to know my enemy and then fight against it” [P19]). Participants indicated that a personal sense of control facilitates health behavior change (“Then an app like that would be great because you can calm down a bit more and have more security for yourself” [P10]). Participants confirmed that they focus on the opportunities rather than the obstacles (“Yes, I usually don’t go right away, but after a short scare I always face all the problems, so I’m more for problem solving rather than suppression” [P3]).

Perceived Threat
Perceived threat was mentioned as a crucial central factor for the use of a DiGA in the context of an impending or existing chronic disease. According to the HBM, perceived threat stems from beliefs about perceived susceptibility to disease and the perceived severity of disease consequences [16]. Susceptibility here refers to a person’s risk of contracting a disease. Severity refers not only to the medical consequences but also to the potential impact of an illness on a person’s daily life, family life, and social relationships. Participants indicated that a threat has a strong influence on health behavior. They explained that if they were seriously ill, they would use any means to support treatment, both digital and analogue (“So, I would say if I got sick now, I would be very interested in a DiGA already” [P29]; “Yes, when I get sick, I try everything possible to get better, whether digital or not” [P3]). Nevertheless, participants indicated that they generally felt very body and health conscious (“I eat healthy and exercise as much as I can, that’s part of my everyday life” [P9, P11, P15, P16, P21, P23, P26, and P29]). Only 1 respondent indicated that they felt an increased fear of possible illness. All participants indicated that they were primarily concerned with their health and well-being and that preventive measures were a natural part of their lives (“I prefer to focus positively on my health instead of worrying daily about illness and negative health” [P23, P29, and P31]).

Trust in the Physician, Attitudes Toward Privacy, and Safety Concerns
Partly different from classical adoption research, trust in the physician, attitudes toward privacy, and safety concerns were mentioned as key characteristics. Due to the medical field, the existing models require expansion. Trust does not describe a direct influencing factor but rather a precondition. For the participants, a deep trust relationship with their physician was crucial to consult a physician in case of a health problem and to receive good medical treatment. If this precondition is not given, an exchange about digital treatment options does not take place (“If I feel that I cannot trust him, then I would change” [P1, P7, P10, P13, and P22]; “The most important thing is the trust relationship. If I don’t trust my physician, I change physicians” [P27]). Subordinated are the characteristics of privacy and security. All participants considered it important to be able to set privacy settings themselves. The vast majority said that they did not particularly care about the content of privacy settings, but wanted to decide for themselves who could access which health data (“I would like to know who knows what about me” [P6, P7, P8, P15, and P26]; “There should be settings options. I don’t want to transmit everyday occurrences” [P9, P10, P11, P13, and P27]). Based on the accreditation of a DiGA as a medical device and the assessment by the BfArM, they have no concerns about the safety of DiGAs and feel confident in using DiGAs (“I trust that our federal system is highly secure” [P11]). On the basis of these correlations, we categorized the 3 concepts together. We rank the trust factor as the most important, as the privacy and security factors can be mitigated by a high level of trust.

Social Influence
The concept of social influence reflects the effect of environmental factors, for example, the opinion of friends and family, and is a significant factor in traditional adoption research [21]. However, in this study, we could not find any relevant results. Here, the medical context seems to have an important role. Most participants indicated that they did not discuss their health behaviors, personal diagnosis, or treatment plans with their social circle (“I don’t want to share all health data, including the fact that I use a health app, with other people” [P9]). We suspect that social influence may be a factor that is difficult to capture due to the sensitive nature of the data.

Outcome Expectations
Outcome expectations are defined as the expected consequences of a certain health behavior, which can be negative or positive [22], that is, what does the patient expect from using a DiGA. Participants stated that they would be able to reduce the frequency of visits to the physician, thus saving travel and time; that an existing illness would be better monitored by the physician; that they would feel safe and well cared for as a result; and that they would receive health-promoting treatment in the form of the DiGA. However, these expectations were tied to the continued use of features of the DiGA. Participants agreed that a DiGA must be simple to use, regardless of age. Many participants expect the app to provide more detailed information about the disease, symptoms, medication, and contraindications. However, this information needs to be understandable,
meaningful, and informative for every patient, regardless of age, education level, and professional background (“It should be easy for me, and I should be able to understand and comprehend it. The ease of use.” [P1, P5, P16, P19, and P21]).

**Facilitating Conditions**

Eventually, these expectations also cross over to the facilitating conditions, which includes the impact of the patient’s knowledge, skills, and resources. Here, the participants considered technical support to be particularly important. This was not because they felt unsure about using the technology but rather because it was a new type of health intervention. Some participants considered the physician to be an appropriate point of contact when difficulties arose in a few cases. Other participants stated that if they had difficulties or questions, they would ask their family and friends for help, as they do with other technical matters (“I think it would be good to have a number that I can call and that can help me” [P2]; “Well, I trust my son, he knows me anyway. And he would also know the diseases I have” [P18]). In summary, this study distinguishes between technical infrastructure and health support, with the health aspect (eg, knowledge and understanding of the diseases and their treatment) being more important here.

**Post Hoc Analysis: Physician Assessment of the Patient’s Ability to Use a DiGA**

During our interviews with the physicians, we discovered a phenomenon that we did not initially anticipate and that has not yet been described in the relevant technology adoption literature: the pre–user adoption decision of another instance, whether the user will be able and willing to voluntarily use a technology. Typically, in technology adoption research analyses, this means whether a user is willing to use a technology in an either voluntary or mandatory environment. In any case, the decision remains with the user.

Now we see a new phenomenon: the assessment of one instance (physician), whether a subsequent instance (patient) would be able to use the DiGA. Only if the assessment is positive would the physician offer the DiGA to the patient, which will then trigger the traditional adoption questions and corresponding behavior of the user, as described in the well-researched technology adoption models such as the technology acceptance model, UTAUT, and HBM [9, 16, 23].

There are no guidelines under which circumstances a patient should be assumed to be able to use a DiGA. Thus, each physician needs to do this assessment individually. If they conclude that the patient will likely not be able to install, maintain, and use the app as foreseen by its developers, there is no point in prescribing the DiGA. As there are no objective guidelines, the assessment is either done explicitly, by asking the patient, or implicitly, by assuming what the patient is capable of.

Interestingly, the physicians interviewed were very consistent regarding the assessment of digital literacy of their patients. They indicated that prioritization of certain patient groups is facilitated by anticipating the digital literacy of their patients. Unfortunately, this often leads to a negative bias toward older users; the physicians described the typical DiGA user as a young and tech-savvy patient (“I would not consider my older patients for the use of DiGAs” [E1, E2, E4, E7, E8, E11, E12, E15, E16, E17, E20, E21, E22, and E28]; “There are certainly exceptions, but most of my older patients are totally overwhelmed with a tablet or a smartphone, because the interest would not even be there” [E11]). Consistently, physicians expressed that they would not even consider an older patient as a DiGA user.

As mentioned before, these findings arose from the data and were not anticipated before. Thus, both issues, the second-order technology adoption process and the (possible) systematic disadvantage of older users, need deeper investigation in further research.

**Discussion**

**Overview**

In this research, we identified the salient factors that were either beneficial or hindering the adoption of DiGAs from the physician and the patient perspective. Furthermore, the results of our study suggest that the adoption process for a DiGA does not only depend on patient behavior but also on the physician’s behavior.

Most informants have a positive attitude toward the digitalization in general. Nevertheless, physicians’ demands on DiGAs are high, and their perceptions can be affected by a lack of facilitating conditions, trust, and digital competence. Certain influencing factors for the adoption of DiGAs by patients are consistent with the literature on established adoption research [9, 24-28].

**Principal Implications**

Our study contributes to the field by investigating factors influencing the adoption of DiGAs to inform future research and guide strategies and efforts for this user group. DiGAs represent a wide range of assistive apps that aim to support disease behaviors, manage various health conditions, and maintain the well-being of those with chronic diseases. There are very few empirical studies addressing the factors influencing users’ adoption of DiGAs [7, 29, 30]; hence, there is limited knowledge and guidance from the existing literature.

First, it is important to demonstrate that existing technology acceptance models reach their limits when used in the context of DiGAs. In contrast to Davis [23] and Venkatesh et al [9], our interviews with physicians and potential patients led to the assumption that, in addition to usefulness and ease of use, there are more constructs that play a significant role. We found that technology and health aspects such as technology- and health-related self-efficacy, trust, and a trustful doctor-patient relationship play a major role in the intention to use DiGAs. So far, these aspects have rarely been brought together. A study by Uncovska et al [30] confirms this finding. However, we note that there are few studies on the adoption process of DiGA. Other studies regarding mHealth app adoption have highlighted that health consciousness of individuals is a factor that directly influences both the intention to use mHealth apps and the actual use behavior [31, 32]. Public trust in the health care system [33] and a strong doctor-patient relationship can empower patients to contribute to treatment decision-making [34].
Second, we found that our interviewees on the patient side distinguished between technology-related and health-related self-efficacy. The consideration of a health component is not integrated into traditional technology adoption research. To date, it has not seemed necessary to consider the health domain in adopting general technologies. However, this is an important difference in the adoption of DiGAs. With DiGAs, the focus is on the health aspect for both the physician and the patient. Patients who may have low self-efficacy with technology do not simultaneously have to have low self-efficacy with their health. Older people, in particular, may have very sophisticated health self-efficacy while lacking technology self-efficacy [35-38]. Distinguishing these forms of self-efficacy provides a more detailed explanation of adoption behavior, increasing our understanding in the context of DiGAs. Personal beliefs, such as outcome expectations and self-efficacy expectations, are among the most critical variables in terms of intention formation and bridging the gap between intention and behavior, according to existing literature [39-41]. Nevertheless, a division into different areas of self-efficacy has not yet been made in information systems research but also in research on health adoption. As technology and health self-efficacy positively impact the adoption of DiGAs, we believe that it is important to consider both factors.

Third, related to the previous aspect, is the construct of perceived threat. Previous research shows that people are concerned about adopting technology in different areas, such as privacy, effort, or performance [7,42,43]. However, when using DiGAs, health is firmly in focus from a medical perspective. Therefore, the perceived threat of diseases and the need to use a DiGA strongly influenced the adoption of a DiGA. In our interviews, this construct was strongly emphasized, and we suspect moderation effects on other constructs, such as technology and health self-efficacy. A high perceived threat can increase the influence of the perceived health self-efficacy on adopting a DiGA because a threat can be better assessed by someone with high health self-efficacy and is, therefore, more likely to act. As a result, it is also possible that DiGAs will have a higher use rate, especially for hazardous diseases. A recent study by Pourhaji et al [44] investigated the perceived threat and stress response to the COVID-19 pandemic and found that the Iranian population’s health behavior was influenced by the perceived severity and susceptibility of the infection, which meant that preventive interventions were more likely to be accepted. Further studies related to COVID-19 found that risk severity also tends to increase with age, but the perception of susceptibility to contracting COVID-19 decreases [45-47]. Thus, risk perception does not seem to increase with age, but vulnerability and severity show opposite patterns [48]. The HBM postulates that individual beliefs about risk can be influenced by various factors such as sociodemographic and sociopsychological variables as well as knowledge, experience, and awareness [16]. However, patient awareness can also become an important issue, as these patients may not perceive a threat and, therefore, not adopt a DiGA.

Fourth, in addition to the patient, this study involves another important stakeholder, the physician. This stakeholder is not considered in the technology adoption models as they do not provide a specific gatekeeper for the technology or basically consider different stakeholders. But in the case of DiGA adoption by the patient, the first step requires the physician’s adoption of a DiGA. This observation has also been noted in previous studies [7,49-52]. Subsequently, the physician’s positive assessment of the patient’s competence to perform a certain behavior is one of the essential conditions for a patient to consistently perform a health intervention. This result relates to previous research without reference to DiGAs as well as with reference to DiGAs [7,53]. Similarly, this view can be developed in a negative direction when doctors decide that the patient is not capable of adopting and using a DiGA, which could be justified by digital ageism. Ageism is a societal bias conceptualized as (1) prejudicial attitudes toward older adults, (2) discriminatory practices toward older adults, or (3) institutionalized policies and social practices that promote these attitudes [54]. Ball et al [55] show that both the development and use of technology have excluded older adults, resulting in a “physical-digital divide,” which exists when a group feels excluded because they are unable to engage with the technologies used around them. Some studies suggest that ageism is widespread in the health care system [56-58]. For example, Walter et al [59] showed that physicians promote less preventive care for older patients. Chu et al [60] emphasized that the exclusion of older people from technology development leads to a broader cycle of inequality and ageist social attitudes, widening the digital divide. In contrast, we noticed that the trust factor impacts a patient’s health behavior, which is in line with Wildenbos et al [61]. Beyond this, we also found that the physician’s trust in the DiGA is equally important for their prescription of a DiGA. A physician needs a strong relationship of trust with the patient to convince the patient of the treatment methods. By motivating their patients to adopt a DiGA and use it to support their therapy, physicians focus on their social influence on the patient [7,62]. We provide justification for the incorporation of the physician as an important influence on adoption behavior in this context. After all, DiGAs live and die with physicians’ willingness to prescribe to their patients and influence them to understand the technology’s necessity. We argue for an adoption model that does not only incorporate a human-technology interaction but also a human-human-technology interaction.

**Implication for Practice**

The results of our study demonstrated that there is insufficient information available and published for both physicians and patients, as well as a lack of comprehensive technical support. Some statements (eg, the involvement of medical professionals in the development process of a DiGA, DiGA list unknown, and lack of evidence-based sources) confirm the knowledge deficit. It is important for health policy makers and public authorities such as the Ministry of Health, the Medical Association, and insurers to address these issues. Extensive information and source references are needed to take into account the needs of physicians and to enable DiGAs to get started more effectively. Including the consideration of conflicting goals in technology development from the beginning seems necessary. Codevelopment can improve app use and effectiveness in the long term by using a user-centered design.
to develop DiGAs that are effective in chronic disease self-management [63,64]. Likewise, we demonstrated that physicians distinguish clearly between digital and traditional treatment options and have significantly higher expectations of digital resources. Despite a previously excluded position of the physician in the development and introduction process of medicines, the physician now anticipates more involvement in digital developments. In order to counteract the feeling of exclusion and disconnection from the supposedly nonmedically focused DiGA development, appropriate education seems to be required. One way to achieve this is through promotion within public networks, local authorities, and medical associations.

Insufficient participation in the follow-up to a prescription was also a concern for medical practitioners. In this context, it is important to fully communicate to physicians the many opportunities DiGAs offer. DiGAs have higher potential than other treatment interventions such as medicines to maintain a meaningful exchange of information and stay in touch with the patient even after prescription. We found that physicians feel a loss of control when the DiGA is prescribed and then used by the patient. A high degree of self-management is demanded of the patient, leaving the physician feeling incapable of action. Digital monitoring with the help of a DiGA also results in an advantage after the prescription compared with the conventional prescription of medical devices and medicines. Furthermore, negative attitudes and lack of digital competence among physicians are major barriers to physician prescription of DiGAs and, thus, patient adoption. At the same time, the lack of facilitating conditions and the high demands regarding the introduction of various digital changes (eg, EHR) exert excessive pressure on physicians. In this case, further information and education of physicians would be useful. In addition, a trial period of the DiGA can demonstrate digital connectivity to the physician.

**Limitations and Future Research**

Because of the exploratory character of the research design, our findings naturally lack generalizability and should be regarded as a first starting point on the investigation of a new phenomenon. As DiGAs are a new option for physicians, several interview partners have not really experienced them yet. We found that some participants, especially those who had not yet heard of DiGAs, found it difficult to properly understand the use and benefits of DiGAs.

Another issue corresponding to the novelty of the phenomenon is that informants’ perceptions change quickly. Therefore, our findings reflect the perceptions of interview partners in the early phase of introductions of DiGAs into the market. It is likely that some of the issues raised will not be present in a couple of years, when DiGAs are more common to the market and perceived as natural to prescribe as all other medicines today.

This poses interesting questions for further research. It would be interesting to conduct longitudinal studies to gain a better understanding about the diffusion of such innovations in the medical space from the legislative setting into the physicians’ toolkit and finally to the patients. This could generate valuable insights for future management of digital innovations in the medical area. In close conjunction to these questions, a cross-national comparison could generate advice for policy to smoothen the introduction phases of digital medical innovations in new other countries.

Finally, the study of the second-order adoption mechanisms—highlighted earlier in the document—could lead to interesting theoretical insights and valuable advice for practitioners to enhance the prescription and adoption of DiGAs and comparable digital innovations.

**Conclusions**

DiGAs provide an opportunity to support people with severe (often chronic) diseases, to live independently with greater confidence and understanding of their condition, better symptom management, and ultimately enhanced quality of life. Our study provides deep insights into the needs and circumstantial evidence that enables a better understanding of the perspectives and preferences for adopting DiGAs by physicians and potential patients. We found that there is a considerable lack of information on both physicians’ and patients’ sides, resulting in poor trust and digital competence. Furthermore, we identified several factors influencing the adoption of DiGAs, which led to a new understanding of adoption research concerning digital health technologies.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Interview guidelines and descriptive results.

[DOCX File, 42 KB - mhealth_v12i1e48345_app1.docx]

**References**


5. The fast-track process for digital health applications (DiGA) according to Section 139e SGB V. A guide for manufacturers, service providers and users. Federal Institute for Drugs and Medical Devices. 2020. URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.html [accessed 2023-12-07]


https://mhealth.jmir.org/2024/11/e48345

JMIR Mhealth Uhealth 2024 vol. 12 e48345 p.229 (page number not for citation purposes)
https://mhealth.jmir.org/2024/1/e48345 JMIR Mhealth Uhealth 2024 | vol. 12 | e48345 | p.230 (page number not for citation purposes)


57. Pritchard KI. Have we been guilty of ageism in the primary treatment of breast cancer? Br J Cancer 2007;96(7):1011-1012 [FREE Full text] [doi: 10.1038/sj.bjc.6603697] [Medline: 17406347]


---

**Abbreviations**

- **BfArM**: Bundesinstitut für Arzneimittel und Medizinprodukte (Federal Institute for Drugs and Medical Devices)
- **DIGA**: Digitale Gesundheitsanwendungen (digital health applications)
- **DIGAV**: Digitale-Gesundheitsanwendungen-Verordnung (Digital Health Applications Ordinance)
- **DVG**: Digitale-Versorgung-Gesetz (Digital Healthcare Act)
- **DVR**: electronic health record
- **EU-GDPR**: European Union General Data Protection Regulation
- **HBM**: health belief model
- **IPA**: interpretative phenomenological analysis
- **mHealth**: mobile health
- **UN**: United Nations
- **UTAUT**: unified theory of acceptance and use of technology
- **WHO**: World Health Organization
Evidence of How Physicians and Their Patients Adopt mHealth Apps in Germany: Exploratory Qualitative Study

Schroeder T, Haug M, Georgiou A, Seaman K, Gewald H

Please cite as:
Schroeder T, Haug M, Georgiou A, Seaman K, Gewald H
Evidence of How Physicians and Their Patients Adopt mHealth Apps in Germany: Exploratory Qualitative Study
JMIR Mhealth Uhealth 2024;12:e48345
URL: https://mhealth.jmir.org/2024/1/e48345
doi:10.2196/48345
PMID:38231550

©Tanja Schroeder, Maximilian Haug, Andrew Georgiou, Karla Seaman, Heiko Gewald. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Original Paper

Engagement With a Remote Symptom-Tracking Platform Among Participants With Major Depressive Disorder: Randomized Controlled Trial

Katie M White1, PhD; Ewan Carr2, PhD; Daniel Leightley1, PhD; Faith Matcham1,3, PhD; Pauline Conde2, BSc; Yatharth Ranjan2, BSc; Sara Simblett1, PhD; Erin Dawe-Lane4, MSc; Laura Williams5, BA, MSc; Claire Henderson6,7, PhD; Matthew Hotopf1,7, PhD

1Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
2Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
3School of Psychology, University of Sussex, Falmer, United Kingdom
4Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
5NIHR MindTech MedTech Co-operative, Institute of Mental Health and Clinical Neurosciences, University of Nottingham, Nottingham, United Kingdom
6Health Services & Population Research Department, King’s College London, London, United Kingdom
7South London and Maudsley NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Katie M White, PhD
Department of Psychological Medicine
Institute of Psychiatry, Psychology and Neuroscience
King's College London
16 De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 020 7848 0002
Email: katie.white@kcl.ac.uk

Abstract

Background: Multiparametric remote measurement technologies (RMTs), which comprise smartphones and wearable devices, have the potential to revolutionize understanding of the etiology and trajectory of major depressive disorder (MDD). Engagement with RMTs in MDD research is of the utmost importance for the validity of predictive analytical methods and long-term use and can be conceptualized as both objective engagement (data availability) and subjective engagement (system usability and experiential factors). Positioning the design of user interfaces within the theoretical framework of the Behavior Change Wheel can help maximize effectiveness. In-app components containing information from credible sources, visual feedback, and access to support provide an opportunity to promote engagement with RMTs while minimizing team resources. Randomized controlled trials are the gold standard in quantifying the effects of in-app components on engagement with RMTs in patients with MDD.

Objective: This study aims to evaluate whether a multiparametric RMT system with theoretically informed notifications, visual progress tracking, and access to research team contact details could promote engagement with remote symptom tracking over and above the system as usual. We hypothesized that participants using the adapted app (intervention group) would have higher engagement in symptom monitoring, as measured by objective and subjective engagement.

Methods: A 2-arm, parallel-group randomized controlled trial (participant-blinded) with 1:1 randomization was conducted with 100 participants with MDD over 12 weeks. Participants in both arms used the RADAR-base system, comprising a smartphone app for weekly symptom assessments and a wearable Fitbit device for continuous passive tracking. Participants in the intervention arm (n=50, 50%) also had access to additional in-app components. The primary outcome was objective engagement, measured as the percentage of weekly questionnaires completed during follow-up. The secondary outcomes measured subjective engagement (system engagement, system usability, and emotional self-awareness).

Results: The levels of completion of the Patient Health Questionnaire-8 (PHQ-8) were similar between the control (67/97, 69%) and intervention (66/97, 68%) arms (P value for the difference between the arms=.83, 95% CI –9.32 to 11.65). The intervention
group participants reported slightly higher user engagement (1.93, 95% CI 1.91 to 5.78), emotional self-awareness (1.13, 95% CI 2.93 to 5.19), and system usability (2.29, 95% CI 5.93 to 10.52) scores than the control group participants at follow-up; however, all CIs were wide and included 0. Process evaluation suggested that participants saw the in-app components as helpful in increasing task completion.

**Conclusions:** The adapted system did not increase objective or subjective engagement in remote symptom tracking in our research cohort. This study provides an important foundation for understanding engagement with RMTs for research and the methodologies by which this work can be replicated in both community and clinical settings.

**Trial Registration:** ClinicalTrials.gov NCT04972474; https://clinicaltrials.gov/ct2/show/NCT04972474

**International Registered Report Identifier (IRRID):** RR2-10.2196/32653

**KEYWORDS**
remote measurement; technology; engagement; app; depression; smartphones; wearable devices; engagement; symptom tracking; self-awareness; community; mobile phone

**Introduction**

**Background**

Multiparametric remote measurement technologies (RMTs), which comprise smartphone apps and wearable devices, have the potential to revolutionize the clinical care of people with chronic, episodic health conditions [1]. Major depressive disorder (MDD) is one such condition, characterized by the relapse and remission of low mood and anhedonia over time [2]. Continuously measured longitudinal RMT data on the symptoms of MDD (mood variability, activity, cognition, and sleep) can capture a less biased picture of clinical state than retrospective self-report data [3]. Research using multiparametric sources might identify signals that could potentially predict future depressive episodes [4]. Such data could be ultimately implemented in patient self-management and shared decision-making in clinical practice [5].

It is important to understand how users engage with RMTs for depression symptom tracking. A recent systematic review found that engagement with RMTs can be measured objectively, for example, as the number of app-based symptom-tracking assessments completed, and subjectively, for example, as the perceived usability of and experience of using the RMT system [6]. Higher levels of objective engagement result in increased data availability, which, in turn, increases the validity of the machine learning approaches used for relapse prediction [7]. Objective engagement can also be used as an indicator of real-world uptake [8,9]. Further evidence suggests that increased satisfaction with mobile health apps is positively associated with the intention to continually use the tools [10]. Therefore, understanding engagement with RMTs is key to realizing their potential for relapse prediction.

Previous studies have reported inconsistent levels of engagement with RMT systems. Data completion, based on the total data expected, ranges from 42% to 82% for app-based symptom reporting and from 50% to 75% for device wear time [11]. The largest, multisite study of multiparametric RMTs for tracking depression to date, Remote Assessment of Disease and Relapse–Major Depressive Disorder (RADAR-MDD) [12], tracked 623 participants for 2 years using a smartphone app for mood tracking and a wrist-worn wearable for continuous passive data collection. The study has recently reported data availability metrics; 55.4% (345/623) of the sample completed >50% of the self-reported mood questionnaires expected to be completed, and 70.1% (437/623) had wearable heart rate data for >50% of the study days. Qualitative analyses from RADAR-MDD have revealed that the presence of a physical research team providing technological support and planned task reminders was a fundamental facilitator of long-term engagement in the study [13]. To ensure the scalability and real-world implementation of RMT systems, it is important to investigate methods that maximize engagement with RMTs while minimizing the human resources needed.

Focusing on the user interface (UI) of RMT systems is the logical first step for promoting engagement. Positioning the design of system UI within a theoretical framework of behavior change could help maximize effectiveness [14]. The Behavior Change Wheel [15] posits that researchers should begin by identifying a target behavior before considering the barriers to and facilitators of this behavior in terms of capability, opportunity, and motivation (the capability, opportunity, motivation, and behavior [COM-B] model). In the case of RMTs, the target behavior can be defined as objective engagement with symptom monitoring tasks. A series of published studies have evaluated both perceived [11,16] and experienced [13] barriers to RMT use in MDD research. Factors such as the knowledge of the utility of the research (capability), motivation linked to mood (motivation), and confirmation of logged data (opportunity) have been suggested to be prominent. The Behavior Change Wheel further provides a series of “intervention functions” best suited to address these factors, each with its own related behavior change techniques. With regard to RMTs, these have been suggested to be the provision of information from credible sources, visual feedback on behavior, and access to support.

The app design literature provides several options for incorporating behavior change techniques into RMT system design. First, following the Fogg behavioral model [17], push notifications can provide a trigger to perform a behavior, such as completing a monitoring task. Notifications can include tailored content, such as insights into the benefits of self-monitoring, which serves to simultaneously motivate the
user to respond to the notification and engage them in future tasks [18]. Second, visual incentives, such as graphs, can be embedded into the app to reflect on patterns in user progress and spark intrinsic motivation to complete future tasks [19]. Visualization can also help users manage uncertainty by attending to information about themselves [20]. A combination of qualitative and single-arm evaluation studies supports the perceived value of data visualization [21,22] and progress viewing [19] in encouraging symptom-tracking completion. Provision of contact details directly within an app can allow the user to directly and immediately access support, if required.

Without a control group, it is difficult to quantify the effect of in-app components on engagement [23]. A randomized controlled trial (RCT) of a substance abuse tracking app [20] suggested that users were 5% more likely to self-report on a day if they received a prior notification with an inspirational quote, although these results were not statistically significant. Conversely, users were 2% less likely to self-report following the provision of personalized visual data summaries; however, this main effect was significantly moderated by the prior day task completion such that those who had not completed the previous task were 36% more likely to self-report after receiving data visualization [20]. Users receiving prompts with tailored health messages, such as those highlighting the beneficial effects of symptom monitoring, were 4% more likely to engage in self-monitoring via another app for mental well-being [24]. It is important to replicate this work with multiparametric symptom monitoring systems, as it is currently unclear which combination of in-app features best promotes engagement with these technologies.

This Study

This study aimed to evaluate whether in-app components could promote engagement with a multiparametric RMT system for symptom tracking in depression. We conducted a 2-arm RCT to compare the system as usual with an adapted system that contained informative notifications, a visual progress report, and access to the research team contact details as a substitute for planned research team contact. We measured engagement as both objective and subjective concepts. This study had the following four specific objectives: (1) to describe data availability in an RCT of a multiparametric RMT system for tracking depression, (2) to test whether in-app components increase the rates of objective data completion, (3) to explore how in-app components influence the subjective experience of using the app, and (4) to understand how the components of the system are used by participants via process evaluation measures.

For objectives 2 and 3, we hypothesized that participants using the adapted app (intervention group) would have higher engagement in symptom monitoring, as measured by both objective engagement (completion of mood questionnaires) and subjective engagement (usability, utility, and emotional self-awareness).

Methods

Ethical Considerations

This study was approved by the Psychiatry, Nursing, and Midwifery Research Ethics Subcommittee at King’s College London (reference number: RESCM-20/21-21083) and registered as a clinical trial (reference number: NCT04972474). A trial protocol has been previously published [25].

Trial Design

This was a single-center, 2-arm, parallel-group RCT (participant-blinded) with 1:1 randomization conducted in London, United Kingdom. We compared a remote symptom-tracking system (RADAR-base [26]; the control arm) with a system that contained additional in-app components (the intervention arm). Both the control and intervention arms were delivered via the RADAR-base system [26] using a smartphone app and a wearable Fitbit Charge (Fitbit Inc) device. Participants in the intervention arm had additional access to (1) theoretically informed notifications, (2) progress visualization, and (3) research team contact details through the study app. All participants were asked to use the system for 12 weeks.

Data were collected at baseline (0 weeks) and follow-up (12 weeks after randomization). Participants in both arms were sent 3 symptom-tracking tasks each week via the app; Fitbit data were collected continuously.

Participants

All participants were recruited from the RADAR-MDD study between April and May 2021. The inclusion criteria were as follows: (1) previous participation in the RADAR-MDD study at the London site (which required experiencing at least 1 episode of MDD in the 2 years before enrollment), (2) consent to be contacted, (3) willingness and ability to continue to use an Android (Google LLC) smartphone (provided for use by RADAR-MDD; see the study by Matcham et al [3] for the full study protocol), and (4) willingness and ability to complete a remote enrollment session owing to the COVID-19 pandemic. Participants were excluded if they were diagnosed with one of the following comorbid psychiatric disorders: bipolar disorder, schizophrenia, psychosis, schizoaffective disorder, or dementia.

Potential participants were invited to take part (up to 3 invitations were sent per participant, as per ethical considerations) and subsequently checked for eligibility, both via email. If eligible, contact details were entered into the REDCap (Research Electronic Data Capture [27]) system, which emailed an automated link to the informed consent form and baseline questionnaires. After participants provided consent and completed the baseline questionnaires, they were sent a link to book an enrollment session (via email, phone call, or video call).

On the day of the enrollment session, the principal investigator (KMW) initiated the REDCap randomization module and generated unique QR codes to link the study devices to the RADAR-base management portal. Each participant was sent a personalized set of instructions for downloading and logging...
into the system using the QR codes at the chosen enrollment time, accompanied by a phone or video call as requested.

Participants were purposefully not contacted by the research team during the follow-up period, aside from sending 1 check-in email at the 6-week time point. However, participants were able to initiate contact with the team if they had any queries during follow-up. The research team did not make withdrawals based on “lost to follow up,” given the fundamental aims of the study; however, participants were aware that they could withdraw at any point.

Suicidal ideation was assessed at baseline and follow-up using the Inventory of Depressive Symptomatology–Self-Report [28] item “thoughts of death or suicide.” Participants who reported suicidal ideation and intent at either time point were contacted via phone call by the principal investigator, advised to contact their treating physician, and emailed a list of signposting resources.

At the 12-week end point, participants were directed to debrief information that explained the aims of the study and provided instructions for logging out of the system.

RADAR-Base

The RADAR-base system is an open-source platform that supports data collection via remote devices [3,26]. It requires users to download and log into an Android smartphone app in addition to wearing and syncing a wearable device. All participants were asked to complete the following three validated symptom-tracking tasks per week via the study app: (1) Patient Health Questionnaire-8 (PHQ-8 [29]); (2) Rosenberg Self-Esteem Scale (RSES [30]); and (3) a speech task, during which the user records themselves reading aloud a short paragraph (Multimedia Appendix 1). All tasks became available on the same day each week, 1 hour apart, beginning at the point of enrollment. All tasks had to be completed within 24 hours.

Interventions

Control Arm

Participants were sent 3 tasks per week via the RADAR-base study app, as outlined in the previous section. For each task, they received a notification on the day that the task was due that read, “Questionnaire Time. Won’t usually take longer than 3 minutes.” They were unable to view any data other than those available on the Fitbit app.

Intervention Arm

The design of the additional in-app components was grounded in behavioral theory and user research on the barriers to and facilitators of RMT use in patients with MDD [11,13,16]. The COM-B [31] framework of behavior change highlighted education, incentivization, and enablement as the most suitable forms of intervention function. Findings from research with users of the RADAR-base system allowed for the translation of these functions into tangible components tailored specifically to the needs and preferences of the target cohort [32]. It was decided that an engaging app should include notifications with information on symptom tracking from a credible source, behavioral feedback via progress visualization, and instant access to researcher contact details (see the study by White et al [25] and Multimedia Appendix 2 for a detailed overview of this process).

Participants in the intervention arm received notifications and tasks at the same time as those in the control arm but with the following additional content (Figure 1):

1. Theoretically informed notifications: the notifications included additional sentences that described the potential benefits of symptom monitoring for emotional self-awareness, clinical practice, and research. Participants were also reminded that they could complete the task “any time today.”
2. Progress visualization: participants were provided with a graph in the app that tracked the completion of the tasks. This graph could be viewed at any time from the main app home page.

3. Researcher contact details: the main app home page included a phone number, an email address, and contact hours of the research team for the reporting of technical issues or requests for support.

Measures
A summary of measures and data collection time points is presented in Table 1. The measures were identical between the intervention and control arms.
Table 1. A summary of measures and data collection points across the 12-week follow-up period.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Baseline</th>
<th>End point</th>
<th>Weekly</th>
<th>Continuously</th>
</tr>
</thead>
<tbody>
<tr>
<td>REDCap(^a) survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consent</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact information</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study devices</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociodemographics</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social environment</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical history</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIDAS(^b)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDS-SR(^c)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The World Health Organization CIDI-SF(^d)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD-7(^e)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSAS(^f)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIPQ(^g)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life events</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSRI(^h)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UES(^i)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESQ(^j)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAUQ(^k)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Active app measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-8(^l)</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>RSES(^m)</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Speech task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fitbit</strong></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Heart rate, step count, and GPS</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td><strong>Process evaluation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App use metrics</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Qualitative interviews</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

\(^a\)REDCap: Research Electronic Data Capture.  
\(^b\)LIDAS: Lifetime Depression Assessment Self-Report.  
\(^c\)IDS-SR: Inventory of Depressive Symptomatology–Self-Report.  
\(^d\)CIDI-SF: Composite International Diagnostic Interview-Short Form.  
\(^e\)GAD-7: Generalized Anxiety Disorder-7.  
\(^f\)WSAS: Work and Social Adjustment Scale.  
\(^g\)BIPQ: Brief Illness Perception Questionnaire.  
\(^h\)CSRI: Client Service Receipt Inventory.  
\(^i\)UES: User Engagement Scale.  
\(^j\)ESQ: Emotional Self-Awareness Questionnaire.  
\(^k\)MAUQ: mHealth App Usability Questionnaire.  
\(^l\)PHQ-8: Patient Health Questionnaire-8.  
\(^m\)RSES: Rosenberg Self-Esteem Scale.
Questionnaires

After registration for the study, participants completed web-based baseline questionnaires via REDCap, providing information on sociodemographics and physical and mental health history, including the presence of depression, recent life events, and service use. The principal investigator also manually extracted data pertaining to previous participation in the RADAR-MDD study, including participation length and technology use. At the 12-week time point, participants repeated these questionnaires.

Outcome Measures

The primary outcome was objective engagement with the system, measured as the number and percentage of weekly PHQ-8 questionnaires completed during follow-up (compared with the total of 12 questionnaires that were sent). Completion of 1 PHQ-8 questionnaire was defined as the completion of all 8 questions.

There were four secondary outcomes, three of which measured subjective engagement with the system:

1. User engagement: this was measured using the User Engagement Scale (UES) [33] adapted to mobile health use [34], a 30-item questionnaire measuring focused attention, perceived usability, esthetic appeal, and reward. All items are scored on a Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). Total scores are calculated by summing the scores for each item in each of the 4 subscales and dividing the resultant value by the number of items in each subscale. An overall engagement score can be calculated using the average of each subscale. A higher average score indicates higher user engagement. The UES has been widely adopted and shows good reliability and construct validity [35].

2. Emotional self-awareness: this was measured using the Emotional Self-Awareness Questionnaire (ESQ) [36], a 33-item scale measuring recognition, contextualization, and decision-making in relation to self-emotion. All items are scored on a 5-point Likert scale ranging from 0 (“never”) to 4 (“a lot”). The total score is calculated as a continuous variable that ranges from 0 to 132, with a higher score reflecting higher emotional self-awareness. The ESQ has a reliability of 0.92 and shows significant positive correlations with the Emotional Intelligence Test [36].

3. System usability: this was measured using the mHealth App Usability Questionnaire (MAUQ) for stand-alone apps [37]. The MAUQ is an 18-item scale that measures the immediate and long-term self-reported usability of an app, including its ease of use and utility for self-management (overall Cronbach α=.914). All items are scored on a 7-point Likert scale ranging from 1 (“disagree”) to 7 (“agree”). The app usability score is calculated as the sum of scores across the items for each participant, ranging from 18 to 126, with a higher score reflecting higher reported usability of the app.

4. Overall adherence to the RADAR-base system: a participant was considered to have adhered to the system if they (1) responded to at least 50% of the 3 weekly tasks and (2) >2 heart rate data points were recorded by the Fitbit device on at least 50% of the days during the 12-week intervention period. This measure was chosen to align with previous data availability reporting [12] and other studies [38].

Process evaluation measures were collected to evaluate the use of the in-app components. Quantitative measures covered app engagement, in-app interactions, and notification engagement. A total of 20 participants, split evenly across the 2 arms, were also invited at the study endpoint to qualitatively discuss their experiences with the components through a 1:1 interview (Multimedia Appendix 3).

Sample Size

Power calculations were performed based on data availability from the RADAR-MDD study [12]. To detect a difference of 25% completion of PHQ-8 tasks between the control and intervention arms, with 80% power and 95% CIs, 132 participants were required (66 per arm). We decided on 25% as the minimum difference that would be practically useful for analyses.

Randomization and Blinding

Participants were randomly allocated in a 1:1 ratio to either the control or intervention arm using simple randomization via the REDCap randomization module.

The principal investigator was unblinded to allocation to ensure that remote enrollments had been carried out successfully and had access to incoming data throughout the study. The trial data manager (DL) was blinded to arm allocation, as this information was stored elsewhere. Participants had previously used the RADAR-base system and, therefore, could not be fully blinded to arm allocation. However, the explicit aims and arm assignments of the study were not revealed until study debrief.

Statistical Methods

Sociodemographic and clinical variables at baseline were described by arm using appropriate summary statistics (counts and percentages for categorical variables and mean and SD or median and IQR for continuous variables). We reported data availability for all outcomes. Data availability for each study app task (PHQ-8, RSES, and speech task) was summarized as the median (IQR) number of weekly tasks completed. Fitbit wear time was summarized as mean (SD) days with >2 heart rate data points. The overall completion of all 4 data sources was also reported, calculated as a percentage of the total expected count (n=12) for the study app tasks and the total expected days of wear time (n=84) for the Fitbit.

The primary outcome, objective engagement, was analyzed using 2-sample 2-tailed t tests, which tested the difference in the mean percentage of PHQ-8 completion over 12 weeks between the study arms.

Three secondary outcomes (UES, ESQ, and MAUQ) were analyzed using separate linear regression models. Each model included the follow-up score as the dependent variable and arm allocation (0=control; 1=intervention) as the only covariate. Models for outcomes measured at baseline and follow-up (UES and ESQ) additionally included the baseline values of the outcomes. Differences in the combined adherence to the system (0=tol0% total data completion; 1=>50% total data completion)
were tested using Pearson chi-square test. The threshold for statistical significance in all the analyses was $P=.05$.

All outcomes were analyzed under the intention-to-treat principle using R (version 4.1; R Foundation for Statistical Computing) [39]. All data, including those from withdrawn participants, were included in the analyses.

### Supplementary Analysis

A supplementary analysis of all outcomes was conducted to estimate complier average causal effect (CACE) [40]. We defined compliers as participants in the intervention group who viewed the progress report page at least once in the 12 weeks. The CACE analyses were performed using a 2-stage least squares regression with arm allocation as the instrumental variable.

### Process Evaluation

We examined quantitative app use measures from data retrieved from Google Analytics (Google LLC). These were reported under the following three categories: (1) app engagement (user-initiated app opening and active weeks), (2) in-app interactions (questionnaire initiation, progress report viewing, viewed progress report >1 time, and progress report viewing duration), and (3) notification engagement (notifications received, notifications opened, and the percentage of notifications opened out of notifications received). The number of active weeks was calculated as the number of weeks the participant was active out of the total 12 weeks, with at least 3 screen view or user engagement metrics recorded per participant. Each indicator was summarized by arm as mean (SD) and median (IQR). The qualitative experiences of the study were thematically analyzed and reported as a brief narrative synthesis. These data will be reported in more detail elsewhere.

### Results

#### Recruitment

A total of 347 individuals were contacted between April and May 2021. Of them, 114 (32.9%) agreed to participate, and 100 (28.8%) completed an enrollment session and were enrolled in the study. Enrollment sessions took place via email (89/100, 89%), video call (9/100, 9%), or phone call (2/100, 2%). Figure 2 details the participation rate and reasons for nonparticipation.

#### Sample Characteristics

All (100/100, 100%) participants completed the baseline outcome assessment, and 87 (87%) participants completed the 12-week follow-up assessment. Among the total 100 participants, 1 (1%) participant in the intervention group withdrew from the study before the 12-week point, citing technological issues with the study apps as the main reason for withdrawal. The follow-up period was from April to September 2021.

Baseline characteristics were similar between the 2 groups (Table 2). The groups contained an equal number of participants (n=50). The mean age of the sample was 53.3 (SD 14.3) years, and 76 (76%) of the 100 participants were female. Most participants reported mild (36/100, 36%) or moderate (29/100, 29%) symptoms of depression at enrollment, as measured by the Inventory of Depressive Symptomatology–Self-Report. Overall, 12 (12%) participants reported suicidal ideation at baseline. Among the 100 participants, 59 (59%) “strongly agreed” that they were confident in using the smartphone they were using for the study, and 51 (51%) “strongly agreed” that they were confident in using the Fitbit device.
Figure 2. Participant flowchart following CONSORT (Consolidated Standards of Reporting Trials) guidelines.
Table 2. Baseline characteristics and demographics of the study sample.

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=50)</th>
<th>Control (n=50)</th>
<th>Overall (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y), mean (SD)</strong></td>
<td>55.3 (12.7)</td>
<td>51.2 (15.7)</td>
<td>53.3 (14.3)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>40 (80)</td>
<td>36 (72)</td>
<td>76 (76)</td>
</tr>
<tr>
<td>Men</td>
<td>10 (20)</td>
<td>14 (28)</td>
<td>24 (24)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or mixed ethnicity</td>
<td>3 (6)</td>
<td>3 (6)</td>
<td>6 (6)</td>
</tr>
<tr>
<td>White British</td>
<td>40 (80)</td>
<td>41 (82)</td>
<td>81 (81)</td>
</tr>
<tr>
<td>White other</td>
<td>5 (10)</td>
<td>4 (8)</td>
<td>9 (9)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (4)</td>
<td>2 (4)</td>
<td>4 (4)</td>
</tr>
<tr>
<td><strong>Total time in education (y), mean (SD)</strong></td>
<td>20.2 (3.34)</td>
<td>20.5 (3.71)</td>
<td>20.4 (3.51)</td>
</tr>
<tr>
<td><strong>Benefit receipt, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>24 (48)</td>
<td>23 (46)</td>
<td>47 (47)</td>
</tr>
<tr>
<td>No</td>
<td>26 (52)</td>
<td>27 (54)</td>
<td>53 (53)</td>
</tr>
<tr>
<td><strong>Income (£; US $), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;15,000 (US $18,828.67)</td>
<td>9 (18)</td>
<td>12 (24)</td>
<td>21 (21)</td>
</tr>
<tr>
<td>15,000-24,000 (US $18,828.67-$30,125.88)</td>
<td>8 (16)</td>
<td>9 (18)</td>
<td>17 (17)</td>
</tr>
<tr>
<td>24,000-40,000 (US $30,125.88-$50,209.8)</td>
<td>15 (30)</td>
<td>10 (20)</td>
<td>25 (25)</td>
</tr>
<tr>
<td>40,000-55,000 (US $50,209.8-$69,038.47)</td>
<td>11 (22)</td>
<td>7 (14)</td>
<td>18 (18)</td>
</tr>
<tr>
<td>&gt;55,000 (US $69,038.47)</td>
<td>7 (14)</td>
<td>12 (24)</td>
<td>19 (19)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>24 (48)</td>
<td>25 (50)</td>
<td>49 (49)</td>
</tr>
<tr>
<td>Sick leave</td>
<td>0 (0)</td>
<td>3 (6)</td>
<td>3 (3)</td>
</tr>
<tr>
<td>Retired</td>
<td>20 (40)</td>
<td>14 (28)</td>
<td>34 (34)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>2 (4)</td>
<td>4 (8)</td>
<td>6 (6)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (8)</td>
<td>4 (8)</td>
<td>8 (8)</td>
</tr>
<tr>
<td><strong>Current depression (continuous), mean (SD)</strong></td>
<td>24.8 (13.7)</td>
<td>26.5 (13.3)</td>
<td>25.7 (13.5)</td>
</tr>
<tr>
<td><strong>Current depression (categorical), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>10 (20)</td>
<td>8 (16)</td>
<td>18 (18)</td>
</tr>
<tr>
<td>Mild</td>
<td>20 (40)</td>
<td>16 (32)</td>
<td>36 (36)</td>
</tr>
<tr>
<td>Moderate</td>
<td>12 (24)</td>
<td>17 (34)</td>
<td>29 (29)</td>
</tr>
<tr>
<td>Severe</td>
<td>5 (10)</td>
<td>7 (14)</td>
<td>12 (12)</td>
</tr>
<tr>
<td>Very severe</td>
<td>3 (6)</td>
<td>2 (4)</td>
<td>5 (5)</td>
</tr>
<tr>
<td><strong>Suicidal ideation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3 (6)</td>
<td>9 (18)</td>
<td>12 (12)</td>
</tr>
<tr>
<td>No</td>
<td>47 (94)</td>
<td>41 (82)</td>
<td>88 (88)</td>
</tr>
<tr>
<td><strong>Current anxiety (continuous), mean (SD)</strong></td>
<td>6.34 (4.62)</td>
<td>7.10 (5.21)</td>
<td>6.72 (4.92)</td>
</tr>
<tr>
<td><strong>Current anxiety (categories), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>22 (44)</td>
<td>17 (34)</td>
<td>39 (39)</td>
</tr>
<tr>
<td>Mild</td>
<td>16 (32)</td>
<td>20 (40)</td>
<td>36 (36)</td>
</tr>
<tr>
<td>Moderate</td>
<td>9 (18)</td>
<td>6 (12)</td>
<td>15 (15)</td>
</tr>
<tr>
<td>Severe</td>
<td>3 (6)</td>
<td>7 (14)</td>
<td>10 (10)</td>
</tr>
</tbody>
</table>
### Medical comorbidity, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=50)</th>
<th>Control (n=50)</th>
<th>Overall (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>25 (50)</td>
<td>34 (68)</td>
<td>59 (59)</td>
</tr>
<tr>
<td>No</td>
<td>25 (50)</td>
<td>16 (32)</td>
<td>41 (41)</td>
</tr>
</tbody>
</table>

### Functional disability, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=50)</th>
<th>Control (n=50)</th>
<th>Overall (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No impairment</td>
<td>14 (28)</td>
<td>17 (34)</td>
<td>31 (31)</td>
</tr>
<tr>
<td>Some impairment</td>
<td>17 (34)</td>
<td>17 (34)</td>
<td>34 (34)</td>
</tr>
<tr>
<td>Significant impairment</td>
<td>19 (38)</td>
<td>16 (32)</td>
<td>35 (35)</td>
</tr>
</tbody>
</table>

### Life events in the past year, mean (SD)

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=50)</th>
<th>Control (n=50)</th>
<th>Overall (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.800 (1.05)</td>
<td>0.920 (1.07)</td>
<td>0.800 (1.05)</td>
</tr>
</tbody>
</table>

### Confidence in smartphone use, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=50)</th>
<th>Control (n=50)</th>
<th>Overall (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly agree</td>
<td>29 (58)</td>
<td>30 (60)</td>
<td>59 (59)</td>
</tr>
<tr>
<td>Agree</td>
<td>15 (30)</td>
<td>16 (32)</td>
<td>31 (31)</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>5 (10)</td>
<td>2 (4)</td>
<td>7 (7)</td>
</tr>
<tr>
<td>Disagree</td>
<td>0 (0)</td>
<td>2 (4)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

### Confidence in Fitbit (Fitbit Inc) use, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=50)</th>
<th>Control (n=50)</th>
<th>Overall (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly agree</td>
<td>27 (54)</td>
<td>24 (48)</td>
<td>51 (51)</td>
</tr>
<tr>
<td>Agree</td>
<td>16 (32)</td>
<td>22 (44)</td>
<td>38 (38)</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>6 (12)</td>
<td>3 (6)</td>
<td>9 (9)</td>
</tr>
<tr>
<td>Disagree</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Not using Fitbit</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>2 (2)</td>
</tr>
</tbody>
</table>

### Existing RADAR-MDD\(^c\) status, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=50)</th>
<th>Control (n=50)</th>
<th>Overall (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finished 2 years</td>
<td>0 (0)</td>
<td>2 (4)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Continuing past 2 years</td>
<td>13 (26)</td>
<td>17 (34)</td>
<td>30 (30)</td>
</tr>
<tr>
<td>Not reached 2 years</td>
<td>36 (72)</td>
<td>31 (62)</td>
<td>67 (67)</td>
</tr>
<tr>
<td>Withdrawn</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

### Existing phone status, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=50)</th>
<th>Control (n=50)</th>
<th>Overall (N=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing Android (Google LLC)</td>
<td>30 (60)</td>
<td>27 (54)</td>
<td>57 (57)</td>
</tr>
<tr>
<td>Switching from iPhone (Apple Inc)</td>
<td>13 (26)</td>
<td>14 (28)</td>
<td>27 (27)</td>
</tr>
<tr>
<td>Switching from nonsmartphone</td>
<td>2 (4)</td>
<td>4 (8)</td>
<td>6 (6)</td>
</tr>
<tr>
<td>Upgrading existing Android</td>
<td>5 (10)</td>
<td>5 (10)</td>
<td>10 (10)</td>
</tr>
</tbody>
</table>

\(^a\)Measured using the Inventory of Depressive Symptomatology–Self-Report. The maximum score possible is 84. The scores are categorized as follows: none=0-13, mild=14-25, moderate=26-38, severe=39-48, and very severe=49-84.

\(^b\)Measured using the Generalized Anxiety Disorder-7 item scale. The maximum score possible is 21. The scores are categorized as follows: none=0-5, mild=6-10, moderate=11-15, and severe=16-21.

\(^c\)RADAR-MDD: Remote Assessment of Disease and Relapse–Major Depressive Disorder.

### Data Availability

Of the 100 participants, 97 (97%) provided any data via the study app, and 93 (93%) had any recorded Fitbit data. A total of 2 (2%) participants were unable to use either the study app or Fitbit immediately following enrollment, 1 (1%) received no notifications from the study app during follow-up, and 4 (4%) were unable to sync the Fitbit with their smartphone. Moreover, 1 (1%) participant opted out of using a Fitbit for the study at enrollment. As data from the sources were unavailable owing to technical limitations, rather than nonengagement, these participants were excluded from the respective analyses (3/100, 3% for primary analysis and 8/100, 8% for the secondary combined adherence analysis).

Each app task had a maximum count of 12 (1 per week). Overall, participants completed a median of 9 (IQR 6-10) PHQ-8 tasks, 9 (IQR 6-10) RSES tasks, and 8 (IQR 6-9) speech tasks. Among the 100 participants, 2 (2%) completed all available tasks, and 7 (7%) completed all available PHQ-8 tasks. A total of 35 (35%)...
participants completed all 3 tasks at each point when they completed a PHQ-8 task.

The participants provided sufficient Fitbit data (at least 2 recordings per day) on a mean average of 74 (SD 19.7) days per participant during the 12-week (84-day) intervention period.

Figure 3 shows the percentage of completion for all 4 data sources across the sample.

**Primary and Secondary Outcomes**

The primary analysis included 97 participants. The levels of completion of the PHQ-8 task were similar between the control (67/97, 69%) and intervention (66/97, 68%) arms (P value for the difference between the arms=.83, 95% CI −9.32 to 11.65).

For the secondary outcomes (Table 3), we found that those in the intervention group reported slightly higher UES (1.93, 95% CI −1.91 to 5.78), ESQ (1.13, 95% CI −2.93 to 5.19), and MAUQ (2.29, 95% CI −5.93 to 10.52) scores than those in the control group at follow-up. However, all CIs were wide and included 0.
Table 3. Linear regression model coefficients for each of the 3 secondary outcomes.

<table>
<thead>
<tr>
<th>Subjective engagement outcome</th>
<th>Treatment effect (95% CI)</th>
<th>Participant (N=100), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UES&gt;b</td>
<td>1.93 (−1.91 to 5.78)</td>
<td>89 (89)</td>
</tr>
<tr>
<td>ESQ&gt;c</td>
<td>1.13 (−2.93 to 5.19)</td>
<td>89 (89)</td>
</tr>
<tr>
<td>MAUQ&gt;d,e</td>
<td>2.29 (−5.93 to 10.52)</td>
<td>87 (87)</td>
</tr>
</tbody>
</table>

*a* For end point measures only.  
*b* UES: User Engagement Scale.  
*c* ESQ: Emotional Self-Awareness Questionnaire.  
*d* MAUQ: mHealth App Usability Questionnaire.  
*e* Only includes end point measure.

The combined adherence secondary analysis included 92 participants. The proportion of participants adhering to the system was similar between both arms (control=32/48, 67%; intervention=35/44, 80%; *P* value for difference between arms=.98).

For the CACE analyses, of the 48 participants in the intervention group, 29 (60%) met the complier definition of viewing the progress report at least once during the intervention period (Table 4). Table 5 presents the CACE treatment estimates for the primary and secondary outcomes. The between-arm difference in PHQ-8 completion was −1.92 (95% CI −19.93 to 15.54; *P*=.83), showing no evidence of a statistical difference. The treatment effect estimates for the UES, ESQ, and MAUQ were larger than the intention-to-treat estimates in favor of the intervention, but the effect sizes remained small.

Table 4. Number of participants and mean percentage of completion among the control group, intervention group compliers, and intervention group noncompliers (n=97).

<table>
<thead>
<tr>
<th>Study arm</th>
<th>Participant, n (%)</th>
<th>Completion (%), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention group compliers*a</td>
<td>29 (30)</td>
<td>75.3 (23.9)</td>
</tr>
<tr>
<td>Intervention group noncompliersb</td>
<td>19 (20)</td>
<td>57.0 (28.0)</td>
</tr>
<tr>
<td>Control group</td>
<td>49 (50)</td>
<td>69.2 (25.1)</td>
</tr>
</tbody>
</table>

*a* Viewed the progress report module at least once during the intervention period.  
*b* Did not view the progress report module during the intervention period.

Table 5. Intention-to-treat (ITT) and complier average causal effect (CACE) treatment estimates for the primary and secondary outcomes.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Participant, n (%)</th>
<th>ITT Treatment effect (95% CI)</th>
<th><em>P</em> value</th>
<th>CACEa Treatment effect (95% CI)</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHQ-8b</td>
<td>97 (100)</td>
<td>−1.16 (−11.65 to 9.32)</td>
<td>.83</td>
<td>−1.92 (−19.39 to 15.54)</td>
<td>.83</td>
</tr>
<tr>
<td>UESc</td>
<td>89 (92)</td>
<td>1.93 (−1.91 to 5.78)</td>
<td>.32</td>
<td>3.49 (−3.75 to 10.73)</td>
<td>.34</td>
</tr>
<tr>
<td>ESQd</td>
<td>89 (92)</td>
<td>1.13 (−2.93 to 5.19)</td>
<td>.58</td>
<td>2.03 (−5.32 to 9.38)</td>
<td>.58</td>
</tr>
<tr>
<td>MAUQe</td>
<td>87 (90)</td>
<td>2.29 (−5.93 to 10.52)</td>
<td>.58</td>
<td>4.21 (−10.87 to 19.28)</td>
<td>.58</td>
</tr>
</tbody>
</table>

*a* Complier average causal effect estimates of intervention group compliers, defined as those who viewed the progress report module at least once during the intervention period.  
*b* PHQ-8: Patient Health Questionnaire-8.  
*c* UES: User Engagement Scale.  
*d* ESQ: Emotional Self-Awareness Questionnaire.  
*e* MAUQ: mHealth App Usability Questionnaire.

Process Evaluation

Table 6 presents the quantitative process evaluation measures collected throughout the study. Over the entire study period, participants in the intervention arm opened the app a mean average of 21.2 (SD 13.5) times, whereas the participants in the control group opened the app a mean average of 19.0 (SD 9.10) times. In total, 60% (35/58) of the participants who were able to view the progress report viewed it multiple times throughout the study, viewing for a mean average of 14.7 (SD 10.9) seconds per time. Participants in both groups received a similar number of notifications, although those in the control group opened a higher mean percentage (39.9, SD 25.9) of the notifications.
Table 6. Process evaluation use statistics by arm over the 12-week follow-up period.

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=48)</th>
<th>Control (n=49)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>App engagement, mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User-initiated app opening</td>
<td>21.2 (13.5)</td>
<td>19.0 (9.10)</td>
</tr>
<tr>
<td>Active weeks(^a)</td>
<td>8.96 (3.14)</td>
<td>8.88 (2.60)</td>
</tr>
<tr>
<td><strong>In-app interactions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaire initiation, mean (SD)</td>
<td>23.6 (10.3)</td>
<td>25.0 (10.5)</td>
</tr>
<tr>
<td><strong>Progress report viewing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD)</td>
<td>3.60 (7.64)</td>
<td>N/A(^b)</td>
</tr>
<tr>
<td>Values, median (IQR)</td>
<td>1.00 (0-2.25)</td>
<td>N/A</td>
</tr>
<tr>
<td>Viewed progress report &gt;1 time , n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>19 (40)</td>
<td>49 (100)</td>
</tr>
<tr>
<td>Yes</td>
<td>29 (60)</td>
<td>N/A</td>
</tr>
<tr>
<td>Progress report viewing duration (seconds), mean (SD)</td>
<td>14.7 (10.9)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Notification engagement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notifications received, mean (SD)</td>
<td>22.0 (13.3)</td>
<td>22.6 (9.82)</td>
</tr>
<tr>
<td>Notifications opened, mean (SD)</td>
<td>6.58 (6.45)</td>
<td>8.69 (6.36)</td>
</tr>
<tr>
<td>Percentage of notifications opened, mean (SD)(^c)</td>
<td>34.3 (31.8)</td>
<td>39.9 (25.9)</td>
</tr>
<tr>
<td>None received, n (%)</td>
<td>2 (4)</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

\(^a\)Calculated as the number of weeks over the 12-week period with at least 3 screen view or user engagement metrics recorded, as per Google Analytics (Google LLC) data.

\(^b\)N/A: not applicable; participants in the control arm were unable to view the progress report.

\(^c\)Percentage of notifications opened based on the total notifications received for each participant.

Participants generally liked the new in-app components. They felt that the progress report could motivate task completion by providing clarity that previous tasks had been successfully logged:

> [It allowed me] certainly to feel more engaged and understand...how it contributes, but also to gain an overview of my own input into it, so not just I enter the data and it disappears. [P99]

Most participants found the notifications somewhat informative:

> The pop up things with little quotes about “doing this helps you”...yeah I liked those, I thought that was really good. You’re doing it for a purpose. [P29]

However, many participants were unsure whether they had seen all the notifications that their phone had received. Some highlighted the potentially demotivating effects of the progress report, depending on previous completion:

> I think it depends what mood you’re in...if I hadn’t completed everything and I wasn’t in a good mind space I could be thinking “ooh I’ve failed.” [P40]

Whereas most participants agreed that the components might motivate others, the impact of the components on participants’ own task completion was more nuanced. Instead, participating in the research study seemed to be the strongest motivation for task completion:

> Because I had committed to do the study it meant that I said I am going to do it so I can’t be half-hearted about it...I want to do the best I could because it was for somebody else’s use.” [P99]

Many participants discussed the beneficial effects of taking part in symptom monitoring generally, such as increased awareness of their depression and communication with others. Several additional in-app components were suggested, including a direct communication channel between the app and research team.

**Harms and Protocol Violations**

No adverse or serious adverse events were reported. Among the 100 participants, 1 (1%) withdrew owing to technological issues.

**Discussion**

**Principal Findings**

This study conducted the first, fully remote RCT of the RADAR-base symptom-tracking system to test the effect of additional in-app components, based on behavioral change theory, on objective and subjective engagement. Overall, objective engagement was high across the sample. We found that the participants who received the adapted system (incorporating theoretically informed notifications, real-time progress reports, and researcher contact details) did not show higher levels of engagement with the system than the participants who received the system as usual. Although
subjective engagement (emotional self-awareness, system utility, and usability) was slightly higher in those who received the adapted app, the difference was small and did not reach statistical significance.

Implications and Links With Previous Work

Previous research, both usability studies [19,21,22] and RCTs [20,24], has suggested that providing notifications and progress visualization can prompt objective engagement in remote symptom monitoring. We propose several explanations as to why our results did not reflect past findings.

First, our findings may reflect the sample used. Participants were recruited from a previous study that used the RADAR-base system. This meant that they had prior experience of and interest in symptom monitoring. Previous work has also highlighted the impact of the academic setting on engagement through altruistic motivations [41]. It is possible that our results reflect a ceiling effect, whereby participants in both groups were motivated to participate in the research and complete symptom monitoring regardless of the changes to the app. This is particularly apparent given that 2 of the in-app components were designed to reflect individual achievement and benefits, aspects that might not have been as relevant in this research context.

Second, the combination of in-app components used in the adapted system might not have been sufficiently tailored to the user. The development of the app was grounded in both behavioral theory [31] and user involvement [21], which suggested that viewing real-time progress and being reminded of the proposed benefits of symptom monitoring might combat the barriers to engagement. However, although these components are proposed to encourage future tracking behavior, in practice, it is unclear how they interact with the motivation section of the COM-B model, in this case, the low motivation linked to low mood in our cohort of people with depression. Previous work has focused on symptom tracking for substance abuse [20] or general population [18,24] cohorts, both of which might react to incentivization in different ways from those with depression. Our qualitative discussions indeed suggested that the impact of viewing data progress might be affected by individual mood and motivational fluctuations. The addition of other virtual incentives, such as gamification [42], might have been more effective in promoting engagement with the tasks here, alongside the ability to personalize which components are seen and when they are seen.

Moreover, our components were static in that they were accessible to all the participants in the intervention group at the same time and frequency. Previous work has suggested that several factors can significantly moderate the relationship between in-app components and engagement. For example, Nahum-Shani and colleagues [20] found that receiving data insights only increased the likelihood of future self-reporting in those who were not frequent users of the app, suggesting that visualizing progress is not incentivizing (or is even perhaps “irritating”) for those who are actively engaged in the task from the beginning. Several studies have found a link between notification timing and engagement [20,24], although attempts at sensor-driven notification sending based on location have so far been unsuccessful in improving data availability [43]. Taken together, this suggests that future work is needed to understand the process of interacting with in-app components in this cohort.

Third, with regard to subjective engagement, the measures used in this study might not have reflected the experience of self-monitoring in the most nuanced manner. We used previous findings [6,41] to inform our operationalization of subjective engagement with RMTs as usability (UES), emotional self-awareness (ESQ), and utility (MAUQ). Our qualitative evaluation suggested that participants generally saw the in-app components as helpful in increasing task completion, which, in turn, might have promoted the feelings of emotional self-awareness they gained from monitoring their symptoms. We also saw that participants who viewed the progress report did so for around 15 seconds at a time, often repeatedly, which suggests a sustained interest in viewing progress. Although we did not see significant differences in either objective or subjective engagement, we did see slight treatment effects for all 3 subjective measures, which were higher still when adjusting for those who viewed the progress report. It is possible that different measures might have revealed a more significant change. For example, the UES is a tool designed primarily for digital health interventions and measures concepts such as focused attention, which are not as relevant to RMTs [6]. Measures tapping into other aspects of the experience of symptom monitoring, such as being seen as an individual [44] or the provision of a safety net [45], might have provided a more detailed understanding of the interaction among the in-app components, objective engagement, and subjective engagement in the study; however, to our knowledge, these have yet to be developed.

Strengths and Limitations

To our knowledge, this was the first study to attempt to quantify the effect of in-app components on objective and subjective engagement with a multiparametric symptom-tracking system for depression. We used an established system that was previously used to conduct the largest, longitudinal study on RMT in MDD to date [12] and demonstrated the successful transference of the system to a remote RCT design. Methodologically, this study laid the foundation for future work to measure both objective and subjective engagements with symptom-tracking devices. We used an adapted system with in-app components, which allowed for an active control group (the system as usual) and embedded data collection to reduce confounding factors associated with the delivery medium [46]. In reference to our first aim, we have shown good data availability in the first fully remote trial of the system, with 87% (87/100) of the participants completing follow-up data collection, a median of 75% completion of symptom-tracking tasks, and a mean of 74 of 84 days of wear time data without planned researcher contact.

There are several limitations to this study. First, as mentioned, the sample was previously engaged in remote symptom tracking and driven by research altruism. This allowed for the recruitment of a large sample from an established group, obtaining results quickly and efficiently. However, it is unclear how far these results might generalize to community cohorts using symptom tracking in their daily life. Second, the study was conducted...
during the COVID-19 lockdown periods in the United Kingdom. A combination of increased free time and interest in health tracking could have resulted in increased engagement rates. Third, despite the large sample size, the study did not reach the intended number of participants needed to achieve the optimum statistical power. Fourth, although the app design was grounded in previous research, working within the confines of an established system gave way to certain design constraints. Some additional facilitators that arose from the COM-B analysis, such as the in-app reporting of technological malfunctions, could not be included or assessed in terms of their impact on engagement.

Avenues for Future Work
Future work should use these findings as a basis for further RCTs quantifying the effects of RMT system design on objective and subjective engagement with remote symptom tracking. Context-specific, dynamic tailoring of notifications and data insights could be key here. Although in-app components reduce the need for human resources, the impact of external factors should not be dismissed. Our system amendments did not promote engagement over and above the system as usual; future work could seek to understand how incentives such as research team support could interact with in-app components to increase engagement, such as the use of supportive chatbots [23]. Of major importance is replicating this work with different cohorts. Using the adapted system with non–help-seeking participants or those with lower technological literacy might affect the impact of the components that we tested. For example, the impact of the theoretically informed notifications might be greater in those who are less aware of the proposed benefits of symptom monitoring. Similarly, engagement with the app is likely to vary if the app is implemented in clinical practice; progress tracking and notification content might be more impactful for those who use the system for their own direct benefit. This work could also seek to complement the RCT design with additional analysis manuscripts for increased insight into the impact of UI features. For example, this could include correlational analyses of in-app component use with the measures of objective and subjective engagement or exploring whether baseline demographics are predictive of engagement in such trials. Another area for exploration is the measurement of the subjective experience of remote symptom tracking. The development of a suitable instrument that encapsulates experiential engagement would propel the understanding of the promotion of engagement across the field.

Conclusions
This study found that a combination of informative notifications, progress visualization, and research team contact details did not increase engagement in remote symptom tracking in our research cohort. However, the system provided good data availability, and the process evaluation measures suggested that participants saw benefits in using the adapted system. We have provided the methodology and scope for future exploration in this area, as well as opportunities to replicate this work in both community and clinical cohorts to further the promotion of engagement in remote health symptom tracking for both data collection and clinical management.

Acknowledgments
This paper represents independent research partly funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or Department of Health and Social Care.

The Remote Assessment of Disease and Relapse–Central Nervous System project received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement number 115902. This joint undertaking receives support from the European Union’s Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations (EFPIA). This communication reflects the views of the RADAR-CNS consortium, and neither Innovative Medicines Initiative nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. The funding body was not involved in the design of the study or the collection, analysis, or interpretation of the data. This research was reviewed by a team with experience of mental health problems and their carers who were specially trained to advise on research proposals and documentation through the Feasibility and Acceptability Support Team for Researchers: a free, confidential service in England provided by the NIHR Maudsley Biomedical Research Centre via King’s College London and South London and Maudsley NHS Foundation Trust. Finally, the authors would like to thank all members of the RADAR-CNS patient advisory board who have experience of living with or supporting those who are living with depression, epilepsy, or multiple sclerosis.

Authors’ Contributions
KMW designed, coordinated, and conducted the study as well as led data processing, data coding, data analysis, and the writing of this manuscript. EC contributed to the design of the study and advised on data analysis. DL contributed to the design of the study and aided with data extraction. FM contributed to the design of the study. PC contributed to in-app component development and data extraction. YR contributed to app development and data extraction. SS contributed to service user involvement research and in-app component development. ED-L contributed to service user involvement research and in-app component development. LW contributed to the coding and thematic analysis of the interview data. CH contributed to the development and design of the study, reviewed data analysis, and supervised the first author. MH contributed to the development and design of the study, reviewed data analysis, and supervised the first author.
Conflicts of Interest

MH is the principal investigator of the RADAR-CNS consortium, a private-public precompetitive consortium with research funding from Janssen, Union Chimique Belge, Merck Sharp & Dohme, Biogen, and Lundbeck. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Speech task paragraph.
[DOCX File, 10 KB - mhealth_v12i1e44214_app1.docx]

Multimedia Appendix 2
Development of the adapted system based on the capability, opportunity, motivation, and behavior framework and service user research.
[DOCX File, 243 KB - mhealth_v12i1e44214_app2.docx]

Multimedia Appendix 3
Semistructured interview schedules.
[DOCX File, 21 KB - mhealth_v12i1e44214_app3.docx]

Multimedia Appendix 4
CONSORT checklist.
[PDF File (Adobe PDF File), 65 KB - mhealth_v12i1e44214_app4.pdf]

References


Abbreviations

- **CACE**: complier average causal effect
- **COM-B**: capability, opportunity, motivation, and behavior
- **ESQ**: Emotional Self-Awareness Questionnaire
- **MAUQ**: mHealth App Usability Questionnaire
- **MDD**: major depressive disorder
- **PHQ-8**: Patient Health Questionnaire-8
- **RADAR-MDD**: Remote Assessment of Disease and Relapse–Major Depressive Disorder
- **RCT**: randomized controlled trial
- **REDCap**: Research Electronic Data Capture
- **RMT**: remote measurement technology
- **RSES**: Rosenberg Self-Esteem Scale
- **UES**: User Engagement Scale
- **UI**: user interface
Engagement With a Remote Symptom-Tracking Platform Among Participants With Major Depressive Disorder: Randomized Controlled Trial


URL: https://mhealth.jmir.org/2024/1/e44214
doi: 10.2196/44214
PMID: 38241070

©Katie M White, Ewan Carr, Daniel Leightley, Faith Matcham, Pauline Conde, Yatharth Ranjan, Sara Simblett, Erin Dawe-Lane, Laura Williams, Claire Henderson, Matthew Hotopf. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org/), 19.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Effectiveness of a Nurse-Led Mobile-Based Health Coaching Program for Patients With Prostate Cancer at High Risk of Metabolic Syndrome: Randomized Waitlist Controlled Trial

Kyoungjin Lee¹,², PhD; Jeongok Park³, PhD; Eui Geum Oh³, PhD; JuHee Lee³, PhD; Chang Park⁴, PhD; Young Deuk Choi⁵, PhD

¹College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea
²College of Nursing, Kyungbok University, Namyangju, Republic of Korea
³Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea
⁴Department of Population Health Nursing Science, College of Nursing, University of Illinois at Chicago, Chicago, IL, United States
⁵College of Medicine, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
Jeongok Park, PhD
Mo-Im Kim Nursing Research Institute, College of Nursing
Yonsei University
Seodaemun-gu, Seoul, Korea
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3390
Fax: 82 2 2227 8303
Email: JOPARK02@yuhs.ac

Abstract

Background: Androgen deprivation therapy (ADT), a standard treatment for prostate cancer (PC), causes many physical side effects. In particular, it causes metabolic changes such as fasting glucose abnormalities or accumulation of body fat, and its continuation can lead to metabolic syndrome (MetS), which is closely related to diabetes and cardiovascular disease. Therefore, it is important to maintain and practice a healthy lifestyle in patients with PC.

Objective: This study aims to evaluate the effectiveness of a nurse-led mobile-based program that aims to promote a healthy lifestyle in patients with PC undergoing ADT with MetS risk factors.

Methods: This was a single-blind, randomized, waitlist control interventional study. A total of 48 patients were randomly assigned to the experimental and waitlist control groups at the urology cancer clinic of a tertiary general hospital in South Korea. The inclusion criteria were patients who had undergone ADT for >6 months, had at least 1 of the 5 MetS components in the abnormal range, and could access a mobile-based education program. The experimental group attended a 4-week mobile-based program on exercise and diet that included counseling and encouragement to maintain a healthy lifestyle, whereas the control group was placed on a waitlist and received usual care during the follow-up period, followed by the intervention. The primary outcome was a change in the lifestyle score. The secondary outcomes were changes in 5 MetS components, body composition, and health-related quality of life. The outcomes were measured at 6 weeks and 12 weeks after the initiation of the intervention. Each participant was assigned to each group in a sequential order of enrollment in a 4x4 permuted block design randomization table generated in the SAS (SAS Institute) statistical program. A linear mixed model was used for statistical analysis.

Results: A total of 24 participants were randomly assigned to each group; however, 2 participants in the experimental group dropped out for personal reasons before starting the intervention. Finally, 46 participants were included in the intention-to-treat analysis. The experimental group showed more positive changes in the healthy lifestyle score ($\beta=29.23; P<.001$), level of each MetS component (fasting blood sugar: $\beta=-12.0; P=.05$ and abdominal circumference: $\beta=-2.49; P=.049$), body composition (body weight: $\beta=-1.52; P<.001$ and BMI: $\beta=-0.55; P<.001$), and the urinary irritative and obstructive domain of health-related quality of life ($\beta=14.63; P<.001$) over time than the waitlist control group.
Conclusions: Lifestyle changes through nurse-led education can improve level of each MetS components, body composition, and ADT side effects. Nurses can induce positive changes in patients’ lifestyles and improve the self-management of patients starting ADT through this program.

Trial Registration: Clinical Research Information Service KCT0006560; http://tinyurl.com/yhvj4vwh

KEYWORDS
nurses; prostate neoplasms; healthy lifestyle; metabolic syndrome; exercise; diet; mobile phone

Introduction

Adverse Effects of Androgen Deprivation Therapy in Prostate Cancer

Prostate cancer (PC) is a commonly occurring cancer in men worldwide, and it is the fourth most common cancer among adult men in South Korea [1]. The incidence of PC in South Korea has increased rapidly in recent years, from 2.2% in 2000 to 14.3% in 2018 [1], owing to changes in dietary patterns and the development of new diagnostic technologies. PC is affected by androgenic activity in the body [2]. Therefore, the goal of treatment is to either remove androgens using castration or neutralize the effects of androgens. Androgen deprivation therapy (ADT), which disrupts the mechanisms that create androgens, is a standard treatment for PC [3]. Initially, ADT was used to treat patients with metastatic PC or older persons with cancer with limited treatment options. However, in recent years, there has been a gradual expansion in the use of ADT to include treating patients in the early stages of PC [4,5].

Repeated ADT can lead to climacteric symptoms in male individuals [6], including sexual dysfunction, hot flashes [7], gynecomastia [8], depression, fatigue, changes in sleep patterns [9], loss of muscle strength [10,11], osteoporosis [12], metabolic syndrome (MetS) [13], and cardiovascular disease [14,15]. These issues can reduce the quality of life of patients and may even be life threatening [16,17]. Specifically, evidence suggests that ADT leads to metabolic changes [13,18]. MetS, also called insulin resistance syndrome, is the clustering of several risk factors associated with obesity. It is closely related to diabetes [14,15,19] and has also been identified as a major cause of cardiovascular disease owing to its association with dyslipidemia, diabetes, and hypertension [14,15]. PC and MetS have a close relationship: patients with PC who received ADT were found to be 2.5 times more likely to have MetS than those who did not receive ADT [18]. A cross-sectional analysis using Korean National Health Insurance Service data found that the prevalence rate of MetS in patients with PC was 40.1%, which is higher than the 34.5% prevalence rate in older Korean male adults [20]. In addition, as age increases, MetS has been found to increase by 8 times in patients with PC [21]. Given that most patients with PC are older adults, ADT-induced metabolic problems coupled with aging can lead to deterioration of health, resulting in cardiovascular disease or death [22]. Despite the necessity of severe side effects management for ADT, >50% of patients with PC are unaware of these problems [23], and most health care providers tend to focus on cancer treatment. Moreover, no protocols have been developed to manage ADT-induced MetS in patients with PC [24,25].
challenging to maintain. In addition, the COVID-19 pandemic has diminished physical activity and nutritional quality worldwide [37]. The World Health Organization has emphasized the importance of exercise and maintaining a healthy lifestyle during the pandemic [38]. Web-based education has relatively few limitations in terms of time and space, and it has the advantage of being able to feature various types of media and teaching and learning materials compared with offline education programs [39].

Aims
This study aimed to evaluate the effectiveness of a nurse-led mobile-based health coaching program that promotes healthy lifestyle changes, normal range of MetS components, and health-related quality of life (HRQoL) for patients with PC receiving ADT. The primary aim was to identify the changes in (1) healthy lifestyle through this program, and the secondary aims were to identify the changes in (1) the levels of each MetS component, including blood pressure, fasting blood sugar (FBS), high-density lipoprotein (HDL) cholesterol, triglyceride, and abdominal circumference (AC); (2) body composition, such as body weight, BMI, skeletal muscle mass, fat mass, and fat percentage; and (3) quality of life.

Methods

Participants
The study population consisted of men who were diagnosed with PC at the urology cancer center of a single tertiary general hospital in South Korea. The inclusion criteria were (1) patients with PC who had been receiving ADT for >6 months at the time of enrollment in this study, (2) those with an abnormal range of at least 1 of the 5 MetS components, (3) those who were fully aware of the object and contents of the study and voluntarily participated, (4) those who understood spoken and written Korean and who could communicate without cognitive impairment, and (5) those who had a smartphone and were able to access the mobile-based education program. The exclusion criteria were (1) patients diagnosed with and treated for cancers other than PC; (2) those who had undergone surgery or chemotherapy for <3 months before the start of the study; (3) those who answered “yes” to ≥1 question of the Physical Activity Readiness Questionnaires; (4) those with cardiovascular diseases such as unstable angina pectoris, uncontrolled blood pressure, myocardial infarction, or comorbidities such as a musculoskeletal or nervous system disease; (5) those who were participating in other research programs; (6) those who had a change in medication to control their blood pressure, cholesterol, or blood sugar <3 months before the start of the study; and (7) those who had difficulty with typical daily activities.

Ethical Considerations
This study was approved by the Institutional Review Board of the Severance Hospital, Yonsei University Health System (4-2020-0889). We informed all participants of the purpose of this study, process, methods, voluntary participation terms, and the possible risks and benefits of participation. Furthermore, we gave the participants 1 week to consider their participation in the study before deciding whether to sign the informed consent form. All study participants were compensated with KRW 50,000 (approximately US $43) as a gift at both the beginning and end of the study.

Enrollment Process
To recruit participants, we posted recruitment announcements on the bulletin board of a urology cancer center. On the day of outpatient visits, we met face-to-face with patients who indicated an interest in participating in the study. We provided potential participants with the opportune time to ask questions. Then, we fully informed the potential participants about all facets of the study and invited them to voluntarily sign the informed consent form. Finally, we registered the participants after screening for the inclusion and exclusion criteria (Figure 1). The allocation process for this study was conducted jointly with a research assistant to ensure that all enrollments were transparent.
Using \textit{F} test in the G\textsuperscript{*}Power program version 3.1 (Heinrich Heine University) [40], we calculated the sample size to identify the variance difference of repeated data between the 2 groups. The minimum number of participants was calculated to be 41 based on an effect size of 0.51, which was derived from the study by Bourke et al [41], with a significance level of .05, a power of 0.08, and 3 repeat measures. Considering a projected dropout rate of 15% [41], the total sample size was set at 48. Therefore, 24 participants were registered in each group. We assigned the participants randomly into experimental and waitlist control groups using a pregenerated sequence in SAS 9.4 (SAS Institute) based on the $2 \times 2$ permuted randomized block method.

\textbf{Research Design and Setting}

This was a single-blind, single-center randomized waitlist controlled trial. It used a pre- and posttest design and a $2 \times 2$ permuted randomization block method. We prepared an allocation table for each group before participants were enrolled. Each participant was assigned to an experimental or a waitlist control group in the order of enrollment using a $2 \times 2$ randomization table generated in the SAS program (SAS Institute). The protocol for this study was registered with the Clinical Research Information Service (registration no. KCT0006560). We recruited the participants from a tertiary general hospital located in Seoul, South Korea.

\textbf{Intervention}

To improve the positive health behaviors of patients with PC receiving ADT, a nurse-led mobile-based health coaching program based on the analyze, design, develop, implement, and evaluate model and the information-motivation-behavioral (IMB) skills model was developed. The analyze, design, develop, implement, and evaluate model, known as an instructional design model, is a representative framework used by teaching and learning methods. The IMB model has been widely used as a theoretical basis for developing interventions that aim to encourage patients with chronic diseases to maintain positive health behaviors. According to the IMB model, acquiring sufficient information, enhancing motivation, acquiring the skills required to perform a behavior, and improving self-efficacy are the factors that lead to long-term behavioral changes that improve subjective and objective health outcomes [42]. The nurse-led mobile-based health coaching program used an individualized approach to ensure that the IMB
skills addressed in the program were best suited to each patient to improve their specific lifestyle patterns. In this program, self-management information included personalized recommendations for diet and exercise and strategies to manage both individual ADT-induced side effects and common side effects from other prescribed drugs. Motivation strategies included health contracting, individual goal setting, encouragement and continuous counseling, and rewarding. In addition, we encouraged the patients’ main caregivers to become involved in their patients’ behavioral self-management. Behavioral self-management skills included sharing vicarious experiences (eg, sharing success stories), self-monitoring through exercises, and maintaining a nutrition diary.

The program consisted of a 4-week intensive program conducted via one-on-one Zoom (Zoom Video Communication, Inc) meetings, followed by an 8-week maintenance program conducted via individual contact through telephone calls and KakaoTalk (Kakao Corporation), which is a popular Korean SMS text messaging app (Figure 2). Participants were provided with relevant information and a to-do list regarding exercise and diet once a week for 4 weeks in an intensive program. We provided an educational package that included a booklet containing PowerPoint (Microsoft Corporation) slides, exercise and nutritional diaries, exercise video files, TheraBand resistance bands (THERABAND), and a pedometer. We focused on overcoming barriers to healthy diet and exercise. For example, we recommended alternative exercises for participants with knee pain, which included a higher proportion of movements performed in a sitting position. Furthermore, the participants received feedback to address problems at the end of every web-based meeting during the intensive program and via SMS text messages and telephone calls during the maintenance program. When participants felt uncertain about an exercise motion, they recorded a video of themselves performing the exercise and sent it to us. We then provided feedback describing the correct motion, using the video as a reference. Regarding diet, we focused on changing unhealthy eating habits, controlling participants’ daily calorie intake, encouraging a low-fat and low-carbohydrate diet, and ensuring adequate protein intake. Participants were educated on the proportions of food from each food group they required per day according to their individual daily recommended calories and the diabetic exchange diet. Feedback and questions related to the diet program were addressed via KakaoTalk. We recommended implementing the following strategies for maintaining healthy behaviors and enhancing self-efficacy once a week or more during the maintenance period: vicarious experiences (eg, sharing success stories), emotional support, encouragement, keeping an exercise and nutrition diary, and goal reminders. To minimize the expected bias that can occur in an intervention study, we conducted the intervention according to the study protocol and used a checklist to ensure consistency between the experimental and waitlist control groups. The control group was placed on a waitlist and only received usual care during the intervention period. After completing data collection, we provided them with the same mobile-based program and educational materials as the experimental group (time point 3; T3). A group of experts evaluated this program to confirm its content and construct validity.

Figure 2. Educational material: (A) education on information and management of metabolic syndrome via the internet using Zoom (Zoom Video Communications, Inc), (B) personalized diet coaching, and (C) personalized exercise coaching: providing feedback describing the correct motion.
Data Collection

Data collection happened at 3 different time points: baseline (time point 1; T1), 6 weeks after the beginning of the intervention (time point 2; T2), and 12 weeks after the beginning of the intervention (T3). We collected T1 and T3 data on a face-to-face basis on the day the patient visited the urology cancer center. Variables that required physical contact, such as AC and grip strength, were measured together using a web-based survey at T1 and T3 but not at T2. At T2, we collected data collection on variables, including lifestyle score and HRQoL, on a non–face-to-face basis using a web-based survey. Clinical data that were obtained included treatment type, ADT exposure time, cancer stage, biochemical data, and Gleason score from electronic medical records (EMRs) at T1.

Outcomes

Lifestyle Score

Lifestyle score was the primary outcome of this study. Lifestyle changes were evaluated using the lifestyle evaluation tool by Kang [43] to assess the health behaviors of patients with MetS. A higher score indicates better self-management behaviors. This assessment consists of 36 items across 6 domains: physical activity and weight control, dietary habits, drinking and smoking, sleep and rest, stress, and drug and health management [43].

We recorded responses on a 4-point scale with the options not at all, sometimes, often, and always. Total scores could range from 36 to 144 points, with a higher score indicating a healthier lifestyle. The Cronbach α of the original study was .92 (physical activity and weight control=.87, dietary habits=.87, drinking and smoking=.87, sleep and rest=.86, stress=.74, and drug and health management=.70). In this study, Cronbach α was .82 (physical activity and weight control=.90, dietary habits=.86, drinking and smoking=.41, sleep and rest=.67, stress=.63, and drug and health management=.54).

MetS Components

The components of MetS consist of FBS, AC, blood pressure (both systolic and diastolic), fasting triglyceride level, and fasting HDL cholesterol. We instructed the participants not to have a meal and not to take antihypertensive drugs on the test day. Then, at the urology cancer center, we first measured AC using a tape at the umbilicus between the highest point of the iliac crest and the lower edge of the 12th rib with an error range of 0.5 cm while the participants fasted. Second, we measured the blood pressure twice on the participant’s nondominant arm using a TM-2657P device (A&D Company Limited) after they had rested for at least 10 minutes. The average systolic blood pressure (SBP) and diastolic blood pressure (DBP) were calculated.

Biochemical Data

Biochemical data included the levels of the following MetS components: fasting triglyceride, fasting HDL cholesterol, and FBS. These variables were measured using blood samples collected while participants had fasted for >6 hours, and the data were obtained from EMRs.

Body Composition

Body composition data were obtained using a body composition analyzer (InBody H20B [Biospace]). We instructed the participants to stand upright and hold the handle attached to the measurement device, which put them into contact with 8 electrodes (2 each on both hands and both feet). Body composition measurements included height (cm), body weight (kg), body fat mass (kg), body fat percentage (%), skeletal muscle mass (kg), and BMI (kg/m²).

HRQoL Measurement Tool

We measured HRQoL using the Korean version of the 26-item Expanded Prostate Cancer Index Composite (EPIC-26). EPIC-26 is a short-form version of the original expanded PC index composite (EPIC) tool, which contains 50 items. The EPIC tool was developed to understand treatment-related symptoms with a higher degree of sensitivity than previous diagnostic tools and the impact of PC treatment on patients’ lives [44]. Higher scores indicated a better HRQoL, with possible scores ranging from 0 to 100. EPIC-26 consists of 5 symptom domains: urinary incontinence, urinary irritation and obstruction, sexual, bowel, and hormonal. There is no Korean version of EPIC-26, but there is a Korean version of the original 50-item EPIC tool. Therefore, the 26 items from EPIC-26 were extracted from the Korean version of the original 50-item EPIC tool [45], and the survey was conducted using this tool. Permission to use both EPIC-26 and the original 50-item EPIC was granted by the original author.

The Cronbach α of EPIC-26 ranged from .70 to .90 in all domains except for the hormonal domain (Cronbach α=.62). The Cronbach α of the Korean version of EPIC was .83 [45]. The Cronbach α of the tool used in this study was .63 (urinary incontinence=.88, urinary irritation and obstruction=.64, sexual=.84, bowel=.13, and vitality or hormonal=.46).

Clinical Data

Disease-related patient information, treatment type, ADT exposure time, cancer stage, and Gleason score were obtained from EMRs.

Data Analysis

Demographic data, disease-related characteristics, and main outcome variables were analyzed using mean, SD, frequency, and percentage. An independent 1-tailed t test and a chi-square test were performed to identify differences between the groups. The equality of variance was assessed before using the pooled variance estimator for the t test. Fisher exact test was performed as appropriate. According to the International Conference on Harmonization E9 guideline [46], which provides guidance on statistical principles for clinical trials, a modified intention-to-treat analysis was conducted. As 2 participants withdrew from the study for personal reasons before initiating the intervention, there was a lack of data that could evaluate the effect of the main outcomes. We judged that there would be no difference in the intention-to-treat analysis owing to the low dropout rate (2/48, 4%) and high compliance rate (22/22, 81%). Hypothesis testing was conducted using a 1-tailed test and the PROC MIXED procedure in SAS (SAS Institute). SE estimates
Results

Overview
The data for this study were collected across 7 months, from March 24 to September 15, 2021. We took approximately 30 minutes per participant to complete data collection. A total of 48 participants were recruited. Two participants in the experimental group declined to participate in the program before starting the intervention, citing personal reasons. A total of 46 participants were finally included in the analysis, with an attrition rate of 4% (2/48).

Results of General and Disease-Related Characteristics
The general and disease-related characteristics are presented in Table 1. The mean age of the participants was 68.83 (SD 7.09) years. Most participants (41/46, 89%) lived with their spouses or families. Approximately 46% (21/46) of the participants were unemployed, and 35% (16/46) had jobs requiring relatively less physical activity, such as office workers, taxi drivers, and service workers. Most participants (27/46, 59%) were exsmokers, and the mean smoking duration was 16.13 (SD 21.38) pack-years.

were obtained as a result of the PROC MIXED procedure using the empirical option to adjust for skewed data from potentially different covariance structures. This method is based on the sandwich estimation approach [47]. It improves variance and covariance with robust and consistent estimates, irrespective of the covariance structure in the actual data. As a follow-up analysis, we determined statistically significant time points within the groups by calculating the difference in the least square means from the baseline at each time point.
Table 1. Homogeneity tests in general characteristics, disease-related characteristics, and main outcome variables between groups (N=46).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total</th>
<th>Experimental group (n=22)</th>
<th>Control group (n=22)</th>
<th>$\chi^2$ (df)</th>
<th>t (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td>68.83 (7.09)</td>
<td>67.59 (6.59)</td>
<td>69.96 (7.48)</td>
<td>N/Aa</td>
<td>−1.13 (44)</td>
<td>.26</td>
</tr>
<tr>
<td>Monthly income (US $), mean (SD)</td>
<td>3679.77 (4605.72)</td>
<td>4813.0 (6188.7)</td>
<td>2640.9 (2035.5)</td>
<td>N/A</td>
<td>1.57 (44)</td>
<td>.13</td>
</tr>
<tr>
<td><strong>Religion, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>21 (46)</td>
<td>10 (22)</td>
<td>11 (24)</td>
<td>0.0 (1)</td>
<td>N/A</td>
<td>.99</td>
</tr>
<tr>
<td>Yes</td>
<td>25 (54)</td>
<td>12 (26)</td>
<td>13 (28)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.40</td>
</tr>
<tr>
<td>Less than or equal to middle school</td>
<td>27 (59)</td>
<td>11 (24)</td>
<td>16 (35)</td>
<td>0.7 (1)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Greater than or equal to college</td>
<td>19 (41)</td>
<td>11 (24)</td>
<td>8 (17)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Living, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.77b</td>
</tr>
<tr>
<td>Alone</td>
<td>5 (11)</td>
<td>3 (7)</td>
<td>2 (4)</td>
<td>0.5 (2)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>With spouse only</td>
<td>21 (46)</td>
<td>9 (20)</td>
<td>12 (26)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>With their family</td>
<td>20 (43)</td>
<td>10 (22)</td>
<td>10 (22)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Job-related physical activities, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.93b</td>
</tr>
<tr>
<td>Unemployed</td>
<td>21 (46)</td>
<td>10 (22)</td>
<td>11 (24)</td>
<td>0.3 (2)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Less active</td>
<td>16 (35)</td>
<td>7 (15)</td>
<td>9 (35)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Highly active</td>
<td>9 (20)</td>
<td>5 (11)</td>
<td>4 (9)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Smoking history</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 pack-year, mean (SD)</td>
<td>16.13 (21.38)</td>
<td>15.78 (21.58)</td>
<td>16.44 (21.65)</td>
<td>N/A</td>
<td>−0.1 (44)</td>
<td>.92</td>
</tr>
<tr>
<td>Nonsmoker, n (%)</td>
<td>16 (35)</td>
<td>6 (13)</td>
<td>10 (22)</td>
<td>1.6 (2)</td>
<td>N/A</td>
<td>.45b</td>
</tr>
<tr>
<td>Exsmoker, n (%)</td>
<td>27 (59)</td>
<td>15 (33)</td>
<td>12 (26)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Current smoker, n (%)</td>
<td>3 (7)</td>
<td>1 (2)</td>
<td>2 (4)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Disease characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of comorbidities, mean (SD)</td>
<td>1.61 (1.04)</td>
<td>1.64 (1.05)</td>
<td>1.58 (2.03)</td>
<td>N/A</td>
<td>0.17 (44)</td>
<td>.87</td>
</tr>
<tr>
<td>0, n (%)</td>
<td>6 (13)</td>
<td>3 (7)</td>
<td>3 (7)</td>
<td>0.8 (3)</td>
<td>N/A</td>
<td>.87b</td>
</tr>
<tr>
<td>1, n (%)</td>
<td>17 (37)</td>
<td>7 (15)</td>
<td>10 (22)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>2, n (%)</td>
<td>14 (30)</td>
<td>8 (17)</td>
<td>6 (13)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>≥3, n (%)</td>
<td>9 (20)</td>
<td>4 (9)</td>
<td>5 (11)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Treatment type, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Operation</td>
<td>16 (35)</td>
<td>5 (11)</td>
<td>11 (24)</td>
<td>2.7 (1)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Operation+radiation</td>
<td>30 (65)</td>
<td>17 (37)</td>
<td>13 (28)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>ADT</strong> type, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antiandrogen</td>
<td>30 (65)</td>
<td>14 (30)</td>
<td>16 (35)</td>
<td>0.0 (1)</td>
<td>N/A</td>
<td>.99</td>
</tr>
<tr>
<td>Antiandrogen+LHRHd</td>
<td>16 (35)</td>
<td>8 (17)</td>
<td>8 (17)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ADT duration (month), mean (SD)</td>
<td>40.63 (24.71)</td>
<td>40.64 (22.61)</td>
<td>40.63 (26.98)</td>
<td>N/A</td>
<td>0.0 (44)</td>
<td>.99</td>
</tr>
<tr>
<td>PSA, mean (SD)</td>
<td>0.07 (0.19)</td>
<td>0.09 (0.27)</td>
<td>0.05 (0.08)</td>
<td>N/A</td>
<td>0.59 (44)</td>
<td>.56</td>
</tr>
<tr>
<td>Gleason score, mean (SD)</td>
<td>7.80 (0.98)</td>
<td>8.00 (1.07)</td>
<td>7.63 (0.88)</td>
<td>N/A</td>
<td>1.31 (44)</td>
<td>.20</td>
</tr>
</tbody>
</table>

Main outcomes

https://mhealth.jmir.org/2024/1/e47102  JMIR Mhealth Uhealth 2024 | vol. 12 | e47102 | p.260  (page number not for citation purposes)
## Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total (n=97)</th>
<th>Experimental group (n=22)</th>
<th>Control group (n=22)</th>
<th>$\chi^2$ (df)</th>
<th>t (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em><em>Lifestyle score</em> mean (SD)</em>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>99.07 (13.56)</td>
<td>100.50 (12.53)</td>
<td>97.71 (14.57)</td>
<td>N/A</td>
<td>0.70 (44)</td>
<td>.48</td>
</tr>
<tr>
<td>Exercise and weight loss</td>
<td>19.67 (5.88)</td>
<td>20.23 (5.45)</td>
<td>19.17 (6.32)</td>
<td>N/A</td>
<td>0.61 (44)</td>
<td>.55</td>
</tr>
<tr>
<td>Diet</td>
<td>42.17 (6.72)</td>
<td>42.77 (6.05)</td>
<td>41.63 (7.38)</td>
<td>N/A</td>
<td>0.57 (44)</td>
<td>.57</td>
</tr>
<tr>
<td>Alcohol and smoking</td>
<td>9.54 (3.40)</td>
<td>10.09 (3.25)</td>
<td>9.04 (3.53)</td>
<td>N/A</td>
<td>1.05 (44)</td>
<td>.30</td>
</tr>
<tr>
<td>Stress management</td>
<td>9.82 (1.79)</td>
<td>9.82 (1.79)</td>
<td>9.33 (2.41)</td>
<td>N/A</td>
<td>0.77 (44)</td>
<td>.45</td>
</tr>
<tr>
<td>Sleep and rest</td>
<td>6.45 (1.60)</td>
<td>6.45 (1.60)</td>
<td>6.71 (1.76)</td>
<td>N/A</td>
<td>–0.51 (44)</td>
<td>.61</td>
</tr>
<tr>
<td>Medication adherence and physical examination</td>
<td>11.52 (2.04)</td>
<td>11.18 (1.84)</td>
<td>11.83 (2.20)</td>
<td>N/A</td>
<td>–1.08 (44)</td>
<td>.28</td>
</tr>
<tr>
<td><em><em>MetS</em> n (%)</em>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>21 (46)</td>
<td>10 (22)</td>
<td>11 (24)</td>
<td>0.0 (1)</td>
<td>N/A</td>
<td>.98</td>
</tr>
<tr>
<td>Yes</td>
<td>25 (54)</td>
<td>12 (26)</td>
<td>13 (28)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>MetS component, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBPb (mm/Hg)</td>
<td>136.42 (13.01)</td>
<td>139.7 (11.44)</td>
<td>133.4 (13.83)</td>
<td>N/A</td>
<td>1.69 (44)</td>
<td>.10</td>
</tr>
<tr>
<td>DBPb (mm/Hg)</td>
<td>83.51 (7.46)</td>
<td>85.18 (6.11)</td>
<td>81.98 (8.35)</td>
<td>N/A</td>
<td>1.47 (44)</td>
<td>.15</td>
</tr>
<tr>
<td>ACj (cm)</td>
<td>95.73 (5.97)</td>
<td>94.30 (5.29)</td>
<td>97.05 (6.36)</td>
<td>N/A</td>
<td>–1.59 (44)</td>
<td>.12</td>
</tr>
<tr>
<td>FBSk (mg/dl)</td>
<td>111.17 (22.22)</td>
<td>112.6 (27.64)</td>
<td>109.8 (16.27)</td>
<td>N/A</td>
<td>0.41 (44)</td>
<td>.68</td>
</tr>
<tr>
<td>HDLj (mg/dl)</td>
<td>52.89 (13.61)</td>
<td>55.50 (13.32)</td>
<td>50.50 (13.71)</td>
<td>N/A</td>
<td>1.25 (44)</td>
<td>.22</td>
</tr>
<tr>
<td>Triglyceride (mg/dl)</td>
<td>114.46 (49.05)</td>
<td>116.5 (44.45)</td>
<td>112.5 (53.82)</td>
<td>N/A</td>
<td>0.27 (44)</td>
<td>.79</td>
</tr>
<tr>
<td><strong>Body composition, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>73.12 (5.99)</td>
<td>72.53 (4.54)</td>
<td>73.66 (7.13)</td>
<td>N/A</td>
<td>–0.64 (44)</td>
<td>.52</td>
</tr>
<tr>
<td>Body fat mass (kg)</td>
<td>22.18 (4.59)</td>
<td>21.75 (3.51)</td>
<td>22.57 (5.44)</td>
<td>N/A</td>
<td>–0.61 (44)</td>
<td>.54</td>
</tr>
<tr>
<td>Body fat percent (%)</td>
<td>30.38 (5.31)</td>
<td>30.50 (4.74)</td>
<td>30.28 (5.88)</td>
<td>N/A</td>
<td>0.14 (44)</td>
<td>.89</td>
</tr>
<tr>
<td>Skeletal muscle mass (kg)</td>
<td>28.22 (3.08)</td>
<td>28.03 (2.96)</td>
<td>28.39 (3.23)</td>
<td>N/A</td>
<td>–0.39 (44)</td>
<td>.70</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td>25.93 (2.04)</td>
<td>25.65 (1.70)</td>
<td>26.19 (2.32)</td>
<td>N/A</td>
<td>–0.89 (44)</td>
<td>.38</td>
</tr>
<tr>
<td><em><em>HRQoL</em> domains, mean (SD)</em>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinary incontinence</td>
<td>67.61 (25.99)</td>
<td>69.17 (26.56)</td>
<td>66.17 (25.93)</td>
<td>N/A</td>
<td>0.39 (44)</td>
<td>.70</td>
</tr>
<tr>
<td>Urinary irritation and obstruction</td>
<td>88.72 (10.59)</td>
<td>87.78 (10.11)</td>
<td>89.58 (11.16)</td>
<td>N/A</td>
<td>–0.57 (44)</td>
<td>.57</td>
</tr>
<tr>
<td>Bowel problem</td>
<td>95.11 (9.31)</td>
<td>95.83 (8.33)</td>
<td>94.44 (10.26)</td>
<td>N/A</td>
<td>0.50 (44)</td>
<td>.62</td>
</tr>
<tr>
<td>Sexual problem</td>
<td>19.28 (19.54)</td>
<td>21.89 (19.18)</td>
<td>16.89 (19.97)</td>
<td>N/A</td>
<td>0.86 (44)</td>
<td>.39</td>
</tr>
</tbody>
</table>
The average number of comorbidities was 1.61 (SD 1.04), and the most common comorbid diseases were cardiovascular diseases such as hypertension or dyslipidemia (31/46, 67%). There was no difference in comorbidities between the experimental and waitlist control groups (P=.87). Of the 65 participants, 30 (65%) received antiandrogen monotherapy, and the rest received antiandrogen therapy with luteinizing hormone–releasing hormone. The average duration of ADT was 40.63 (SD 24.71) months, the mean prostate-specific antigen level was 0.07 (SD 0.19) ng/ml, and the mean Gleason score, which determines the prognosis and pathological status of cancer, was 7.80 (SD 0.98). The experimental and waitlist control groups showed no statistical differences between general and disease-related characteristics. The mean healthy lifestyle score was 99.07 (SD 13.56). The MetS components with mean values in an abnormal range were SBP (mean 136.42, SD 13.01 mm Hg), AC (mean 95.73, SD 5.97 cm), and FBS (mean 111.17, SD 22.22 mg/dL). There were no statistical differences in MetS components between the groups. The mean body weight and BMI were approximately 73 (SD 5.99) kg and 26 (SD 2.04) kg/m^2, respectively. The participants responded that they had problems in the urinary incontinence (mean 67.61, SD 25.99) and sexual (mean 19.28, SD 19.54) domains of HRQoL.

**Results of Primary Outcome Variables Between Groups Over Time**

Table 2 and Figure 3 show the results of the lifestyle score variables between the groups over time. The study found that there were no group differences in lifestyle scores at baseline (Table 1). However, over time, the experimental group’s lifestyle scores consistently increased (T1=100.55, T2=125.82, and T3=130.27), whereas the waitlist control group’s lifestyle scores showed no consistent increase (T1=97.71, T2=95.92, and T3=98.21). Over time, the lifestyle scores of the experimental group significantly increased (β=29.23; P<.001) compared with the waitlist control group. The experimental group developed a healthier lifestyle with time, and remarkable changes were observed during the intervention period.
Table 2. Group comparison in lifestyle score, MetS\(^\text{a}\) components, body composition, and HRQoL\(^\text{b}\) domain parameters (N=46).

<table>
<thead>
<tr>
<th>Outcome and group</th>
<th>T1(^\text{c}), mean (SD)</th>
<th>T2(^\text{d}), mean (SD)</th>
<th>T3(^\text{e}), mean (SD)</th>
<th>Difference of changes between groups over time(^\text{f})</th>
<th>Estimate (SE; 95% CI)</th>
<th>F test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lifestyle score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>100.55 (12.53)</td>
<td>125.82 (11.73)</td>
<td>130.27 (10.15)</td>
<td>29.23 (3.50; 22.26 to 36.19)</td>
<td>38.49 (88)</td>
<td>≤.001</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>97.71 (14.57)</td>
<td>95.92 (18.64)</td>
<td>98.21 (15.07)</td>
<td>29.23 (3.50; 22.26 to 36.19)</td>
<td>38.49 (88)</td>
<td>≤.001</td>
<td></td>
</tr>
<tr>
<td><strong>MetS components</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBP(^\text{g})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>139.73 (11.44)</td>
<td>N/A(^\text{h})</td>
<td>126.25 (14.08)</td>
<td>−5.64 (3.72; −13.12 to 1.85)</td>
<td>2.30 (44)</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>133.38 (13.83)</td>
<td>N/A(^\text{h})</td>
<td>125.54 (12.58)</td>
<td>−5.64 (3.72; −13.12 to 1.85)</td>
<td>2.30 (44)</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>DBP(^\text{i})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>85.18 (6.11)</td>
<td>N/A(^\text{h})</td>
<td>79.36 (9.35)</td>
<td>−0.94 (2.21; −5.40 to 3.52)</td>
<td>0.18 (44)</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>81.98 (8.35)</td>
<td>N/A(^\text{h})</td>
<td>77.10 (7.36)</td>
<td>−0.94 (2.21; −5.40 to 3.52)</td>
<td>0.18 (44)</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>FBS(^\text{j})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>112.64 (27.64)</td>
<td>N/A(^\text{h})</td>
<td>102.59 (12.42)</td>
<td>−12.0 (6.02; −24.14 to 0.13)</td>
<td>3.98 (44)</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>109.83 (16.27)</td>
<td>N/A(^\text{h})</td>
<td>111.79 (16.88)</td>
<td>−12.0 (6.02; −24.14 to 0.13)</td>
<td>3.98 (44)</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>HDL(^k) cholesterol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>55.50 (13.33)</td>
<td>N/A(^\text{h})</td>
<td>55.32 (12.46)</td>
<td>−0.27 (1.48; −3.25 to 2.72)</td>
<td>0.03 (44)</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>50.50 (13.71)</td>
<td>N/A(^\text{h})</td>
<td>50.58 (12.34)</td>
<td>−0.27 (1.48; −3.25 to 2.72)</td>
<td>0.03 (44)</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>Triglyceride</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>116.55 (44.45)</td>
<td>N/A(^\text{h})</td>
<td>107.73 (43.19)</td>
<td>−11.65 (12.04; −35.91 to 12.61)</td>
<td>0.94 (44)</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>112.54 (53.82)</td>
<td>N/A(^\text{h})</td>
<td>115.38 (62.71)</td>
<td>−11.65 (12.04; −35.91 to 12.61)</td>
<td>0.94 (44)</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>AC(^l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>94.30 (5.29)</td>
<td>N/A(^\text{h})</td>
<td>91.02 (3.94)</td>
<td>−2.49 (1.23; −4.98 to −0.01)</td>
<td>4.09 (44)</td>
<td>.049</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>97.05 (6.36)</td>
<td>N/A(^\text{h})</td>
<td>96.27 (6.83)</td>
<td>−2.49 (1.23; −4.98 to −0.01)</td>
<td>4.09 (44)</td>
<td>.049</td>
<td></td>
</tr>
<tr>
<td><strong>Body composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Body weight (kg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>72.53 (4.54)</td>
<td>N/A(^\text{h})</td>
<td>70.04 (4.53)</td>
<td>−1.52 (0.46; −2.45 to −0.58)</td>
<td>10.71 (44)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>73.66 (7.13)</td>
<td>N/A(^\text{h})</td>
<td>72.68 (7.04)</td>
<td>−1.52 (0.46; −2.45 to −0.58)</td>
<td>10.71 (44)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td><strong>BMI (kg/m(^2))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>25.65 (1.70)</td>
<td>N/A(^\text{h})</td>
<td>24.76 (1.49)</td>
<td>−0.55 (0.17; −0.88 to −0.21)</td>
<td>10.54 (44)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>26.19 (2.32)</td>
<td>N/A(^\text{h})</td>
<td>25.84 (2.27)</td>
<td>−0.55 (0.17; −0.88 to −0.21)</td>
<td>10.54 (44)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td><strong>Skeletal muscle mass (kg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>28.03 (2.96)</td>
<td>N/A(^\text{h})</td>
<td>27.84 (4.19)</td>
<td>0.53 (0.83; −1.15 to 2.20)</td>
<td>0.40 (44)</td>
<td>.53</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>28.39 (3.23)</td>
<td>N/A(^\text{h})</td>
<td>27.67 (2.84)</td>
<td>0.53 (0.83; −1.15 to 2.20)</td>
<td>0.40 (44)</td>
<td>.53</td>
<td></td>
</tr>
<tr>
<td><strong>Fat mass (kg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>21.75 (3.51)</td>
<td>N/A(^\text{h})</td>
<td>19.76 (5.01)</td>
<td>−1.93 (1.35; −4.65 to 0.78)</td>
<td>2.06 (44)</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>22.57 (5.44)</td>
<td>N/A(^\text{h})</td>
<td>22.52 (4.66)</td>
<td>−1.93 (1.35; −4.65 to 0.78)</td>
<td>2.06 (44)</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td><strong>Fat percentage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>30.50 (4.74)</td>
<td>N/A(^\text{h})</td>
<td>28.32 (7.27)</td>
<td>−2.75 (1.90; −6.58 to 1.09)</td>
<td>2.08 (44)</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>30.28 (5.88)</td>
<td>N/A(^\text{h})</td>
<td>30.84 (4.61)</td>
<td>−2.75 (1.90; −6.58 to 1.09)</td>
<td>2.08 (44)</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td><strong>HRQoL domains</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

\(^a\) MetS components

\(^b\) HRQoL domain parameters

\(^c\) T1 = baseline

\(^d\) T2 = week 2

\(^e\) T3 = week 8

\(^f\) Difference of changes between groups over time

\(^g\) SBP = Systolic Blood Pressure

\(^h\) N/A = Not Available

\(^i\) DBP = Diastolic Blood Pressure

\(^j\) FBS = Fasting Blood Sugar

\(^k\) HDL = High-Density Lipoprotein

\(^l\) AC = A1c
<table>
<thead>
<tr>
<th>Outcome and group</th>
<th>T1&lt;sup&gt;c&lt;/sup&gt;, mean (SD)</th>
<th>T2&lt;sup&gt;d&lt;/sup&gt;, mean (SD)</th>
<th>T3&lt;sup&gt;e&lt;/sup&gt;, mean (SD)</th>
<th>Difference of changes between groups over time&lt;sup&gt;f&lt;/sup&gt;</th>
<th>F test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Urinary incontinence</strong></td>
<td></td>
<td></td>
<td></td>
<td>Estimate (SE; 95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>69.17 (26.56)</td>
<td>81.19 (18.87)</td>
<td>83.38 (18.67)</td>
<td>7.70 (6.55; −5.31 to 20.72)</td>
<td>0.70 (88)</td>
<td>.50</td>
</tr>
<tr>
<td>Control</td>
<td>66.18 (25.93)</td>
<td>72.29 (23.00)</td>
<td>72.68 (29.53)</td>
<td>7.70 (6.55; −5.31 to 20.72)</td>
<td>0.70 (88)</td>
<td>.50</td>
</tr>
<tr>
<td><strong>Urinary irritation and obstruction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>87.78 (10.11)</td>
<td>94.60 (7.54)</td>
<td>97.73 (4.93)</td>
<td>14.63 (4.05; 6.57 to 22.69)</td>
<td>7.01 (88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Control</td>
<td>89.58 (11.16)</td>
<td>89.32 (11.43)</td>
<td>84.90 (18.14)</td>
<td>14.63 (4.05; 6.57 to 22.69)</td>
<td>7.01 (88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Bowel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>95.83 (8.33)</td>
<td>95.64 (7.21)</td>
<td>98.48 (4.37)</td>
<td>2.30 (2.98; −3.61 to 8.22)</td>
<td>0.34 (88)</td>
<td>.71</td>
</tr>
<tr>
<td>Control</td>
<td>94.44 (10.26)</td>
<td>91.67 (14.33)</td>
<td>94.79 (10.15)</td>
<td>2.30 (2.98; −3.61 to 8.22)</td>
<td>0.34 (88)</td>
<td>.71</td>
</tr>
<tr>
<td><strong>Sexual</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>21.89 (19.18)</td>
<td>18.86 (16.02)</td>
<td>25.49 (16.71)</td>
<td>1.52 (4.50; −7.43 to 10.47)</td>
<td>0.08 (88)</td>
<td>.92</td>
</tr>
<tr>
<td>Control</td>
<td>16.89 (19.97)</td>
<td>12.15 (12.18)</td>
<td>18.97 (15.03)</td>
<td>1.52 (4.50; −7.43 to 10.47)</td>
<td>0.08 (88)</td>
<td>.92</td>
</tr>
<tr>
<td><strong>Hormonal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>85.00 (11.65)</td>
<td>89.32 (12.94)</td>
<td>92.73 (7.03)</td>
<td>6.27 (3.48; −0.64 to 13.18)</td>
<td>1.85 (88)</td>
<td>.16</td>
</tr>
<tr>
<td>Control</td>
<td>85.00 (15.88)</td>
<td>86.88 (12.58)</td>
<td>86.46 (13.87)</td>
<td>6.27 (3.48; −0.64 to 13.18)</td>
<td>1.85 (88)</td>
<td>.16</td>
</tr>
</tbody>
</table>

<sup>a</sup>MetS: metabolic syndrome.
<sup>b</sup>HRQoL: health-related quality of life.
<sup>c</sup>T1: time point 1.
<sup>d</sup>T2: time point 2.
<sup>e</sup>T3: time point 3.
<sup>f</sup>Reference: interaction between control group and T1.
<sup>g</sup>SBP: systolic blood pressure.
<sup>h</sup>N/A: not applicable.
<sup>i</sup>DBP: diastolic blood pressure.
<sup>j</sup>FBS: fasting blood sugar.
<sup>k</sup>HDL cholesterol: high-density lipoprotein cholesterol.
<sup>l</sup>AC: abdominal circumference.
**Figure 3.** Group comparison among the main outcome variables (Lifestyle score) over time. T1: time point 1; T2: time point 2; T3: time point 3.

### Results of Secondary Outcome Variables Between Groups Over Time

The difference in the prevalence of MetS before and after the intervention between the experimental group and the waitlist control group was not statistically significant ($\chi^2 = 1.1; P = .31$ at T3 [not presented in the tables]). Among the MetS components, the parameters for FBS ($\beta = -12.0; F_{44} = 3.98; P = .05$) and AC ($\beta = -2.49; F_{44} = 4.09; P = .049$) showed significant interactions between group and time (Table 2 and Figures 4 and 5). Regarding body composition, the mean body weight and BMI in the experimental group decreased significantly by 1.52 kg and 0.55 kg/m$^2$, respectively ($P < .001$) compared with the baseline values. Group, time, and group and time interactions were also statistically significant between these variables ($P < .001$). Over time, the mean body weight ($P < .001$) and BMI ($P < .001$) decreased more in the experimental group than in the waitlist control group (Table 2 and Figures 6 and 7). Regarding HRQoL domains (Table 2 and Figure 8), a more significant improvement was observed in the experimental group than in the waitlist control group for the urinary irritative and obstructive domain. The mean changes in the urinary irritative and obstructive domain of HRQoL were statistically significant at each time point from the baseline, whereas the mean changes in the control group for this domain were not statistically significant. Group and time interactions were not significant, except in the urinary irritative and obstructive domain of HRQoL in the experimental group ($\beta = 14.63; F_{8,8} = 7.01; P < .001$).
Figure 4. Group comparison among the main outcome variables (fasting blood sugar) over time. T1: time point 1; T3: time point 3.

Figure 5. Group comparison among the main outcome variables (abdominal circumference) over time. T1: time point 1; T3: time point 3.
**Figure 6.** Group comparison among the main outcome variables (body weight) over time. T1: time point 1; T3: time point 3.

**Figure 7.** Group comparison among the main outcome variables (BMI) over time. T1: time point 1; T3: time point 3.
Discussion

Principal Findings

This study was conducted to determine the effectiveness of a nurse-led mobile-based education program for patients with PC who are at risk of MetS. This study showed that a nurse-led mobile-based health coaching program promoted a healthy lifestyle in patients with PC receiving ADT, which ultimately improved the levels of some MetS components (eg, reduction in FBS and AC), body composition (eg, reduction in weight and BMI), and HRQoL (eg, the urinary irritative and obstructive domain).

In this study, with a 3-month intervention, the results indicated that each variable required a different amount of time to show changes. In a previous study, a period of at least 3 to 8 weeks was required to confirm significant changes in weight, BMI, and the levels of each MetS component following lifestyle intervention programs [48-50]. Specifically, AC, FBS, body weight, and BMI have been found to decrease significantly over short periods [49,51]. However, in studies with intervention periods of ≥6 months, significant results were confirmed for MetS components including HDL cholesterol, SBP, and DBP [52,53]. Even in weight control programs that included strict diet control and exercise intervention guidelines, MetS components, including HDL cholesterol, did not change over a period of 8 weeks and only showed a significant change in both men and women 10 months after the end of the intervention [54].

The study by Dawson et al [55] observed decreases in body weight and AC but not in HDL cholesterol level among the MetS components in patients with PC. Focht and colleagues [56] also reported decreased bodyweight in patients with PC. These studies included lifestyle interventions lasting 3 months for patients with PC who had received ADT. The mean duration of ADT was 14 months in the study by Focht et al [56] and 30 months in the study by Dawson et al [55]. Specifically, Focht et al [56] confirmed that the body composition of body weight and body fat decreased significantly at 3 months compared with the control group. Furthermore, Dawson et al [55] conducted a program that emphasized exercise and protein supplementation for 3 months in patients with PC receiving ADT. As a result, of the MetS components, the AC of participants decreased most significantly in the experimental group than in the waitlist control group.

Reductions in body weight, AC, and FBS levels were closely related. When beginning to lose weight, the body temporarily lowers its metabolic function to maintain homeostasis and first metabolizes glucose, which is a basic energy source. When stored glycogen is broken down, the insulin mechanism is activated for additional energy consumption [57]. Weight loss causes a decrease in FBS level along with the action of insulin, which reactivates FBS, stored in the form of excess fat in the liver or abdomen. This fat is continuously used to generate energy, consequently, the fat accumulated in the liver or abdomen is consumed, leading to a reduced volume [58]. In this study, although no significant decrease in body fat was observed, a decrease in AC was observed. Similarly, in previous
studies, decreases in body weight [56], AC [55], and FBS levels [49] were confirmed 3 months after the intervention.

In this study, there were no significant changes in body composition related to body fat and skeletal muscle mass or in MetS components related to blood pressure (SBP and DBP) and lipids (HDL cholesterol and triglyceride). Insulin resistance develops over a long period, which increases the risk of obesity, diabetes, and MetS. In addition, it affects the lipid ratio, leading to an increase in low-density lipoprotein cholesterol and a decrease in HDL cholesterol [58], and causes inflammatory changes, resulting in changes in blood pressure owing to an increase in the residual amount of sodium in the blood [59]. Moreover, in a study that examined prediabetic patients diagnosed with impaired fasting glucose levels over the course of a 10-year follow-up period, 37% of the patients developed diabetes [60]. This finding indicates that the disease mechanism does not change over a short period but rather progresses slowly. Therefore, a long-term follow-up study is required to more accurately confirm the effectiveness of a lifestyle intervention program for MetS.

Among the HRQoL domains, an effect was observed only in the urinary irritative and obstructive domain, which may be caused by the education on how to manage the side effects of ADT treatment in the third and fourth weeks of the intervention. Patients with PC who have undergone multiple surgeries or frequent radiation therapy complained of side effects [61], such as urinary irritation and frequency. During the intervention in this study, information about appropriate water intake and pelvic floor muscle exercises to relieve urinary irritative and obstructive symptoms such as urinary irritation and frequency were included in an educational brochure [57], and appropriate water intake and pelvic floor muscle exercise were recommended for the participants depending on the presence or absence of symptoms. Thus, participants with these symptoms may have experienced relief through this intervention. Of the remaining HRQoL domains, the urinary incontinence, bowel, sexual, and hormonal domains did not show statistically significant changes. HRQoL was assessed using a questionnaire on treatment side effects and symptoms during the preceding 4 weeks. Most participants in this study were receiving long-term treatment after surgery, with an average of 40 (SD 24.71) months of ADT treatment. The bowel domain contains questions about whether the participants had diarrhea and bloody stools, which are acute side effects of patients who have undergone PC resection surgery. This domain had a higher mean score than the other HRQoL domains, with an average of ≥95 beginning at the baseline; therefore, these questions might not be relevant for participants of this study who were not in the acute stage after surgery. In addition, most questions in the hormonal domain examined the side effects that appear toward the start of ADT such as breast tenderness, bloody stools, and weight loss. Therefore, the insignificant changes in this domain might be because the participants in this study had already been taking medication for these side effects. At 19 points, the sexual domain had the lowest mean score of all the HRQoL domains. Sexual function might be limited in recovery as a result of a short-term health coaching program [62,63]. Although a previous study found that exercise can improve sexual function [64], sexual function requires interaction with a partner [65]. To restore sexual function, it is necessary to combine both psychological intervention and drug treatment [66].

This study found lifestyle changes to be critical in reducing the risk of MetS and that improved exercise and nutritional regimens should be implemented consistently for at least 3 months. Furthermore, a patient-centered, individualized approach that considers the side effects of ADT is needed to increase adherence.

Limitations
This study had some limitations. First, the criteria for participant selection included patients with at least 1 MetS component. To more thoroughly examine the effectiveness of this program on MetS management, stricter inclusion criteria should have been applied, such as only including those who were being treated with luteinizing hormone–releasing hormone, which is a type of ADT that causes the occurrence of more MetS components MetS components; those with ≥3 MetS components; those with a sedentary lifestyle; or those with a specific time after radical prostatectomy. Owing to data unavailability, this study only collected the Gleason score but not the information on TNM classification as the patient’s pathological data. It is important to establish accurate patient pathological criteria; hence, future studies should incorporate both the TNM classification and the Gleason score with accurate criteria. Second, the 3-month intervention period planned in this study was limited to evaluating the continuation of self-management and improvement in MetS risk factors. It is necessary to extend the intervention period to at least 6 months to further evaluate the persistence of self-management. As older men with PC receiving ADT are at high risk of osteoporosis, measuring bone health–related indicators in addition to changes in body composition is highly recommended for future studies. Third, it is difficult to generalize the results of this study because the participants were recruited from a single hospital and not from a multicenter. Fourth, obtaining low Cronbach α values for some items in the lifestyle score may indicate a mismatch between the measurement and the participant’s lifestyle. This is because the lifestyle score was created for the general adult population; however, the participants in this study were older adults who had PC, had undergone surgery, and were taking hormone-suppressing medications.

Conclusions
A nurse-led mobile-based health coaching program was developed to promote a healthy lifestyle among patients with PC receiving ADT. The research findings confirmed that lifestyle changes through nurse-led education can improve the components of MetS, body composition, and ADT’s side effects. Therefore, by participating in this program, nurses were capable of creating changes in patients’ lifestyles and improving the self-management of patients who were beginning ADT for the first time. In addition, this study can provide a basis for the development of other mobile-based education programs and tools.
Acknowledgments

This work was supported by the Korea Medical Device Development Fund grant funded by the Korean Government (the Ministry of Science and Information and Communications Technology; the Ministry of Trade, Industry and Energy; the Ministry of Health and Welfare; and the Ministry of Food and Drug Safety; project number: 1711174508, KMDF_PR_20200901_0077). This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1A6A3A13060728). This work was supported by the NRF of Korea grant funded by the Korean Government (MSIT; Ministry of Science and Information Communications Technology number 2019R1F1A1062769 and number 2022R1A2C1092084). KL received a scholarship from Brain Korea 21 FOUR Project funded by National Research Foundation NRF of Korea, Yonsei University College of Nursing.

Authors' Contributions

KL contributed to data curation, formal analysis, investigation, methodology, project administration, software work, statistical analysis, visualization, original draft, writing, review and editing, funding acquisition. JP contributed to funding acquisition, investigation, methodology, project administration, visualization, review and editing, and supervision. All authors contributed to the conceptualization and have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1218 KB - mhealth_v12i1e47102_app1.pdf ]

References


Abbreviations

AC: abdominal circumference
ADT: androgen deprivation therapy
DBP: diastolic blood pressure
EMR: electronic medical record
EPIC: expanded prostate cancer index composite
EPIC-26: 26-item Expanded Prostate Cancer Index Composite
FBS: fasting blood sugar
HDL: high-density lipoprotein
HRQoL: health-related quality of life
IMB: information-motivation-behavioral
MetS: metabolic syndrome
PC: prostate cancer
SBP: systolic blood pressure
T1: time point 1
T2: time point 2
T3: time point 3

Edited by L Buis; submitted 22.03.23; peer-reviewed by Y Kim, SH Kim; comments to author 13.09.23; revised version received 29.10.23; accepted 20.12.23; published 01.02.24.

Please cite as:
Lee K, Park J, Oh EG, Lee J, Park C, Choi YD
Effectiveness of a Nurse-Led Mobile-Based Health Coaching Program for Patients With Prostate Cancer at High Risk of Metabolic Syndrome: Randomized Waitlist Controlled Trial
JMIR Mhealth Uhealth 2024;12:e47102
URL: https://mhealth.jmir.org/2024/1/e47102
doi:10.2196/47102
PMID:
Weight Loss Using an mHealth App Among Individuals With Obesity in Different Economic Regions of China: Cohort Study

Xinru Huang1,*; Yefei Shi2,*; Hongyun Yao1,*; Mingjie Li1; Zhijun Lei2, PhD; Jiayun Shi2; Bo Li2, PhD; Weixia Jian1, MD, PhD

*these authors contributed equally

Corresponding Author:
Weixia Jian, MD, PhD

Abstract

Background: With the increasing prevalence of obesity, weight loss has become a critical issue in China. Self-managed weight loss through a mobile health (mHealth) app may be a prospective method. However, its practicability in different economic regions of China is unknown.

Objective: This study aims to evaluate the effectiveness of self-managed weight loss through an mHealth app among individuals with obesity in different economic regions of China and to demonstrate the feasibility of online self-management for weight loss.

Methods: A total of 165,635 Chinese adults who signed up for the mHealth app were included to analyze the body composition characteristics of individuals from different economic regions by χ² analyses. Furthermore, 2 types of participants with obesity using mHealth monitoring, including 74,611 participants with a BMI ≥24.0 kg/m² and 22,903 participants with a normal BMI but an excessive percentage of body fat (PBF), were followed for 6 months to explore the weight loss and fat loss effects in different economic regions of China and to find independent predictors associated with weight loss success by 2-tailed Student t test and multivariable logistic regression analysis.

Results: There were 32,129 users from low-income regions and 133,506 users from high-income regions. The proportion of users with obesity in low-income regions was higher than in high-income regions, both based on BMI (15,378/32,129, 47.9% vs 59,233/133,506, 44.4%; P < .001) and PBF classification (19,146/32,129, 59.6% vs 72,033/133,506, 54%; P < .001). Follow-up analyses showed that the weight loss effect among participants with overweight or obesity in low-income regions was greater than in high-income regions (mean –4.93, SD 6.41 vs mean –4.71, SD 6.14 kg; P < .001), while there was no significant difference in fat loss (mean –2.06%, SD 3.14% vs mean –2.04%, SD 3.19%; P = .54). In the population with normal-weight obesity, the weight loss (mean –2.42, SD 4.07 vs mean –2.23, SD 4.21 kg; P = .004) and fat loss effects (mean –1.43%, SD 2.73% vs mean –1.27%, SD 2.63%; P < .001) were stronger in high-income regions than in low-income regions. In addition, multivariable logistic regression analyses showed that age, baseline PBF, skeletal muscle rate, and measurement frequency were related to weight loss, whereas gender and baseline body metabolic rate only showed a correlation with weight loss in the population in high-income regions.

Conclusions: This study found a high proportion of mHealth app users with obesity in low-income regions. Individuals with overweight and obesity in different economic regions of China experienced significant weight loss and fat loss using an mHealth app. Moreover, individuals in high-income regions paid more attention to body fat and had better fat reduction effects. Therefore, promoting self-monitoring of weight and PBF through an mHealth app could be an important intervention that could be implemented across all regions of China.

doi:10.2196/48675

KEYWORDS
weight loss; obesity; normal-weight obesity; economic regions; mHealth app; mobile health; China; mHealth

Introduction

Obesity is a global health crisis that has reached pandemic proportions in many countries, including China [1, 2]. The World Health Organization estimated that over 2.1 billion people worldwide are overweight or obese, and this number is expected to continue rising in the coming years [3, 4]. Obesity is a significant risk factor for several chronic health conditions, including heart disease, stroke, diabetes, and certain types of cancer, and has been linked to decreased quality of life and premature death [5-7]. In China, the prevalence of overweight and obesity has substantially increased in recent years. This has
been attributed to several factors, including increased urbanization, changes in dietary habits, and decreased physical activity levels [8,9]. The rising burden of obesity in China has led to growing concerns about the health and economic consequences of this trend and has prompted the need for effective weight loss.

Standard behavioral treatment for obesity included dietary and physical activity counseling and self-monitoring of body weight, activity, and diet [10,11]. Consistent self-weighing over time promoted the awareness of behaviors, environments, or situations that might lead to desired or undesired changes in weight. Researchers have established a correlation between self-weighing and successful weight loss, with studies showing that self-weighing significantly improved weight loss outcomes during the first 6 months of a weight loss intervention [12,13].

A new online weight management system, the smart body fat scale, calculates the percentage of body fat (PBF), records and synchronizes the data to mobile health (mHealth) apps, and offers more advantages than a traditional scale [14,15]. In addition to measuring PBF, smart body fat scales usually offer additional features such as the ability to track weight and body fat over time and the ability to measure other health-related metrics such as muscle mass [16]. The use of mHealth with smart body fat scales has become increasingly popular in recent years, as individuals seek to monitor their health and fitness more closely and make more informed decisions about their diets and physical activity patterns.

Currently, the increase in obesity rates is decelerating in high-income areas of China. In contrast, obesity is showing a significant increase in low-income areas, which indicates a requirement for targeted health policies to prevent a further increase in obesity among the general population [17]. Despite the growing popularity of smart body fat scales and the cost-effective potential of digital platforms for reaching a large number of individuals, there was limited research on their impact on weight loss across different economic regions. Here, we conducted a cohort study that analyzed the data of obesity-related anthropometric indices recorded through an mHealth app, which connected to the body fat scale, in individuals with overweight and obesity from low- and high-income regions of China. Our study aimed to investigate the weight loss effects of mHealth connected to smart body fat scales in different economic regions and to test our hypothesis that using mHealth in different economic regions could achieve significant weight loss and fat loss. This would also provide important insights into the potential for self-managed mHealth methods to promote healthy weight loss and improve health outcomes in this population.

**Methods**

**Participants**

This study analyzed the data of 165,635 adults aged 18 to 79 years who signed up for the Qingniu Health app between January 2020 and July 2022 and lived in different economic regions of China. Users with a baseline BMI outside the 95% range were excluded. The baseline data were used to determine the general and body composition characteristics of mHealth users from different economic regions. Furthermore, we followed 74,611 participants with a BMI ≥24.0 kg/m² and 22,903 participants with a normal BMI but an excessive PBF. All participants were followed for 6 months to assess weight and fat loss (Figure 1).
Ethical Considerations
This study was approved by the Ethics Committee of Xinhua Hospitals (approval XHEC-D-2022-195). Electronic consent was provided by all participants. The data were deidentified, and our study did not involve compensation.

Data Collection
The Qingniu Health app was used for the data collection. Participants were required to provide their baseline information, including gender, age, height, and city of residence, at the time of registering in the app. Body composition data, including weight, PBF, skeletal muscle rate (SMR), and basal metabolic rate (BMR), were measured using the same body fat scales (CS10C; Yolanda Technology Co., Ltd). These scales used multifrequency bioelectrical impedance analysis, as previously noted in another study [14].

Subgroups
For the analyses, participants were divided into groups based on gender, age, BMI, PBF, measurement frequency, and economic class of the residential city. Participants younger than 40 years, 40-59 years of age, and 60 years and older were categorized as adults, middle-aged adults, and older adults, respectively. The BMI classification followed standards set by the Ministry of Health’s Disease Control Department for Chinese People. Those with a BMI less than 18.5 kg/m² were considered underweight, those with a BMI between 18.5 and 23.9 kg/m² were considered normal, those with a BMI between
24.0 and 27.9 kg/m$^2$ were considered overweight, and those with a BMI of 28.0 kg/m$^2$ or more were considered obese [9]. Male adults with a PBF $\geq$ 25.0% and female adults with a PBF $\geq$ 30.0% were considered obese [18], whereas the others were considered nonobese. Individuals with normal weight but had a high PBF were considered to have “normal-weight obesity.” Participants were also evenly divided into groups based on their frequency of measurements: low, medium, and high. First-tier, new first-tier, and second-tier cities were classified as high-income regions, whereas third-tier, fourth-tier, and fifth-tier cities were classified as low-income regions.

Follow-Up Outcomes

The results of the follow-up were measured for 6 months (180±30 d) after the initial registration. A previous study showed that a 5% weight loss reduced the incidence of obesity-related diseases [19]. Therefore, participants who lost more than 5% of their initial body weight were considered to have achieved effective weight loss.

Statistical Analyses

The continuous variables were reported as means and SDs, and the categorical variables were reported as counts and percentages. The frequency of measurements was characterized by medians and IQRs. Differences among groups were analyzed using the unpaired, 2-tailed Student $t$ test, and the differences in the constituent ratios were evaluated with the $\chi^2$ test. Multivariable logistic regression analysis was used to identify the independent factors that influenced the results. The statistical analyses were performed using SPSS 26.0 (IBM Corp). A $P$ value <.05 was considered statistically significant.

Results

General Characteristics of Participants

A total of 165,635 Chinese adults registered in the Qingniu Health app were analyzed, consisting of 32,129 (19.4%) citizens from low-income regions and 133,506 (80.6%) citizens from high-income regions (Table 1). Users were mainly distributed in eastern coastal cities and inland provincial capitals (Figure 2). The majority of users were women (n=133,377, 80.5%) and those aged 18 to 40 years (n=126,796, 76.6%), and the proportion of women in low-income regions was greater than that in high-income regions (27,119/32,129, 84.4% vs 106,258/133,506, 79.6%; $P$<.001). Body composition analyses revealed that the proportion of individuals with overweight and obesity, as classified by BMI, and the proportion of individuals with obesity with a PBF above the upper limit were higher in low-income regions than in high-income regions ($P$<.001). On average, the baseline BMR and SMR of users in low-income regions were lower than those in high-income regions (BMR: mean 1339.99, SD 160.81 vs mean 1350.04, SD 174.93 kcal; $P$<.001; SMR: mean 41.33, SD 3.93 vs mean 42.06, SD 4.22%; $P$=.002). Based on the frequency of measurements during the 6 months, all users were equally divided into 3 groups: low (median 2.29, IQR 1.76 times), medium (median 7.26, IQR 3.95 times), and high (median 22.88, IQR 14.70 times).
Table. General characteristics of participants grouped by regional economic classification.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Low-income regions (n=32,129)</th>
<th>High-income regions (n=133,506)</th>
<th>( P ) value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total participants (n=165,635), n (%)</td>
<td>32,129 (19.4)</td>
<td>133,506 (80.6)</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Women, n (%)</td>
<td>27,119 (84.4)</td>
<td>106,258 (79.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age group (years), n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>18-40</td>
<td>24,382 (75.9)</td>
<td>102,414 (76.7)</td>
<td></td>
</tr>
<tr>
<td>40-60</td>
<td>7419 (23.1)</td>
<td>29,438 (22)</td>
<td></td>
</tr>
<tr>
<td>60-80</td>
<td>328 (1)</td>
<td>1654 (1.2)</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Underweight</td>
<td>696 (2.2)</td>
<td>3662 (2.7)</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>16,055 (50)</td>
<td>70,611 (52.9)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>10,625 (33.1)</td>
<td>41,389 (31)</td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>4753 (14.8)</td>
<td>17,844 (13.4)</td>
<td></td>
</tr>
<tr>
<td>Percentage of body fat, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Normal</td>
<td>12,983 (40.4)</td>
<td>61,473 (46)</td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>19,146 (59.6)</td>
<td>72,033 (54)</td>
<td></td>
</tr>
<tr>
<td>Baseline skeletal muscle rate (%), mean (SD)</td>
<td>41.33 (3.93)</td>
<td>42.06 (4.22)</td>
<td>.002</td>
</tr>
<tr>
<td>Baseline basal metabolic rate (kcal), mean (SD)</td>
<td>1339.99 (160.81)</td>
<td>1350.04 (174.93)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Frequency of measurements, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Low</td>
<td>10,638 (33.1)</td>
<td>44,608 (33.4)</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>10,488 (32.6)</td>
<td>44,768 (33.5)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>11,003 (34.2)</td>
<td>44,130 (33.1)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>All \( P \) values were compared between different economic regions.

<sup>b</sup>N/A: not applicable.
**Figure 2.** The number of users in mainland China. The pink areas have more than 2000 registrants, the purple areas have 1000-2000 registrants, and the blue areas have fewer than 1000 registrants.

**Baseline Characteristics of the Participants With Obesity**

In our study, 2 types of obesity were defined based on BMI and PBF: overweight and obese with a BMI $\geq 24.0$ kg/m$^2$ and normal-weight obese with a normal BMI but an excessive PBF. A total of 74,611 participants with overweight and obesity and 22,903 participants with normal-weight obesity were included in the analysis of baseline characteristics. The proportion of users with overweight and obesity in low-income regions was close to half (15,378/32,129, 47.9%), rising to 62.3% (20,021/32,129) when users with normal-weight obesity were included. High-income regions (77,493/133,506, 58%) had a slightly lower proportion compared to low-income regions. There was no significant difference in the distribution of measurement frequency among individuals with overweight and obesity between different economic areas ($P=.08$). The percentage of middle-aged adults was higher (3958/15,378, 25.7% vs 14,748/59,233, 24.9%; $P=.002$) and the mean baseline BMR was lower in low-income areas compared to high-income areas (Table 2).
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Overweight and obesity</th>
<th>Normal-weight obesity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low-income regions (n=15,378)</td>
<td>High-income regions (n=59,233)</td>
</tr>
<tr>
<td>Total participants (low-income regions: n=32,129; high-income regions: n=133,506), n (%)</td>
<td>15,378 (47.9)</td>
<td>59,233 (44.4)</td>
</tr>
<tr>
<td>Baseline basal metabolic rate (kcal), mean (SD)</td>
<td>1420.76 (165.97)</td>
<td>1446.57 (181.40)</td>
</tr>
<tr>
<td>Age group (years), n (%)</td>
<td>.002</td>
<td>.001</td>
</tr>
<tr>
<td>18-40</td>
<td>11,226 (73)</td>
<td>43,543 (73.5)</td>
</tr>
<tr>
<td>40-60</td>
<td>3958 (25.7)</td>
<td>14,748 (24.9)</td>
</tr>
<tr>
<td>60-80</td>
<td>194 (1.3)</td>
<td>942 (1.6)</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), n (%)</td>
<td>.06</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Overweight</td>
<td>10,625 (69.1)</td>
<td>41,389 (69.9)</td>
</tr>
<tr>
<td>Obesity</td>
<td>4753 (30.9)</td>
<td>17,844 (30.1)</td>
</tr>
<tr>
<td>Percentage of body fat, n (%)</td>
<td>&lt;.001</td>
<td>N/A</td>
</tr>
<tr>
<td>Normal</td>
<td>875 (5.7)</td>
<td>5462 (9.2)</td>
</tr>
<tr>
<td>Obesity</td>
<td>14,503 (94.3)</td>
<td>53,771 (90.8)</td>
</tr>
<tr>
<td>Frequency of measurements, n (%)</td>
<td>.08</td>
<td>.06</td>
</tr>
<tr>
<td>Low</td>
<td>4308 (28)</td>
<td>16,779 (28.3)</td>
</tr>
<tr>
<td>Medium</td>
<td>4850 (31.5)</td>
<td>19,073 (32.2)</td>
</tr>
<tr>
<td>High</td>
<td>6220 (40.4)</td>
<td>23,381 (39.5)</td>
</tr>
</tbody>
</table>

<sup>a</sup>All P values were compared between different economic regions.

<sup>b</sup>N/A: not applicable.

Moreover, the proportion of individuals with normal-weight obesity was also slightly higher in low-income regions than in high-income regions (4643/32,129, 14.5% vs 18,260/133,506, 13.7%; P<.001), and there was no significant difference in the mean baseline BMR of individuals with normal-weight obesity (mean 1234.80, SD 77.04 vs mean 1236.48, SD 77.93 kcal; P=.19; Table 2). Compared with the percentage of middle-aged adults with overweight and obesity (defined by BMI; 18.748/74.611, 25.1%), the percentage of middle-aged adults with normal-weight obesity was higher (9754/22,903, 42.6%).

### Weight Loss and Fat Loss of the Participants With Obesity

We followed the aforementioned individuals with obesity for 6 months and found that participants with overweight and obesity who used the self-monitoring mHealth app experienced significant weight loss and fat loss in both regions (low-income regions: mean –4.93, SD 6.41 kg; high-income regions: mean –4.71, SD 6.14 kg; P<.001). Men with overweight and obesity in low-income regions lost weight markedly more than those in high-income regions (P=.01; Table 3). When grouped by age, baseline BMI, and measurement frequency, weight loss was greater among users in low-income regions than in high-income regions. Additionally, individuals with normal-weight obesity showed less weight loss than individuals with overweight and obesity. In contrast, in the population with normal-weight obesity, individuals lost more weight in high-income regions than those in low-income regions.
Considering that the goal of participants with normal-weight obesity was to reduce body fat, we further explored the differences in the efficiency of fat loss in different economic regions. We found that the fat loss effect on the population with normal-weight obesity was greater in high-income regions. However, grouped by gender, age, and measurement frequency, we conducted unpaired 2-tailed Student $t$ tests for each group and found that there were no differences in the fat loss effect on people with overweight and obesity in different economic regions (Table 4).

### Table 4. Weight loss grouped by regional economic classification.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Weight loss in the overweight and obesity group (kg), mean (SD)</th>
<th>Weight loss in the normal-weight obesity group (kg), mean (SD)</th>
<th>$P$ value$^a$</th>
<th>Low-income regions</th>
<th>High-income regions</th>
<th>$P$ value$^a$</th>
<th>Low-income regions</th>
<th>High-income regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>All participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-$4.93 (6.41)</td>
<td>$-$4.71 (6.14)</td>
<td>$&lt;.001$</td>
<td>$-$2.23 (4.21)</td>
<td>$-$2.42 (4.07)</td>
<td>$.004$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>$-$4.40 (6.81)</td>
<td>$-$4.08 (6.51)</td>
<td>$.01</td>
<td>$-$1.62 (9.31)</td>
<td>$-$1.72 (4.55)</td>
<td>$.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>$-$5.08 (6.27)</td>
<td>$-$4.99 (5.94)</td>
<td>$.11</td>
<td>$-$2.24 (4.14)</td>
<td>$-$2.43 (4.06)</td>
<td>$.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-40</td>
<td>$-$5.08 (6.68)</td>
<td>$-$4.82 (6.37)</td>
<td>$&lt;.001$</td>
<td>$-$2.51 (4.39)</td>
<td>$-$2.72 (4.30)</td>
<td>$.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-60</td>
<td>$-$4.56 (5.59)</td>
<td>$-$4.39 (5.39)</td>
<td>$.07</td>
<td>$-$1.91 (3.95)</td>
<td>$-$2.11 (3.68)</td>
<td>$.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-80</td>
<td>$-$3.68 (4.86)</td>
<td>$-$4.33 (5.75)</td>
<td>$.14</td>
<td>$-$1.30 (4.10)</td>
<td>$-$1.31 (4.55)</td>
<td>$.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>$-$4.07 (5.48)</td>
<td>$-$3.92 (5.16)</td>
<td>$.01</td>
<td>N/A$^b$</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>$-$6.85 (7.77)</td>
<td>$-$6.53 (7.64)</td>
<td>$.01</td>
<td>N/A$^b$</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of body fat (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>$-$5.08 (6.40)</td>
<td>$-$4.94 (6.19)</td>
<td>$.01</td>
<td>$-$2.23 (4.21)</td>
<td>$-$2.42 (4.07)</td>
<td>$.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of measurements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>$-$3.24 (7.44)</td>
<td>$-$2.92 (6.74)</td>
<td>$.009</td>
<td>$-$0.84 (5.00)</td>
<td>$-$1.28 (4.71)</td>
<td>$.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>$-$3.96 (5.72)</td>
<td>$-$3.84 (5.53)</td>
<td>$.20</td>
<td>$-$1.94 (3.61)</td>
<td>$-$2.09 (3.72)</td>
<td>$.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>$-$6.86 (5.58)</td>
<td>$-$6.70 (5.57)</td>
<td>$.046</td>
<td>$-$3.50 (3.64)</td>
<td>$-$3.51 (3.57)</td>
<td>$.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$All $P$ values were compared between different economic regions.

$^b$N/A: not applicable.
Table. Fat loss of participants with obesity grouped by regional economic classification.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Fat loss in the overweight and obesity group (PBF(^a); %), mean (SD)</th>
<th>Fat loss in the normal-weight obesity group (PBF; %), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low-income regions</td>
<td>High-income regions</td>
</tr>
<tr>
<td>All participants</td>
<td>–2.06 (3.14)</td>
<td>–2.04 (3.19)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>–2.00 (3.57)</td>
<td>–2.03 (3.76)</td>
</tr>
<tr>
<td>Women</td>
<td>–2.07 (3.01)</td>
<td>–2.04 (2.91)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-40</td>
<td>–2.13 (3.26)</td>
<td>–2.11 (3.30)</td>
</tr>
<tr>
<td>40-60</td>
<td>–1.87 (2.82)</td>
<td>–1.85 (2.86)</td>
</tr>
<tr>
<td>60-80</td>
<td>–1.31 (2.12)</td>
<td>–1.70 (2.83)</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>–1.85 (2.92)</td>
<td>–1.82 (2.95)</td>
</tr>
<tr>
<td>Obesity</td>
<td>–2.52 (3.55)</td>
<td>–2.55 (3.65)</td>
</tr>
<tr>
<td>PBF (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>–2.13 (3.09)</td>
<td>–2.15 (3.13)</td>
</tr>
<tr>
<td>Frequency of measurements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>–1.42 (3.80)</td>
<td>–1.33 (3.60)</td>
</tr>
<tr>
<td>Medium</td>
<td>–1.64 (2.92)</td>
<td>–1.68 (2.94)</td>
</tr>
<tr>
<td>High</td>
<td>–2.81 (2.61)</td>
<td>–2.84 (2.90)</td>
</tr>
</tbody>
</table>

\(^a\)PBF: percentage of body fat.
\(^b\)All \(P\) values were compared between different economic regions.
\(^c\)N/A: not applicable.

Independent Factors Linked With Successful Weight Loss in Different Economic Regions

We conducted multivariable logistic regression analyses to examine the relationship between various baseline factors and successful weight loss in individuals with overweight and obesity. In the population from low-income regions, younger participants with higher baseline PBF, baseline SMR, and measurement frequency were more likely to succeed in weight loss (Table 5). In addition, women in high-income regions were more likely to achieve successful weight loss. The results indicated that the frequency of measurement was the most critical independent factor in both low-income (odds ratio 5.036, 95% CI 4.618-5.491; \(P<.001\)) and high-income (odds ratio 5.271, 95% CI 5.042-5.511; \(P<.001\)) regions, particularly a high measurement frequency. Similarly, we further estimated the dependent factors associated with successful fat loss in the populations with overweight and obesity. In different economic regions, younger men with a higher baseline PBF and measurement frequency and lower BMR were more likely to lose body fat (Table 5).
Table. Factors linked with successful weight loss and fat loss in overweight and obese group.

| Characteristics | Weight loss | | Fat loss | |
|-----------------|-------------|-----------------|-------------|
|                 | Low-income regions | High-income regions | Low-income regions | High-income regions |
|                 | OR (95% CI) | P value | OR (95% CI) | P value |
| Gender          | N/A b | N/A | 0.748 (0.680-0.823) | <.001 | 2.568 (2.139-3.081) | <.001 |
| Age             | 0.981 (0.977-0.985) | <.001 | 0.983 (0.981-0.985) | <.001 | 0.971 (0.967-0.975) | <.001 |
| Baseline PBF c  | 1.109 (1.087-1.132) | <.001 | 1.079 (1.071-1.088) | <.001 | 1.047 (1.034-1.060) | <.001 |
| Baseline SMR d  | 1.052 (1.029-1.075) | <.001 | 1.036 (1.025-1.048) | <.001 | N/A | N/A |
| Baseline BMR e  | N/A | N/A | 1.001 (1.000-1.001) | <.001 | 0.999 (0.999-1.000) | <.001 |
| Frequency of measurements classification | | | | |
| Medium-low      | 1.611 (1.479-1.755) | <.001 | 1.771 (1.695-1.851) | <.001 | 1.557 (1.427-1.700) | <.001 |
| High-low        | 5.036 (4.618-5.491) | <.001 | 5.271 (5.042-5.511) | <.001 | 4.606 (4.224-5.023) | <.001 |

aOR: odds ratio.  
bN/A: not applicable.  
cPBF: percentage of body fat.  
dSMR: skeletal muscle rate.  
eBMR: basal metabolic rate.

**Discussion**

**Principal Findings**

After 6 months of follow-up, the results showed that significant weight and fat loss were found in participants with overweight and obesity using the mHealth app in different economic regions. Furthermore, individuals in low-income regions lost more weight than individuals in high-income regions in the population with overweight and obesity, and there was no difference between individuals in fat loss. Providing further health education and online weight loss monitoring in low-income regions was beneficial to the population. Interestingly, in the population with normal-weight obesity, individuals in high-income regions lost more fat than individuals in low-income regions.

**Comparison With Prior Work**

Obesity has become a major public health challenge worldwide, with strong links to metabolic disorders such as cardiovascular diseases [5] and diabetes [20]. With the transformation of China’s economic and social structure, dietary patterns and nutritional status have undergone significant changes, and the problem of obesity has become increasingly serious [21]. Addressing obesity can help reduce the incidence of chronic diseases [22], and the economic benefits of long-term nonsurgical weight loss in individuals with obesity have been well established [23]. In the previous study, we found that a large number of Chinese individuals with overweight and obesity were able to achieve weight loss goals through an mHealth app during a long-term follow-up, leading us to conclude that mHealth with body fat scales might be a promising method for weight loss and fitness [24]. However, given the vast territory and large population of China (Figure 2), we aimed to investigate whether self-management through mHealth could achieve consistent results across different economic regions, further demonstrating the feasibility of online self-management for weight loss.

The mHealth app with the body fat scale is a widely used self-weighing tool due to its affordability and convenience, different from the traditional paper record [25]. We analyzed the user data of the Qingniu Health app linked to the body fat scale. We divided all the included users into 2 groups: high-income regions and low-income regions according to the economic classification of their living regions. Results showed that the number of users in high-income regions was more than quadruple that in low-income regions (n=133,506 vs n=32,129).
and the vast majority of users were women in both groups, which was similar to other research [26]. This revealed that people in high-income regions paid more attention to obesity prevention and monitoring, particularly women. Furthermore, the age distribution of users showed a majority of young and middle-aged people. When grouped by baseline BMI and PBF, both the proportion of people with an excessive BMI and the proportion of people with an excessive PBF in low-income regions were significantly higher than those in high-income regions, as noted in a comment published in Nature [27]. In all high-income countries, overweight and obesity levels were already higher in rural areas than in urban areas, and the same phenomenon might be occurring in China [27,28]. These results suggest that self-management and lifestyle interventions are required to prevent further development of metabolic-related diseases in people with excess body weight and body fat in low-income areas.

Previous studies have established a close association between socioeconomic status and the risk of obesity [29-31]. Despite this, the relationship between weight loss and economic regions among Chinese adults remains largely unknown. Our study of the user data of the Qingniu Health app linked to the body fat scale found that nearly half of the users met the criteria for a diagnosis of obesity based on their BMI. The average baseline BMR for users in low-income regions was lower, and the individual BMR was correlated with multiple factors such as age, gender, body composition, and BMI [32]. There were several users with overweight and obesity with a normal baseline PBF and some people with a normal BMI but excessive PBF, which we called the population with normal-weight obesity. Normal-weight obesity was closely associated with metabolic and cardiovascular diseases [33]. Our findings showed that the proportion of the population with normal-weight obesity in low-income regions was higher than in high-income regions. Moreover, the increase in the proportion of middle-aged adults with normal-weight obesity was associated with an increase in biological age, a lack of physical activity, and other factors. This phenomenon highlighted the need for increased focus on body fat control in this age group [34]. As expected, we found no statistical difference in the composition of measurement frequency between participants in different economic regions, showing that users were able to maintain a certain frequency of use with the app, which suggested that promoting the use of mHealth in our country was a feasible and effective approach to monitoring obesity.

In our previous study [24], we found that a younger age was the most important contributing factor to fat loss success. In this study, the age composition of participants with normal-weight obesity in high-income regions was not different from that in low-income regions but showed a more obvious trend of weight loss and fat loss, reflecting that participants from high-income regions had focused on body fat in addition to body weight. To better understand the factors affecting successful weight loss in different economic regions, we found that age, baseline PBF, SMR, and measurement frequency were significant predictors for weight loss success, whereas gender and baseline BMR only showed a correlation with weight loss in the population in high-income regions, which suggests women in high-income regions paid more attention to self-management than those in low-income regions.

Limitations
Our study had a large sample size and a prolonged follow-up period, which accurately reflected the real-world situation of online self-management in China. Nevertheless, there were certain limitations in our study. We must acknowledge that our study was a nonrandomized cohort study without controls. Nevertheless, we observed a significant weight loss outcome in individuals with overweight and obesity from the 2 different economic regions in this large-scale follow-up study. Furthermore, individuals who used the smart body fat scale with a higher frequency exhibited better weight loss results compared to those with a lower use frequency. This suggests that the use of an mHealth app connected to a smart body fat scale has a certain impact on effective weight reduction, regardless of economic region. In the future, we can consider adding control groups using alternative weight loss methods to further clarify the role of using an mHealth app connected to a smart body fat scale in self-initiated weight and fat loss. Additionally, we did not consider the 2 important influencing factors of participants’ diet and exercise volume. We also did not consider the overall health status of the participants; factors such as hydration level might impact the accuracy of body fat measurement using bioelectrical impedance analysis.

Conclusions
A high proportion of individuals with obesity from low-income regions was found in our study, and individuals with overweight and obesity who used body fat scales in different economic regions of China experienced significant weight loss and fat loss. Individuals from high-income regions paid more attention to body fat and had better fat loss than those from low-income regions, and in the middle-aged population, the issue of normal-weight obesity required more attention. Therefore, promoting self-monitoring of weight and fat through the use of body fat scales connected to an mHealth app could be an important intervention measure for the population with overweight and obesity across all regions of China.

Acknowledgments
This study was supported by the Chinese National Natural Science Foundation (grant 82270867) and the Natural Science Foundation of Shanghai (grant 20ZR1435300). We also acknowledge all the users involved in the study and the staff of the Qingniu Health app, Yolanda Technology Co., Ltd.
Data Availability
Data are available from the corresponding author with the permission of the Qingniu Health app, Yolanda Technology Co., Ltd.

Authors' Contributions
XH, YS, and HY contributed equally. XH, YS, HY, and WJ designed and wrote the manuscript. ML and ZL analyzed the data. JS, BL, and WZ reviewed the analyses and manuscript. All authors approved the final version.

Conflicts of Interest
None declared.

References


Abbreviations

BMR: basal metabolic rate
mHealth: mobile health
PBF: percentage of body fat
SMR: skeletal muscle rate

© Xinru Huang, Yefei Shi, Hongyun Yao, Mingjie Li, Zhijun Lei, Jiayun Shi, Bo Li, Weiwei Zhang, Weixia Jian. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.1.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Use and Engagement With Low-Intensity Cognitive Behavioral Therapy Techniques Used Within an App to Support Worry Management: Content Analysis of Log Data

Paul Farrand¹²*, PhD; Patrick J Raue³*, PhD;Earlise Ward⁴*, PhD;Dean Repper⁵*, MSc; Patricia Areán³*, PhD

¹Clinical Education, Development and Research, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom
²Department of Psychology, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom
³AIMS CENTER, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States
⁴School of Medicine and Public Health, Carbone Comprehensive Cancer Center, University of Wisconsin-Madison, Madison, WI, United States
⁵Trent PTS, Improving Access to Psychological Therapies, Derby, United Kingdom

*all authors contributed equally

Corresponding Author:
Paul Farrand, PhD
Clinical Education, Development and Research
Faculty of Health and Life Sciences
University of Exeter
Sir Henry Wellcome Building for Mood Disorders Reserach
Perry Road
Exeter, EX4 4QG
United Kingdom
Phone: 44 01392725793
Email: p.a.farrand@exeter.ac.uk

Abstract

Background: Low-intensity cognitive behavioral therapy (LICBT) has been implemented by the Improving Access to Psychological Therapies services across England to manage excessive worry associated with generalized anxiety disorder and support emotional well-being. However, barriers to access limit scalability. A solution has been to incorporate LICBT techniques derived from an evidence-based protocol within the Iona Mind Well-being app for Worry management (IMWW) with support provided through an algorithmically driven conversational agent.

Objective: This study aims to examine engagement with a mobile phone app to support worry management with specific attention directed toward interaction with specific LICBT techniques and examine the potential to reduce symptoms of anxiety.

Methods: Log data were examined with respect to a sample of “engaged” users who had completed at least 1 lesson related to the Worry Time and Problem Solving in-app modules that represented the “minimum dose.” Paired sample 2-tailed t tests were undertaken to examine the potential for IMWW to reduce worry and anxiety, with multivariate linear regressions examining the extent to which completion of each of the techniques led to reductions in worry and anxiety.

Results: There was good engagement with the range of specific LICBT techniques included within IMWW. The vast majority of engaged users were able to interact with the cognitive behavioral therapy model and successfully record types of worry. When working through Problem Solving, the conversational agent was successfully used to support the user with lower levels of engagement. Several users engaged with Worry Time outside of the app. Forgetting to use the app was the most common reason for lack of engagement, with features of the app such as completion of routine outcome measures and weekly reflections having lower levels of engagement. Despite difficulties in the collection of end point data, there was a significant reduction in severity for both anxiety (t_{53}=5.5; P<.001; 95% CI 2.4-5.2) and low mood (t_{53}=2.3; P=.03; 95% CI 0.2-3.3). A statistically significant linear model was also fitted to the Generalized Anxiety Disorder–7 (F_{2,51}=6.73; P<.001), while the model predicting changes in the Patient Health Questionnaire–8 did not reach significance (F_{2,51}=2.33; P=.11). This indicates that the reduction in these measures was affected by in-app engagement with Worry Time and Problem Solving.

Conclusions: Engaged users were able to successfully interact with the LICBT-specific techniques informed by an evidence-based protocol although there were lower completion rates of routine outcome measures and weekly reflections. Successful interaction
with the specific techniques potentially contributes to promising data, indicating that IMWW may be effective in the management of excessive worry. A relationship between dose and improvement justifies the use of log data to inform future developments. However, attention needs to be directed toward enhancing interaction with wider features of the app given that larger improvements were associated with greater engagement.

(JMIR Mhealth Uhealth 2024;12:e47321) doi:10.2196/47321

KEYWORDS
cognitive behavioral therapy; low-intensity; mCBT; app; log data; worry management; CBT; management; application; therapy; implementation; treatment; symptoms; anxiety; worry; engagement

Introduction

Excessive worry represents a core characteristic associated with generalized anxiety disorder (GAD) [1], characterized as 6 or more months of chronic worry about several different events and little belief worry can be controlled, and is associated with general somatic symptoms of anxiety [2]. It is highly pervasive in high-income countries, with a lifetime prevalence rate of 7.8% in the United States [3] and a median age of onset between 24 and 50 years, and is more common in women and people who are unemployed [4]. Excessive worry is deleterious to the individual, and if it manifests into GAD, it places a significant burden on society and employers with an average of 6.3 days per month of work absenteeism [5]. Furthermore, treatment is associated with increased service use [6], thereby placing a significant burden on primary care settings within both the United Kingdom [7] and United States [8].

Despite excessive worry impacting on the individual quality of life with progression to GAD representing a major public health problem [9], treatment availability remains limited. The treatment gap for GAD has been reported to be as high as 67% in the United Kingdom and 70% in the United States [10]. In an attempt to address the treatment gap [11], innovations in the delivery of evidence-based psychological therapy have been witnessed, for example, by broadening the workforce delivering cognitive behavioral therapy (CBT) for late-life GAD with no differences in effectiveness demonstrated when delivered by bachelor-level lay providers compared with PhD-level experienced therapists [12]. Further attempts to address the treatment gap have been addressed within the Improving Access to Psychological Therapies (IAPT) program, implementing low-intensity cognitive behavioral therapy (LICBT) self-help for the management of mild to moderate common mental health problems alongside therapist-delivered high-intensity CBT for moderate to severe presentations [13].

While improved access has been facilitated through the IAPT program, predictions indicate that access will only be increased to 25% of the community prevalence of depression and anxiety disorders by 2023-2024 [14]. Furthermore, between 2021 and 2022, only 37% of patients completed a course of therapy [15]. Difficulties in reducing the treatment gap are further dependent on a number of individual-level barriers such as stigma, desire to handle the problem independently, and limited willingness to disclose problems [11].

Greater implementation of digital health technologies such as smartphone apps [16] has potential to address barriers at the level of the individual. Furthermore, with high use of smartphones, for example, 81% of adults within the United States [17], apps offer the promise of delivering mental well-being interventions at scale and derive cost savings with respect to delivery and increased productivity within the workplace [18]. The expansion of apps to manage challenges with well-being has predominantly been based on CBT [19]. CBT is particularly well suited to informal mental well-being apps with emphasis placed on active engagement with specific techniques rather than exclusive reliance on a supportive relationship with a behavioral health coach or therapist [13]. This is especially salient with regard to an approach based on “collaborative empiricism,” whereby after engagement with specific CBT techniques, people are encouraged to explore outcomes for themselves [20]. In doing so, a better understanding of the way their mental health difficulty is affecting them can be derived through an appreciation of the cognitive behavioral model.

Despite CBT providing a compatible evidence-based approach for mental well-being apps with the potential to deliver at scale to close the treatment gap, implementation and uptake remain limited [21]. This is partly dependent on poor engagement with apps relying on factors such as poor usability, techniques inconsistent with user expectations, and poor health literacy [22]. Exploring ways to enhance engagement is of significance given that greater engagement has been reported to predict improvements in mental well-being [23]. In particular, focusing on obstacles and difficulties encountered in completing LICBT techniques is of significance given their effectiveness [24], while not dismissing common factors to establish a good “therapeutic relationship” generating a sense of genuineness, warmth, and collaborative working [25]. Focusing on both common and specific LICBT techniques used within the Iona Mind Well-being app for Worry management (IMWW) is therefore of importance given that the combination of both is crucial for bringing about therapeutic change [26].

This paper reports log data [27] to examine engagement with IMWW based on LICBT to help in the management of excessive worry. It has been proposed that rather than looking at overall engagement regarding areas such as number of sessions engaged with or session duration, it is better to focus attention on specific user interactions [28]. The focus of this paper is therefore directed toward appreciating engagement and interaction with specific LICBT techniques and wider features of IMWW to examine potential enhancements alongside wider usability. Furthermore, the relationship between engagement and outcomes...
will be explored to inform ongoing development to maximize effectiveness.

**Methods**

**Design**

Areas representing the focus of log data analysis have been informed by recommendations identified as useful when seeking to address the effectiveness of eHealth technology alongside behavioral and theoretical models [27]. Analysis was only undertaken on data collected regarding “engaged users” and their log data associated with engagement with the LICBT techniques. To be considered an engaged user, the user was required to have completed at least 1 lesson related to the Worry Time and Problem Solving in-app modules within any time period. These criteria represented the “minimum dose” [29] necessary for the user to be considered to have engaged enough to be able to understand the in-app CBT techniques and apply them outside of the app to manage excessive worry. This approach to represent “engaged users” has previously been adopted. For example, with respect to a feasibility trial examining internet-administered LICBT for parents of children treated for cancer [30].

A $\chi^2$ test of homogeneity was undertaken to compare demographic data provided by the engaged users and those who failed to engage with specific LICBT techniques to receive a minimum dose.

**Sample**

Over 6 months (July 19, 2022, to February 19, 2023), 956 adults 18 years and older with a GAD-7 score of 6 and above downloaded and opened IMWW and completed the sign-up process. Of these, 803 (84%) adults did not engage sufficiently with the app to be considered an “engaged user,” resulting in 153 (16%) users engaging with the app sufficiently to be considered to have received a minimum dose (Table 1).

Approximate data regarding the continent the user was accessing the app from were automatically collected by the app from the time zone set on the user’s phone and therefore collected on all 956 adults.

$\chi^2$ tests of homogeneity indicate that differences between engaged users and those who had downloaded the app but failed to receive a minimum dose were not significant at the 5% level across any of the demographic variables. In addition to the collection of demographic data, 41 of 153 (27%) engaged users responded to a question regarding receipt of other therapy, of whom 34 (83%) users indicated that they were not.

Table 1. Demographic questionnaire responses completed (N=956).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Engaged users (n=153)</th>
<th>Not received minimum dose (n=803)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>43 (72)</td>
<td>50 (85)</td>
</tr>
<tr>
<td>Men</td>
<td>14 (23)</td>
<td>9 (15)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Age range (years; n=53), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>18 (34)</td>
<td>16 (30)</td>
</tr>
<tr>
<td>25-34</td>
<td>20 (38)</td>
<td>19 (36)</td>
</tr>
<tr>
<td>35-44</td>
<td>9 (17)</td>
<td>12 (23)</td>
</tr>
<tr>
<td>45-54</td>
<td>2 (4)</td>
<td>6 (11)</td>
</tr>
<tr>
<td>55-64</td>
<td>1 (2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>≥65</td>
<td>3 (6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Continent, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Americas</td>
<td>78 (51)</td>
<td>369 (46)</td>
</tr>
<tr>
<td>Europe</td>
<td>33 (22)</td>
<td>196 (24)</td>
</tr>
<tr>
<td>Asia</td>
<td>28 (18)</td>
<td>161 (20)</td>
</tr>
<tr>
<td>Africa</td>
<td>5 (3)</td>
<td>22 (3)</td>
</tr>
<tr>
<td>Australasia</td>
<td>4 (3)</td>
<td>32 (4)</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (3)</td>
<td>23 (3)</td>
</tr>
</tbody>
</table>

A total of 60 engaged users and 59 users who did not receive the minimum dose responded.

**Iona Mind Well-Being App for Worry Management**

The IMWW is, in part, informed by the techniques described in the LICBT written self-help intervention for managing excessive worry [31], based on the CBT protocol for the management of GAD [32]. LICBT is recommended for the management of GAD [33] and is one of the most commonly adopted written self-interventions used within the IAPT program [34]. The focus of IMWW is explained during onboarding where the user is required to explicitly acknowledge its purpose as a
well-being tool. Users wishing to continue engagement acknowledge that they understand conditions related to use and consent to have their data processed.

**Collection of Demographic Data**

Demographic data were not used to inform the delivery or functionality of the app. Therefore, a screen requesting demographic data, or a question regarding receipt of other therapy, was only presented once the user had engaged with IMWW on 2 occasions at least 6 hours apart following enablement, and no other higher priority messages were pending. If higher priority messages were pending, the request to provide demographic data was repeatedly postponed to the following day until the user had supplied all data, completed specific questions, or declined the request to open the screen (Table 1). Due to the optionality and logic surrounding whether demographic data were requested from users who downloaded the app, such data were therefore not requested from all users and were not prioritized over other more useful app functionality. Collecting demographic data was not prioritized given that answering such questions on an app can increase the risk of disengagement [35].

**Supporting Interaction**

Interaction is supported by an algorithmically driven chatbot simulating a “conversation” between the conversational agent (CA) and the user. Users interact by entering raw text or selecting a predetermined response. The type of response depends on the type of content being delivered and varies between selecting a button from a list of options or entering free text when a personalized response is requested. User feedback is collected at the level of a user message supplied within the “conversation” through selecting a “thumbs up” or “thumbs down” icon.

Key principles associated with user-centered systems design were adopted to inform the development of IMWW [36]. The app was developed to manage excessive worry and support emotional wellness with an “SOS” button prominently displayed for users finding themselves in significant emotional distress. If selected, signposting information to a comprehensive list of local and international crisis helplines is presented alongside free text when a personalized response is requested. User feedback is collected at the level of a user message supplied within the conversation through selecting a “thumbs up” or “thumbs down” icon.

**Progress Through IMWW**

Engagement begins with users landing on the Today home screen and progressing through 6 educational modules through which they learn about and interact with the LICBT techniques (Multimedia Appendix 1).

Educational modules are chronologically ordered and unlocked as engagement is initiated. The order in which they are unlocked is dependent on user choice, reflecting whether they wish to initially address practical or hypothetical worries. However, the user is able to move forward and backward between these specific factors to address the different types of worries where preferred. After onboarding, the user is given the opportunity to complete the GAD-7 and Patient Health Questionnaire-8 (PHQ-8) weekly during their weekly review, with scores presented on a progress screen. However, following the completion of these measures during onboarding, subsequent completion is voluntary.

**Home Screen**

Informed by the CBT protocol for the management of GAD, which can also be used in the context of improving emotional well-being by supporting the management of worries [37], the Today (Home) screen supports the user to record their worries. This screen also presents a timer that counts down to the user’s scheduled worry time and offers tips to complete Worry Time (Multimedia Appendix 1). Should the user not have completed the lesson, a placeholder is displayed inviting them to learn more about Worry Time. The screen has been designed to make it as easy as possible for users to record worries, plan Worry Time, and access the CBT content. A navigation bar at the bottom of the screen links to the screens related to the LICBT techniques through which users can engage depending on preference.

**CBT Techniques**

To promote engagement, the IMWW is informed by collaborative empiricism where the user actively engages with techniques associated with CBT [38]. Collaborative empiricism has been identified as core to the therapeutic relationship [39], supporting “learning by doing” fundamental to CBT [40]. Accordingly, LICBT techniques are presented as skills to be mastered through regular practice. The user is initially introduced to the CBT model followed by supporting them to record and categorize their worries. The user is then provided with the choice of Problem Solving or Worry Time to address practical or hypothetical worries, respectively.

**CBT Model**

An interactive CBT Five Areas model (introduced in October 2022) is presented [41], and the user is encouraged to interact and identify a current situation in the “here and now.” In response to this current situation, the user is also encouraged to interact with boxes reflecting “Thoughts” that go through their head, “Behaviors” engaged in, “Physical Feelings,” and to recognize “Emotions” (Multimedia Appendix 1). Additionally, the model helps them appreciate ways in which the specific LICBT techniques presented to address practical and hypothetical worries may be helpful. The model serves as psychoeducation, enabling the user to recognize the interaction between each of these areas, understand the nature of their worry, and appreciate factors that maintain their worry behavior and the impact of physical symptoms associated with anxiety.

**Worry Diary**

The primary function of the Worry Diary is to enable the user to actively add new worries as they arise throughout the day and as a record of worries for subsequent review (Multimedia Appendix 1). Prompts and predefined categories are used to
enable the user to differentiate between practical worries that have a solution and hypothetical worries that do not.

**Problem Solving**

Where practical worries are logged, the user is guided to list potential solutions, consider strengths and weaknesses for each solution, and select the most appropriate one. A time to try the solution out is then optionally scheduled by the user with a reminder given to complete it at the chosen time. After the chosen time has passed or 30 hours elapsed, on opening the app, the CA will ask the user to review how their solution went. Prompts ask the user if the problem was resolved and if not request further information regarding the challenges encountered. Advice is offered where problems have been encountered, putting the solution into action (eg, to break the problem down or work through and apply another solution).

**Worry Time**

Unless explicitly overridden by the user, all worries identified as hypothetical are displayed only during Worry Time at a time determined by the user and are blurred out at all other times. Users are reminded that their worry time is starting with a push notification to their phone. If the user opens the app during Worry Time, they are prompted to work through the time they have set aside to worry with the CA. If they choose to do so, the CA will list out all user worries and request that these be worried about for the specified period of time. Subsequently, the CA will review each worry with the user, asking them if it still remains an issue or if it now better represents a practical worry. Worry Time represents a form of cognitive exposure with users exposed to hypothetical worries written down during the day. This is proposed to overcome avoidance behavior and reduce intolerance of uncertainty when it is recognized that there are no solutions to the worry [37].

**Maintaining Engagement**

A chatbot informed by theoretically driven techniques is adopted to help establish a “therapeutic approach” to maintain and promote user engagement embedded within IMWW [38]. Such techniques help to establish an approach based on collaborative empiricism [39], whereby the user is encouraged to explore outcomes arising from engagement for themselves.

**Conversational Agent**

Support is omnipresent throughout the engagement and comes in the form of an algorithmically driven chatbot stimulating engagement between IMWW, the CA, and users. This helps them overcome difficulties encountered with the specific LICBT techniques and uses common factors to maintain engagement. Upon recognition that difficulties are experienced with any of the specific techniques, the CA is deployed to enable users to work through the specific techniques. If the user reports difficulties in trying out a solution they have planned, the CA will ask questions to determine the nature of the difficulty encountered and direct them to the appropriate parts of the app. For example, the user would be directed back to Problem Solving should they need to break the problem down, or Worry Diary if the worry appears to be hypothetical rather than practical (Table 1).

On other occasions, the CA provides the user with helpful tips and advice or the opportunity to ask FAQs to navigate difficulties experienced. For example, 2 days after learning about Worry Time, the CA will check back in with the user and ask how the exercise has been going. Depending on user response, advice will be given. For example, if the user forgets to engage with Worry Time, they are reminded to turn on their notifications and set an alarm on their phone to serve as a prompt. Consistent with the delivery of CBT, during engagement with the LICBT techniques, the CA brings the user back to the CBT model to reinforce their understanding of the intervention and maintain motivation for continued engagement.

**Common Factors**

The CA uses nontherapeutic common factor skills in the form of “therapeutic empathy” to instill a sense of hopefulness and encouragement to maximize engagement with the specific factors linked to symptom reduction [42]. Statements include those demonstrating an empathic stance highlighting a desire to help alongside empathic attunement where statements demonstrate an appreciation of the user’s emotional experience [42]. When recognizing that the user is experiencing difficulties in engaging or is not improving, the CA uses empathy to maintain engagement.

**Behavior Change Techniques**

Within the module on recording worries, Behavioral Contracting [43] encourages the user to sign an agreement to consistently engage with IMWW throughout the 6-week program with a separate Goal Setting lesson guiding the user to set approach, rather than avoidance, goals [44]. Behavior “Push” notifications serve to prompt or maintain behavior change while engaging with the app. Furthermore, constructs derived from self-determination theory [45] promote autonomy and intrinsic motivation that serve to facilitate collaborative empiricism [37].

**Monitoring Progress**

Throughout engagement, a progress screen presents a summary of the user’s app use and engagement with in-app lessons, previously entered goals, and scores regarding symptom severity associated with anxiety (GAD-7) [46] and low mood (PHQ-8) [47]; it also presents links to the settings page, which houses operational features such as typing speed (Multimedia Appendix 1). Given a potential association between providing feedback and improved outcomes, all data collected are repeatedly presented to the user throughout engagement [48].

**Weekly Reflection**

Consistent with face-to-face CBT [39], on a weekly basis, the CA prompts the user to reflect on their engagement with IMWW and the features found most helpful (Table 1). Using reflective learning within the app facilitates learning, with the CA encouraging engagement to promote self-discovery [49]. Furthermore, during the weekly reflection, the CA requests information on the LICBT techniques engaged with and highlights those found most helpful. In the event an identified technique was not engaged with, the CA also requested information as to the main reasons from a range of options provided.
Data Collection and Analysis

Log Data
Consistent with the aims of the study, analysis was undertaken on log data collected from engaged users to reflect their engagement with the specific LICBT techniques. Log data were collected by IMWW automatically logging the actions of each engaged user and requests to complete surveys throughout the use of the app. From these data, summary statistics for use in this paper were extracted. Progress of engaged users through IMWW was monitored and informed by data regarding the number of sessions completed, session duration, weekly reflections, and completion of LICBT techniques alongside summary statistics recorded. Engagement with IMWW was explored with respect to the number of users who reached the CBT model and interacted with it, text entered into each area, completion of the LICBT techniques, and general input and behavior during the engagement. Specific worry management techniques were examined with respect to the number of worries entered and the proportion classified as practical or hypothetical problems. With respect to Problem Solving, data analysis included the number of users who completed the lesson, the number of times the in-app tool was used to solve a practical problem, the number of practical problems entered, the number of users prompted to follow up on their problem-solving with the CA, and the number who engaged with it. Furthermore, analysis was undertaken on the number of engaged users who sought to manage hypothetical worries by learning about Worry Time, set a time for Worry Time, and started an in-app session alongside the number of hypothetical worries entered being recorded.

User responses from the Weekly Reflection conversation within IMWW were also collected and analyzed to gauge general engagement with the specific techniques. As a proxy for behavior change approaches adopted to maintain engagement within the app, the number of users who were delivered at least 1 push notification and the number of those who interacted were also examined. For engaged users completing more than 1 GAD-7 or PHQ-8 at assessment, the log of assessments and the number of times IMWW was used for more than 10 seconds, which is defined as a “session,” were analyzed.

Potential Effectiveness
To examine the potential effectiveness of IMWW for engaged users, separate paired samples t-tests were undertaken to examine the difference between outcome data collected regarding the severity of anxiety (GAD-7) and low mood (PHQ-8). This analysis was only undertaken for the 54 of 153 (35%) engaged users who completed the outcome measures during onboarding and at the end of the engagement.

Impact of Engagement on Potential Effectiveness
A multivariate linear regression model was used to investigate the impact that engagement with IMWW had on improvement in anxiety and low mood. In particular, the extent to which specific features were used to complete therapeutic exercises impacted on scores over time. Engagement with, and completion of, Worry Time and Problem Solving was expected to lead to improvements in the symptoms of anxiety, and hence a model to analyze this was specified. Because there are multiple discrete interventions being applied within IMWW and the dependent variable is not univariate, the multiple regression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ was adopted. Within this model, $y$ is the change in GAD-7 or PHQ-8 from the initial score at onboarding to the final input during progress review, $x_1$ is the binary variable indicating whether the user completed Problem Solving and resolved their problem, $x_2$ is the binary variable indicating whether the user completed at least 1 instance of Worry Time in-app, and $\epsilon$ is the stochastic error term. Additional controls were added to the model to examine the extent to which the number of in-app sessions completed, and the number of worries, problems, and solutions recorded predicted improvement in GAD-7 and PHQ-8. All models met OLS model assumptions associated with multicollinearity, heteroskedasticity, and normality of residuals.

Ethical Considerations
Users were only able to download IMWW after agreeing to Iona Mind’s Terms of Service and Privacy policy, which required them to acknowledge that they understand conditions related to use and consent to have their anonymized data processed. Being based on anonymous, routinely collected log data from a nonclinical population, research ethics was not required for this study.

Results
General Engagement
Analysis of log data collected from the 153 engaged users indicated engagement with 1108 sessions (mean 7.2, SD 7.7) with an average session length of 6.2 (SD 6.2) minutes. The number of sessions and session length varied significantly across users with a median session length of 4.5 minutes and 6 being the median number of sessions (Multimedia Appendix 2).

Engagement With LICBT Techniques
Since inclusion (October 2022), 36 users started filling out the CBT model to reflect their current difficulties with anxiety, and of these, 31 (86%) users completed all areas in an average of 2.4 minutes (SD 1.7; median 1.7 minutes). The lesson on the Worry Time technique was successfully completed by almost all users (147/153, 96%). However, of these users, only 50 (33%) were observed to have performed Worry Time at their chosen time using the in-app tools.

Problem Solving was engaged with by 114 of 153 (74.5%) users; however, only 89 of 153 (58.2%) users actually completed the lesson. This indicates that 25 of 114 (21.9%) users engaged with the in-app tools to problem-solve one of their practical worries without completing the lesson. This behavior is permitted within the IMWW user experience because the user is able to choose the specific features of the app they wish to engage with. The majority of app features start in an unlocked state to encourage exploration and self-discovery. Only 42 of 153 (27%) users completed a follow-up conversation to review their solutions and progress using the Problem Solving protocol on their worries. During the interaction, engaged users recorded a total of 720 worries (mean 4.7, SD 6.1), and a median of 3
worries were recorded for each user. Of the worries recorded, 399 (55%) were categorized by the user as practical, 306 (43%) as hypothetical, and only 15 (2%) worries were not categorized. With respect to practical worries, 244 (61%) worries were problem-solved using the in-app tools with at least 1 possible solution added.

**Weekly Reflection**

In response to the CA asking the user to reflect on their experience of engaging with IMWW, in-app Weekly Reflections were completed by 58 of 153 (38%) users who recorded 206 responses (mean 3.6) identifying LICBT techniques engaged with, alongside 48 responses identifying the technique found most helpful (Table 2).

During the Weekly Reflection, 27 of 58 (47%) individual users reflected on engaging with Worry Time, of whom 19 (70%) were observed to have used the in-app tooling to complete it at their chosen time. A total of 8 of 58 (14%) users therefore engaged with Worry Time without using the in-app tools. In addition to asking which features of IMWW the user had engaged with, the CA also asked which feature they found most helpful. The users were asked this question during the weekly review, and for each weekly review, they could give at most 1 response.

**Table 2.** Weekly reflection techniques engaged with and found most helpful.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Engaged with (n=206), n (%)</th>
<th>Most helpful (n=48), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journaling worries</td>
<td>50 (24)</td>
<td>16 (33)</td>
</tr>
<tr>
<td>Worry time</td>
<td>46 (22)</td>
<td>11 (23)</td>
</tr>
<tr>
<td>Problem-solving</td>
<td>35 (17)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Avoiding worry behaviors</td>
<td>28 (14)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Watching out for different worry types (Worry categorization)</td>
<td>17 (8)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>CBTb modelc</td>
<td>8 (4)</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

*aUsers can respond multiple times.

bCBT: cognitive behavioral therapy.

cIntroduced in October 2022.

**Maintaining Engagement**

To maintain engagement with IMWW, 142 of 153 (93%) users were sent at least 1 push notification with 113 of 153 (74%) users responding. An average of 84 (SD 75) push notifications were sent to each engaged user throughout their engagement, although the quantity of push notifications per user varied substantially with use pattern and duration.

**Potential Effectiveness**

Separate paired sample 2-tailed $t$ tests were conducted to examine the difference between the GAD-7 and PHQ-8 scores for 54 of 153 (35%) engaged users who completed the measures during onboarding and the final score provided. There was a significant reduction in both anxiety ($t_{53}=5.5; P<.001; 95\%$ CI 2.4-5.2) and low mood ($t_{53}=2.3; P=.03; 95\%$ CI 0.2-3.3), with severity dropping from moderate to mild in both instances (Figure 1).

Examination of individual-level data indicates that the vast majority of users (43/53, 81%) experienced a reduction in anxiety between baseline and final observation with the score of 2 (4%) users remaining unchanged. The majority of users (35/53, 66%) also saw a reduction in PHQ-8 with no difference arising for 4 (8%) users. Deterioration in GAD-7 was experienced by 9 (17%) users and rose to 15 (28%) users for low mood.
Impact of Engagement on Potential Effectiveness

The multivariate linear regression predicting changes in GAD-7 based on engagement reached statistical significance ($F_{2,51}=6.73; P=.03$), but the model predicting changes in the PHQ-8 did not ($F_{2,51}=2.33; P=.11$). Two-sided 2-tailed t tests were performed on the slope estimates in the models. The model predicting changes in GAD-7 estimates that the marginal effect of a user completing in-app Worry Time ($\beta_2$) is a –3.3 change in GAD-7 and is significant ($P=.02$). The constant $\beta_0$ and the marginal effect of the user successfully completing Problem Solving $\beta_1$ had respective values of –1.6 and –3.0. However, they failed to reach significance at $\alpha=.05$ ($P_0=.07$, $P_1=.08$). The model had an $R^2$ of 0.21. Furthermore, an improvement in the GAD-7 and PHQ-8 was not predicted by the number of in-app sessions completed ($P=.09$) or the number of worries ($P=.36$), problems ($P=.27$), and solutions ($P=.16$) recorded.
**Discussion**

**Principal Findings**

While engaged users represented a minority of those who downloaded IMWW, a large number of these interacted with the LICBT techniques associated with the CBT protocol to manage excessive worry and support emotional well-being [31]. The vast majority of those who engaged completed all areas presented with the CBT model and were able to successfully record worries and categorize them as practical or hypothetical. The CA was commonly used to help engaged users overcome difficulties when engaging with practical worries. Worry Time was engaged with to a much lesser extent within the app; however, several users reported engaging with it outside of the app. Forgetting to engage with the LICBT techniques was identified as the most common reason for lack of engagement, while experiencing them as too difficult to comprehend was only reported by a small minority of engaged users. The change in the user GAD-7 score was predominantly explained by engagement with the LICBT techniques as opposed to the number of times they used IMWW.

Poor engagement with an app following download is not uncommon, with only 14% of people often using it the following day [50] and even lower rates typically associated with mental health apps [22]. Despite using common factors and behavior change techniques, however, only a minority of users who engaged with IMWW had enough engagement with the LICBT techniques to be considered engaged users. This is of some concern given that users failing to engage to a point where they have received an adequate dose to bring about recovery may serve as a barrier to seeking further support.

While engagement following download was poor, log data identified that engaged users had moderate to good levels of interaction and fidelity [51], with the CBT model alongside recording and categorizing worries. Fidelity and interaction with Problem Solving were also good, potentially arising from support provided through the CA. When engaging with Problem Solving, the CA was commonly used to support users to overcome difficulties in engaging with the LICBT techniques and to encourage continued engagement. There was less within-app engagement with Worry Time; however, some users reported engaging with it outside of the app. Engagement with the LICBT techniques included within IMWW may therefore have been greater than log data alone suggest. This supports the additional benefits of exploring out-of-app engagement with specific techniques to get a full appreciation of interaction [28]. Exploring ways to promote out-of-app engagement is of benefit given that engagement with techniques in face-to-face CBT between support sessions as “homework” is identified as important to improve clinical outcomes related to anxiety [52].

Although there were moderate levels of interaction with LICBT techniques used within IMWW, exploring additional ways to enhance interaction across all techniques and promote prolonged engagement would be highly beneficial. Enhancing engagement through approaches such as involving personalized support, guidance, and feedback regarding engagement has also been associated with improved effectiveness for mental well-being digital tools [48]. Furthermore, recommendations to enhance out-of-app homework compliance to deliver better outcomes have also been proposed [53]. These include ensuring that app content is congruent to the therapeutic approach adopted, learning is consolidated through engagement, and emphasis is placed on completion. Additionally, recommendations include ensuring that the app is tailored to specific populations and building connections with others has been identified as supporting engagement with homework [53]. Within IMWW the CA was used to enhance engagement through the use of common factor skills to encourage and motivate the user. However, greater focus needs to be directed toward maximizing the ability of the CA to enhance engagement within and outside of the app.

Maximizing engagement may be achieved by implementing mental well-being apps for use adjuvant to health professional support and integrated into clinical settings [54]. Benefits associated with providing support are recognized by the National Institute of Health and Care Excellence recommendations for supported LICBT for anxiety and depression [14]. This has resulted in Psychological Practitioner support for LICBT adopted by the IAPT program implemented across England [14]. Support enables the patient to engage with the interventions by using personalized common factor skills, monitor progress, and provide encouragement during weekly support sessions. However, it does not include a therapeutic role in the delivery of LICBT techniques within the clinical sessions [13].

However, nonprofessional forms of support have also been demonstrated to enhance engagement and improve outcomes with LICBT. For example, group support within community settings is provided by trained volunteers with varying backgrounds [55]. Furthermore, forms of support through technology such as web-based communities providing constructive peer support [55] and discussion forums [56] have been identified to enhance engagement with digital tools [54]. Potentially, therefore, using IMWW adjuvant to some form of minimal-contact support provided by a practitioner, volunteers within community organizations, or mediated through technology offers promise to result in enhanced effectiveness at reduced delivery costs.

With respect to outcomes, the average level of anxiety and low mood improved among users who engaged with IMWW to a degree they would be considered to have received a minimum dose of the LICBT techniques [29]. That anxiety and low mood are identified to share mechanisms has led to recommendations to combine techniques within a single app to reduce the commitment needed by users to maximize engagement [19]. However, when exploring recovery at the level of the individual user, the low mood of several more users deteriorated compared with anxiety. However, it would remain possible to develop a single app that included protocol-informed LICBT techniques to target low mood or anxiety once the main emotional difficulty being experienced was determined.

**Strengths and Limitations**

Providing a clear description of the LICBT techniques contained within IMWW informed by a theoretical basis represents a real strength of the paper. This has enabled the analysis of log data
to be interpreted with respect to interactions with the techniques. Clearer conclusions regarding the relationship between engagement and outcomes regarding the management of symptoms associated with anxiety were therefore able to be reached. This facilitates specific targeting of future development work on IMWW to ensure greater levels of engagement to derive improved outcomes.

There was a large difference between the number of people who downloaded IMWW and those who interacted with at least 1 lesson related to Worry Time and Problem Solving for them to be considered engaged users. While it is known that the background demographics of these 2 groups did not significantly differ, it is unclear as to why a large number of those who downloaded IMWW never went on to engage with one of these specific LICBT techniques. Unfortunately, reasons behind failing to engage with IMWW were not requested, and therefore the extent to which poor usability may have been a relevant factor is unknown. As the use of digital health technologies continues to increase [16], understanding the usability of apps is of increasing interest [57]. Future research exploring log data could therefore consider using a measure of usability, such as the mHealth Usability Questionnaire [58], alongside the collection of log data to gain a better understanding of the way in which an app is used as well as potential barriers to usability.

Finally, while data regarding outcomes can be considered promising with respect to IMWW as a tool to support worry management, this study does not enable definitive conclusions regarding effectiveness to be reached. As a consequence of the lack of clear end points when using log data, users can stop using the app at any time without completing outcome measures. This makes it difficult to reach conclusions regarding effectiveness. The use of multivariate regression with terms to represent proxy use of techniques was adopted to compensate for this. However, this cannot be considered to represent a substitute for the collection of clear and reliable data within a trial design comparing IMWW with an appropriate control [59]. Furthermore, reaching conclusions regarding effectiveness is further confounded given that only a minority of engaged users responded to a question regarding the current receipt of treatment.

**Conclusions**

While a large number of people downloaded IMWW, only a minority engaged with the app to be considered engaged users. Of these users, however, analysis of log data identified good interaction with the LICBT techniques associated with an evidence-based protocol to support worry management [31]. Although there were good levels of interaction, exploring additional ways to promote interaction with the LICBT techniques and other features of the app to result in prolonged engagement remains beneficial. This could involve adopting a “user-centric” design process whereby potential users are directly involved in ongoing development [22]. Considering log data as part of a user-centric design process may enhance engagement to a point where more users receive an appropriate “dose” to bring about improvement [29]. Log data can therefore be used to inform ongoing development to maximize engagement and protocol fidelity [51]. This is significant given the relationship between engagement and effectiveness. While effectiveness data associated with IMWW can only be seen as promising, capturing log data will serve to enhance ongoing intervention development. A high-quality randomized controlled trial would then enable definitive conclusions regarding effectiveness to be reached [54]. This would help address concerns that the current level of evidence derived from poor-quality trials does not enable recommendations regarding apps to enhance mental well-being to be reached [60].

**Authors’ Contributions**

PF, PA, and PJR conceptualized and designed the project with EW and DR providing theoretical input informing the background of the paper. PF wrote the initial draft of the paper with all authors contributing to the development of the paper, interpretation of the analysis, and editing of the final manuscript. PF was not involved in any part of data analysis.

**Conflicts of Interest**

PF is currently on a paid sabbatical with Iona Mind from the University of Exeter. All other authors have confirmed they have no conflicts of interest to declare.

**Multimedia Appendix 1**
Progress through the Iona Mind Well-being app for Worry management.
[DOCX File, 951 KB - mhealth_v12i1e47321_app1.docx ]

**Multimedia Appendix 2**
Number of sessions of varying length.
[PNG File, 60 KB - mhealth_v12i1e47321_app2.png ]

**References**


17. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019. URL: http://tinyurl.com/3erjjbjt [accessed 2023-12-01]


Abbreviations

CA: conversational agent
CBT: cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder–7
IAPT: improving access to psychological therapies
IMWW: Iona Mind Well-being app for Worry management
LICBT: low-intensity cognitive behavioral therapy
PHQ-8: Patient Health Questionnaire–8

©Paul Farrand, Patrick J Raue, Earlise Ward, Dean Repper, Patricia Areán. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
A Behaviorally Informed Mobile App to Improve the Nutritional Quality of Grocery Shopping (SwapSHOP): Feasibility Randomized Controlled Trial

Carmen Piernas1,2, MSc, PhD; Charlotte Lee1, MSc, DPhil; Alice Hobson1, MSc; Georgina Harmer1, MSc; Sarah Payne Riches1, MPH, DPhil; Michaela Noreik1,3, MSc, PhD; Susan A Jebb1, PhD

1Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
2Department of Biochemistry and Molecular Biology II, Center for Biomedical Research (CIBM), Institute of Nutrition and Food Technology (INYTA), University of Granada, Granada, Spain
3Faculty of Food and Nutrition Sciences, Hochschule Niederrhein, Mönchengladbach, Germany

Abstract

Background: Interventions targeting the nutritional quality of grocery shopping have the potential to help improve diet and health outcomes.

Objective: This study aims to assess the feasibility and acceptability of receiving advice on healthier food purchases through SwapSHOP, a behaviorally informed smartphone app that allows users to scan barcodes of grocery products from the United Kingdom, providing nutritional information and personalized swap suggestions to encourage healthier purchases.

Methods: We randomized adult volunteers in a 6-arm parallel-group controlled feasibility trial. Participants used the SwapSHOP app to record their grocery shopping during a 2-week run-in period and were individually randomized in a 3:1 ratio to either intervention or control arms within 3 strata related to a nutrient of concern of their choice: saturated fat (SFA), sugar, or salt. Participants randomized to the intervention received the SwapSHOP app with a healthier swap function, goal setting, and personalized feedback. Participants in the control group were instructed to use a simpler version of the app to log all their food purchases without receiving any guidance or advice. The primary outcome was the feasibility of progression to a full trial, including app use and follow-up rates at 6 weeks. The secondary outcomes included other feasibility outcomes, process and qualitative measures, and exploratory effectiveness outcomes to assess changes in the nutrient content of the purchased foods.

Results: A total of 112 participants were randomized into 3 groups: SFA (n=38 intervention and n=13 control), sugar (n=40 intervention and n=15 control), and salt (n=5 intervention and n=1 control, not analyzed). The 2 progression criteria were met for SFA and sugar: 81% (30/37) and 87% (34/39) of intervention participants in the SFA and sugar groups, respectively, used the app to obtain healthier swaps, and 89% (68/76) of intervention participants and 96% (23/24) of control participants completed follow-up by scanning all purchases over the follow-up period. The process and qualitative outcomes suggested that the intervention was acceptable and has the potential to influence shopping behaviors. There were reductions of −0.56 g per 100 g (95% CI −1.02 to −0.19) in SFA and −1 g per 100 g (95% CI −1.97 to −0.03) in total sugars across all food purchases in the intervention groups.

Conclusions: People were willing to use the SwapSHOP app to help reduce sugar and SFA (but not salt) in their grocery shopping. Adherence and follow-up rates suggest that a full trial is feasible. Given the suggestive evidence indicating that the intervention resulted in reductions in sugars and SFA, a definitive trial is necessary to target improvements in health outcomes.

Trial Registration: International Standard Randomised Controlled Trial Number ISRCTN13022312; https://doi.org/10.1186/ISRCTN13022312

doi:10.2196/45854
Introduction

Background

Consumption of saturated fat (SFA), sugars, and salt in the United Kingdom and most high-income countries currently exceeds dietary recommendations for good health [1]. These nutrients of concern contribute to the burden of diabetes and cardiovascular disease, either directly or through their effects on blood cholesterol, blood pressure, insulin sensitivity, and body weight [2-8]. Despite decades of nutrition education and efforts to promote healthier behaviors, significant and sustained dietary changes are yet to be observed. Evidence suggests that people need more than general knowledge of dietary recommendations to change behavior [9-11]; however, there is limited evidence for individual-level interventions that are feasible and effective in supporting dietary change at the population level.

Food purchasing is a key antecedent of food consumption, and improving the nutritional quality of food purchases presents a clear opportunity to intervene. Grocery stores account for 71% of the weekly expenditure on food and drinks, including a substantial proportion of foods high in SFA, sugar, and salt [12]. For many foods, there are usually alternatives with less SFA, sugar, or salt, which are functionally similar. This variation can be attributed partly to the natural differences in recipes, such as those found in ready meals or variances in ingredient costs. In addition, it can also be linked to product reformulation driven by consumer demand, such as the introduction of low-fat yogurts, or policy-driven actions, such as changes in soft drink formulations [13].

Systematic reviews have identified some intervention components to support individual dietary change, including providing information, tailored dietary advice, self-monitoring, and personalized feedback [9,14,15]. Specifically, recommending healthier swaps at the point of purchase or as part of individually tailored regular feedback has shown potential to help improve the nutritional quality of grocery shopping [16-20]. Technological advances such as smartphone apps can facilitate this by providing scalable, lower-cost support to help people make healthier choices while shopping.

A systematic review of smartphone apps identified only 2 studies that tested apps that can provide healthier alternatives at the point of choice [21]. Although there is evidence of similar apps in the scientific literature, these previous apps mostly provide healthier swaps to consumers without the option of self-monitoring the quality of their grocery shopping, and many apps have limitations regarding how information is presented to consumers or the limited comprehensiveness of the databases covering the UK food market [22,23]. Our recent proof-of-concept study tested the functionality of the SaltSwap app to help find lower-salt foods when grocery shopping, which included behavioral components such as nutritional information, prompts to lower-salt options, goal setting on swaps, feedback on swaps and salt reduction, and history of swaps [20].

Objectives

For this study, we have further developed the SaltSwap app into SwapSHOP, a new app that also provides swaps for SFA, sugars, and salt. This study aimed to assess the feasibility and acceptability of receiving dietary advice through the SwapSHOP app. It is a stand-alone intervention that allowed users to scan barcodes of grocery products from major UK stores to obtain nutritional information and suggestions for personalized swaps with lower SFA, sugar, or salt and also enabled users to set goals and monitor their shopping behavior.

Methods

Study Design

This was a prospectively registered (International Standard Randomised Controlled Trial Number: ISRCTN13022312) randomized 6-arm parallel-group controlled feasibility trial conducted in the United Kingdom. After giving written informed consent to participate in the study and completing screening and baseline assessment, participants entered a 2-week run-in period where they used a basic version of the SwapSHOP app (no swap or behavioral functionality shown) to record their grocery shopping. They were individually randomized to 1 of the intervention arms or control following a 3:1 ratio (intervention: control) within 3 strata related to a nutrient of concern of their choice (SFA, sugar, or salt). Individuals participated in the study for 6 weeks from screening to the final follow-up.

Ethical Considerations

This study was reviewed and approved by the University of Oxford Medical Sciences Interdivisional Research Ethics Committee (R67216/RE001). Informed consent was obtained from all the participants.

Participants and Procedures

Participants were recruited between April and October 2021 through the community (eg, word of mouth) and through social networks and media, online newsletters and newspapers, and electronic mailing lists. Participants were eligible if they were aged ≥18 years, were willing and able to give informed consent, were English speaking and able to understand the demands of the study, were looking for support to improve their diet, owned a smartphone with access to the internet and an email account, were willing to download and use a smartphone app to scan and track their grocery shopping for the 6 weeks during their participation in the study, were responsible for at least some of their household grocery shopping, and shopped regularly (eg, at least once a week in a physical store or online at any of the 6 major food retailers in the United Kingdom: Tesco, Sainsbury’s, Waitrose, Asda, Morrisons, or Iceland). People were not eligible if they were already on a clinician-supervised diet or a restricted diet, were currently using apps to monitor the quality of their food shopping (excluding apps to track and monitor food intake), were currently participating in another study aimed at dietary change or asking them to change the way...
they shop for food, or were planning on going away from home (holiday or other) for ≥2 consecutive weeks following enrollment in the study.

Eligible participants were invited to complete a baseline web-based questionnaire that collected demographic information, relevant self-reported medical history, and details about their shopping behaviors. Participants also reported the nutrient (SFA, sugar, or salt) they were most concerned about in their diet, which was used as a stratification factor during the randomization process. Participants then entered a 2-week run-in period during which they were asked to record all their food purchases, either by scanning the barcode of purchased grocery products or manually inputting what was purchased using a simpler version of the SwapSHOP app. Only participants who completed this task with good engagement with the app were randomized. Good engagement was defined as scanning products in at least 2 shops, each with products from at least 3 different predefined food categories (eg, fresh meat, chilled ready meals, and soft drinks), and scanning products from ≥5 of the food categories across the 2 weeks.

**Randomization and Blinding**

After the 2-week run-in period, participants were individually randomized to either an intervention arm or control following a 3:1 ratio (intervention:control) within each stratum depending on their chosen nutrient of concern (sugar, salt, or SFA). Participants who did not complete the baseline data collection were not randomized. Randomization was performed with a computer using a random allocation sequence (concealed to the investigators) and was stratified by the nutrient of concern using block randomization with block sizes of 4 and 8.

It was not possible to blind the participants to the intervention because of the nature of the intervention; however, randomization was performed remotely via computer-generated randomization, and the researchers were not aware of the treatment group until after randomization was complete.

**Intervention**

Participants randomized to the intervention arms were able to access an enhanced version of the SwapSHOP app, which included a healthier swap function, goal setting, and personalized feedback. Participants scanned the barcode of grocery products to receive nutritional information about the product using the UK traffic light system and were presented with healthier alternatives or swaps that were lower in SFA, sugar, or salt (depending on group allocation). The swaps that appeared after scanning the original product were chosen from a list of products within the same food category, which were ranked from a larger to a smaller reduction in the specific nutrient that was initially chosen. Participants could also set goals for the number of swaps they wanted to make in each shopping trip and record the products they swapped in the app. The app had specific functions to provide feedback on the achievement of their goals and feedback on the overall nutrient reduction achieved by making swaps. The app also had a section to display the ranking of all purchased foods contributing the most to each nutrient of concern (Figure S1 in Multimedia Appendix 1).

The SwapSHOP app included a comprehensive database of >70,000 grocery products available in January 2021 within major UK grocery stores (including Morrisons, Sainsbury’s, Tesco, Waitrose, Iceland, and Asda) [24] and a bespoke system for categorizing products and selecting suitable alternative swap suggestions. SwapSHOP was based on a previous version developed exclusively for salt, and its theoretical basis is grounded in the Behavior Change Wheel [20,25], a framework that integrates the elements of capability, opportunity, and motivation, which are key for behavior change. A range of behavior change techniques from the taxonomy groups goals and planning and feedback and monitoring that were incorporated into this intervention have been associated with successful dietary change [9,26].

To enable assessment of changes in the nutritional composition of the shopping baskets, participants received weekly reminders to continue scanning and recording all their purchased products for the remaining 4 weeks of follow-up (with a minimum of 2 full weeks of grocery shopping data during the follow-up period).

**Comparator**

Participants randomized to the control arm were asked to continue using the simpler version of the app with no product nutrition information, swaps, or behavioral components, solely to record all their food purchases as part of the outcome assessment at the end of the 6-week follow-up.

**Outcome Measures**

**Primary Outcomes**

The primary objective of this study was to determine the feasibility of a larger study to evaluate the effectiveness of an app that offers healthier swaps while grocery shopping to improve the nutritional quality of food purchasing with respect to sugar, salt, or SFA. The primary outcomes were prespecified progression criteria as follows:

1. App use: at least 70% of participants in the active intervention group use the app to obtain swaps on at least 1 occasion by the end of the second week after randomization.
2. Follow-up rate: at least 60% of participants in total (intervention and control) complete follow-up by scanning all their purchases for a minimum of 2 weeks over the entire follow-up period (4 weeks).

**Secondary Outcomes**

Feasibility outcomes included (1) recruitment rates: total recruited (including number signed up, eligible, consented, and randomized), (2) time needed for recruitment of the final sample, (3) outcome reporting: number of participants who failed to scan their purchases for a minimum of 2 weeks during follow-up, and number of participants who failed to complete the end-of-study questionnaires.

**Process Evaluation and Qualitative Outcome Measures**

A summary of app-related use (within-app automatic recording) and acceptability measures was collected through the end-of-study questionnaires at follow-up: (1) average number
of shopping trips where the app was used to scan products to obtain a swap per week; (2) number of occasions the app was used to scan products for a swap per trip; (3) number of swaps made overall per week and per shopping trip; (4) nutrient (SFA, sugar, and salt) reduction per swap made; (5) use of specific functionality, for example, goal setting and feedback; (6) end-of-study questionnaires with free text to understand app use and acceptability of the swaps, app functionality (eg, if app scans most products), and if this prompted other behaviors such as reading nutrition labels.

**Exploratory Effectiveness Outcomes**

Measures included changes in the nutrient content (SFA and sugar in grams per 100 g) of household food purchases recorded in the app between baseline and follow-up in the intervention group compared with the control group.

**Statistical Analysis and Sample Size**

This study was planned as a feasibility study. A sample size of 120 (n=approximately 30 participants per intervention group and n=approximately 10 participants per control group) was prespecified as sufficient to enable testing of the trial methodology and outcome measures using parametric statistical models. The protocol was prospectively registered (International Standard Randomised Controlled Trial Number: ISRCTN13022312).

Baseline characteristics overall and by trial arm were summarized as means and SDs for continuous variables and counts and percentages for categorical variables. Baseline characteristics were coded as age (years), gender (man, woman, other, or not specified), ethnic group (Asian, Black, White, mixed, or other), education (no formal qualifications, secondary education, or higher education), income (>£15,000, £15,000-£24,999, £25,000-£39,999, £40,000-£75,000, and >£75,000; GBP £1=US $1.38), household size (1, 2-4, and ≥5), grocery shopping habits (eg, spending >£25) on groceries more than once a week, once a week, once a fortnight, once a month, and less than once a month), and relevant self-reported health history (eg, concerns related to weight, high blood pressure, diabetes, and heart disease).

Descriptive statistics were used to present the primary and secondary outcomes using all available data, regardless of whether the participants completed the trial or withdrew.

For the exploratory effectiveness measures, we used data from products purchased over the 2 weekly periods recorded at the beginning and end of the trial, with available information on weight or volume as well as the nutrient content (grams per 100 g) to estimate changes in SFA and sugar for food purchases recorded in the app. The excluded items in this analysis included those categorized as fresh fruits, vegetables, or with no or minimal nutrient content (eg, sugar-free gum) as well as products that were manually entered in the app or those missing nutrient or volume information. We used linear regression models to investigate changes in the nutrient content of food purchases between baseline and follow-up in the intervention and control groups. Tests for linear regression assumptions were conducted and met. The main models included adjustment for baseline values of the dependent variable. Potential confounding because of imbalance in baseline characteristics was investigated, and the following variables were adjusted in sensitivity analyses: age (years), gender (man, woman, other, or not specified), ethnic group (Asian, Black, White, mixed, and other), and income (>£15,000, £15,000-£24,999, £25,000-£39,999, £40,000-£75,000, and >£75,000). StataSE (version 16; StataCorp) was used for all the analyses. A P value of <.05 was set to denote statistical significance.

A descriptive analysis of the qualitative outcome measures was conducted using the method by Braun and Clarke [27] for thematic analysis in NVivo 1 software (Lumivero). Each response was line-by-line coded, and codes were inductively constructed based on the aim of the study. We then organized codes into subthemes using the One Sheet of Paper technique and produced top-level themes [28].

**Patient and Public Involvement**

The SwapSHOP app was based on a previous version developed exclusively for salt reduction [20], which was conceived and tested using input from a patient and public involvement panel. People told us that they would also value an app to help change other aspects of their diet. The current app was also beta tested by members of the public recruited from the community.

**Results**

**Participant Characteristics**

Study participants were recruited between April and October 2021. Of the 289 interested participants, 190 (65.7%) were eligible, provided consent, and entered the 2-week run-in period. Of these 190 participants, 78 (41%) were excluded because they did not download or use the app (n=49, 63%) or did not fulfill the progression criteria based on their engagement with the app (n=29, 37%). The final sample of 38.8% (112/289) of participants successfully completed the task and were randomized (Figure 1). A total of 51 participants were randomized to the SFA group (n=38 to intervention and n=13 to control), 55 participants were randomized to the sugar group (n=40 to intervention and n=15 to control), and 6 participants were randomized to the salt group (n=5 to intervention and n=1 to control). Of the randomized participants, 100 (89%) completed the study and were analyzed as follows: 49 (96%) in the SFA group and 51 (93%) in the sugar group. A total of 2 participants in the SFA group and 4 participants in the sugar group were withdrawn from the study and excluded from the analysis because their data indicated fraudulent activity (eg, fake phone numbers and implausible shopping patterns). Data from 6 participants who were randomized to the salt group were not analyzed, as this group did not reach the target sample by the end of the recruitment period.
Participants in the analytic sample were, on average, aged 35 (SD 12) years, mostly women (80/100, 80%) of White ethnicity (79/100, 79%), and from higher education backgrounds (73/100, 73%; Table 1). Participants mostly lived in households with 2 to 4 other members, shopped once (61/100, 61%) or more than once a week (26/100, 26%), shopped mostly in larger grocery stores (89/100, 98%), or shopped on the internet (39/100, 39%). Most participants looked at SFA or sugar information on nutrition labels always, often, or sometimes and approximately half of the sample (54/100, 54%) reported concerns related to weight among the relevant medical conditions.
Table 1. Baseline characteristics of participants in the analytic sample.

<table>
<thead>
<tr>
<th></th>
<th>Total population (n=100)</th>
<th>Sugar group</th>
<th>Saturated fat group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (n=12)</td>
<td>Intervention (n=39)</td>
<td>Control (n=12)</td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td>35 (12)</td>
<td>36 (8)</td>
<td>35 (10)</td>
</tr>
<tr>
<td>Gender (woman), n (%)</td>
<td>80 (80)</td>
<td>9 (75)</td>
<td>33 (85)</td>
</tr>
<tr>
<td>Ethnic group, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or Asian</td>
<td>9 (9)</td>
<td>1 (8)</td>
<td>6 (15)</td>
</tr>
<tr>
<td>White</td>
<td>79 (79)</td>
<td>10 (83)</td>
<td>27 (69)</td>
</tr>
<tr>
<td>Mixed, other, or not specified</td>
<td>12 (12)</td>
<td>1 (8)</td>
<td>6 (15)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No qualifications or not specified</td>
<td>2 (2)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Secondary education</td>
<td>25 (25)</td>
<td>0 (0)</td>
<td>12 (31)</td>
</tr>
<tr>
<td>Higher education</td>
<td>73 (73)</td>
<td>12 (100)</td>
<td>27 (69)</td>
</tr>
<tr>
<td>Individual income a, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;£15,000</td>
<td>28 (28)</td>
<td>3 (25)</td>
<td>9 (23)</td>
</tr>
<tr>
<td>£15,000-£24,999</td>
<td>24 (24)</td>
<td>4 (33)</td>
<td>10 (26)</td>
</tr>
<tr>
<td>£25,000-£39,999</td>
<td>31 (31)</td>
<td>4 (33)</td>
<td>10 (26)</td>
</tr>
<tr>
<td>£40,000-£75,000</td>
<td>11 (11)</td>
<td>1 (8)</td>
<td>6 (15)</td>
</tr>
<tr>
<td>&gt;£75,000</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Household size, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>29 (29)</td>
<td>2 (17)</td>
<td>11 (28)</td>
</tr>
<tr>
<td>2-4</td>
<td>64 (64)</td>
<td>8 (67)</td>
<td>24 (62)</td>
</tr>
<tr>
<td>≥5</td>
<td>7 (7)</td>
<td>2 (17)</td>
<td>4 (10)</td>
</tr>
<tr>
<td>Frequency of grocery shopping (≥£25 per trip), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than once a week</td>
<td>26 (26)</td>
<td>7 (58)</td>
<td>8 (21)</td>
</tr>
<tr>
<td>Once a week</td>
<td>61 (61)</td>
<td>4 (33)</td>
<td>26 (67)</td>
</tr>
<tr>
<td>Once a fortnight</td>
<td>11 (11)</td>
<td>1 (8)</td>
<td>3 (8)</td>
</tr>
<tr>
<td>Once a month</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Less than once a month</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Type of grocery shop usually visited, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supermarkets</td>
<td>98 (98)</td>
<td>12 (100)</td>
<td>38 (97)</td>
</tr>
<tr>
<td>Online supermarkets</td>
<td>30 (39)</td>
<td>5 (42)</td>
<td>11 (28)</td>
</tr>
<tr>
<td>Corner shop or convenience store</td>
<td>32 (32)</td>
<td>1 (8)</td>
<td>14 (36)</td>
</tr>
<tr>
<td>Greengrocers or fruit and vegetable shop</td>
<td>12 (12)</td>
<td>3 (25)</td>
<td>5 (13)</td>
</tr>
<tr>
<td>Butchers or meat market</td>
<td>11 (11)</td>
<td>0 (0)</td>
<td>6 (15)</td>
</tr>
<tr>
<td>Other fresh food markets</td>
<td>17 (17)</td>
<td>4 (33)</td>
<td>7 (18)</td>
</tr>
<tr>
<td>Looking at salt in nutrition labels, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always or often</td>
<td>14 (14)</td>
<td>2 (17)</td>
<td>3 (8)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>24 (24)</td>
<td>3 (25)</td>
<td>12 (31)</td>
</tr>
<tr>
<td>Rarely or never</td>
<td>62 (62)</td>
<td>7 (58)</td>
<td>24 (62)</td>
</tr>
<tr>
<td>Looking at sugar in nutrition labels, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always or often</td>
<td>42 (42)</td>
<td>2 (17)</td>
<td>19 (49)</td>
</tr>
</tbody>
</table>
For participants randomized to the SFA and sugar groups, the 2 progression criteria were met above the set thresholds (Table 2). Overall, most intervention participants (30/37, 81% in the SFA group and 34/39, 87% in the sugar group) used the app to obtain swaps on at least 1 occasion by the end of the second week after randomization. In addition, 89% (68/76) of the intervention participants and 96% (23/24) of the control participants completed follow-up by scanning all purchases for a minimum of 2 weeks over the entire follow-up period.

### Primary Outcomes

The number of participants recruited for the salt group did not reach the target sample size despite additional recruitment attempts, suggesting that a full-scale trial for salt reduction would not be feasible. Data from the participants in this group were not analyzed further.

---

<table>
<thead>
<tr>
<th></th>
<th>Total population (n=100)</th>
<th>Sugar group</th>
<th>Saturated fat group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (n=12)</td>
<td>Intervention (n=39)</td>
<td>Control (n=12)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>30 (30)</td>
<td>6 (50)</td>
<td>10 (26)</td>
</tr>
<tr>
<td>Rarely or never</td>
<td>28 (28)</td>
<td>4 (33)</td>
<td>10 (26)</td>
</tr>
</tbody>
</table>

**Looking at fat in nutrition labels, n (%)**

- Always or often: 29 (29) 1 (8) 6 (15) 8 (67) 14 (38)
- Sometimes: 44 (44) 7 (58) 21 (54) 1 (8) 15 (41)
- Rarely or never: 27 (27) 4 (33) 12 (31) 3 (25) 8 (22)

**Relevant health conditions, n (%)**

- Concerns related to weight: 54 (54) 7 (58) 18 (46) 5 (42) 24 (65)
- High blood pressure: 5 (5) 0 (0) 2 (5) 0 (0) 3 (8)
- Diabetes: 11 (11) 2 (17) 4 (10) 1 (8) 4 (11)
- Heart disease: 4 (4) 0 (0) 1 (3) 1 (8) 2 (5)

**Relevant current medications, n (%)**

- High blood pressure: 4 (4) 0 (0) 1 (3) 0 (0) 3 (8)
- Diabetes: 6 (6) 0 (0) 2 (5) 1 (8) 3 (8)
- Heart disease: 1 (1) 0 (0) 0 (0) 1 (8) 0 (0)

---

a The number of participants recruited for the salt group did not reach the target sample size. Data from the participants in this group were not analyzed further.

b GBP £1=US $1.38.
Table 2. Primary and secondary outcomes—progression criteria and feasibility outcomes\(^a,b\).

<table>
<thead>
<tr>
<th></th>
<th>Control (n=24)</th>
<th>Intervention (n=76)</th>
<th>Control (n=12)</th>
<th>Intervention (n=39)</th>
<th>Control (n=12)</th>
<th>Intervention (n=37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population, n (%)</td>
<td>N/A</td>
<td>64 (84)</td>
<td>N/A</td>
<td>34 (87)</td>
<td>N/A</td>
<td>30 (81)</td>
</tr>
<tr>
<td>Saturated fat group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control (n=12)</td>
<td>23 (96)</td>
<td></td>
<td>11 (92)</td>
<td>36 (92)</td>
<td>12 (100)</td>
<td>32 (86)</td>
</tr>
<tr>
<td>Sugar group</td>
<td>289 (100)</td>
<td>N/A</td>
<td>13 (100)</td>
<td>40 (100)</td>
<td>13 (100)</td>
<td>38 (100)</td>
</tr>
<tr>
<td>Total (N=100)</td>
<td>226 (78)</td>
<td>N/A</td>
<td>197 (68)</td>
<td>141 (49)</td>
<td>106 (100)</td>
<td>16 (15)</td>
</tr>
<tr>
<td>Consented participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Failed to complete</td>
<td>0 (0)</td>
<td>16 (21)</td>
<td>0 (0)</td>
<td>9 (23)</td>
<td>0 (0)</td>
<td>7 (18)</td>
</tr>
<tr>
<td>baseline assessments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>78 (100)</td>
<td>15 (100)</td>
<td>40 (100)</td>
<td>13 (100)</td>
<td>38 (100)</td>
</tr>
<tr>
<td>Signed-up participants</td>
<td>256 (89)</td>
<td>N/A</td>
<td>17 (100)</td>
<td>29 (100)</td>
<td>17 (100)</td>
<td>24 (100)</td>
</tr>
<tr>
<td>Eligible participants</td>
<td>226 (78)</td>
<td></td>
<td>197 (68)</td>
<td>141 (49)</td>
<td>106 (100)</td>
<td>16 (15)</td>
</tr>
<tr>
<td>Located participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>78 (100)</td>
<td>15 (100)</td>
<td>40 (100)</td>
<td>13 (100)</td>
<td>38 (100)</td>
</tr>
</tbody>
</table>

\(^a\)Percentage of eligible, consented, and completed participants was calculated based on the number of signed-up participants originally; the number of randomized participants (total population) excluded those randomized to the salt group, as these were not analyzed.

\(^b\)The number of participants recruited for the salt group did not reach the target sample size. Data from the participants in this group were not analyzed further.

\(^c\)N/A: not applicable.

**Secondary Outcomes**

The secondary outcomes provided evidence that a larger trial aiming at reducing sugars or SFA would recruit enough participants that adhere to the trial procedures (ie, completing baseline assessments) within the set time frames and that 84% (16/100 failed to complete follow-up) of those randomized would complete follow-up assessments.

**Process Evaluation and Qualitative Outcomes**

Participants randomized to the SwapSHOP intervention used the app regularly (Table 3). The average number of shopping trips where the app was used to obtain a swap was 5 in the SFA group (92% of shopping trips) and 5.4 in the sugar group (83% of shopping trips). The average number of occasions when the app was used to scan products for a swap in each shopping trip was 2.5 times in the SFA group and 3.3 in the sugar group. Overall, participants set goals averaging approximately 2 swaps per shopping trip. The results showed that individual product swaps were associated with an average reduction in total sugars (−12.5 g of sugar per 100 g, SD 6.97) or in SFA (−4 g of SFA per 100 g, SD 2.16) in the sugar and SFA groups, respectively.
Table 3. Secondary outcomes—process evaluation measures among participants using the swap function\(^a,b\).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sugar group (n=34), mean (SD)</th>
<th>Saturated fat group (n=30), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of shopping trips where the app was used to</td>
<td>5.41 (5.14)</td>
<td>5.07 (3.39)</td>
</tr>
<tr>
<td>obtain a swap</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of total shopping trips where the app was used to</td>
<td>83.15 (24.74)</td>
<td>91.62 (18.83)</td>
</tr>
<tr>
<td>obtain a swap</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occasions the app was used to scan products for a swap per</td>
<td>3.31 (2.30)</td>
<td>2.49 (1.94)</td>
</tr>
<tr>
<td>shopping trip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average swap goals set in the app per shopping trip</td>
<td>2.14 (1.15)</td>
<td>2.01 (1.54)</td>
</tr>
<tr>
<td>Sugar reduction per swap (grams per 100 g)</td>
<td>−12.47 (6.97)</td>
<td>−0.48 (6.70)</td>
</tr>
<tr>
<td>Saturated fat reduction per swap (grams per 100 g)</td>
<td>−1.62 (2.57)</td>
<td>−4.02 (2.16)</td>
</tr>
</tbody>
</table>

\(^a\)Sugar and SFA reduction per swap per 100 g were calculated as the average change in nutrient per 100 g of product across all swaps made during the intervention period.

\(^b\)The number of participants recruited for the salt group did not reach the target sample size. Data from the participants in this group were not analyzed further.

The results of the qualitative research relating to the acceptability of the app, the feedback and swaps provided through the app, and the usefulness and comprehensiveness of the app were summarized into 4 themes (Table S1 in Multimedia Appendix 1). Overall, most participants shared positive experiences of the intervention, noting that the app was helpful and the swap suggestions were acceptable. They valued the novelty of the app and the traffic light food labeling system, which encouraged them to read food labels. The key barriers to usability were that some grocery stores were not supported by the app and that there was poor product coverage within the app database. Although participants could manually enter their food purchases to find alternatives, they reported that this was time consuming and a barrier to engagement. Some participants also found it difficult to locate the suggested swaps in the store or were frustrated that the suggested swaps were not available in the store or that the product database in the app did not include own-label products in some stores. One significant obstacle to accepting the intervention was the specificity of certain swap suggestions. These suggestions were not always direct replacements or tailored to personal or household dietary preferences, which made it challenging for shoppers to act upon the prompts. In other instances, some participants found it challenging to reduce one nutritional component if it involved an increase in another, and they expressed a desire for a swap suggestion that recognized the overall healthiness of the product. Participants suggested that more information on the fiber composition of swaps, recommended portion sizes, and price comparison would inform their purchasing behavior. Most participants noted that other visual self-monitoring techniques would improve engagement with the intervention.

**Exploratory Effectiveness Outcomes**

Baseline and follow-up data on food purchases with available information on volume and nutrient content were available for 86% (44/51) of the participants in the sugar group and 82% (40/49) of the participants in the SFA group. This analysis included all food purchases recorded in the app but excluded purchases that were entered manually (22%) or those with no, minimal, or missing nutritional information; 16%). The degree of missingness in the food purchasing data was comparable between the groups (manually entered products: 23% intervention vs 19% control group and missing nutrient information: 15% intervention vs 16% control group).

There was evidence of changes in the intended direction in both the intervention groups (Table 4). The sugar group reduced total sugars in their grocery purchases by −1 (95% CI −1.97 to −0.03) g/100 g, whereas the control group reduced total sugars in their grocery purchases by 0.32 (−1.47 to 2.11) g/100 g, though the differences between groups were not significant (−0.68, 95% CI −1.94 to 0.58 g/100 g; P=.28 adjusted for baseline values). The SFA group reduced total SFA in food purchases by −0.56 (95% CI −1.02 to −0.19) g/100 g, and the control group increased total SFA in food purchases by 0.52 (95% CI −0.19 to 1.22) g/100 g, with a significant between-group difference of −1.05 (95% CI −1.83 to −0.27) g/100 g and P=.009 adjusted for baseline values. These results were robust in sensitivity analyses adjusted for age, gender, ethnicity, and income (Table S2 in Multimedia Appendix 1).
swaps at the point of choice helps improve the nutritional quality of food shopping [16,17,19].

This SwapShop intervention showed promising signs of early effectiveness, with observed reductions of $-1.97 \pm 1.22$ in total sugars and $-0.56 (95\% \text{ CI} -1.02$ to $-0.19$ in SFA in all food purchases in their respective intervention group, though this feasibility study was not powered to detect an effect in purchased nutrients. Data also show that smartphone app use is increasing across a wide range of demographic and age groups [31], and consumers are showing an increasing interest in healthier options [32]. However, although the app may be a useful tool for promoting healthier food choices through personalized advice and support, it is unlikely to be sufficient to achieve dietary recommendations. It is plausible that the impact could be enhanced by interventions to encourage product reformulation to offer greater availability of healthier alternatives. Moreover, price, promotions, positioning, and availability strategies within supermarkets have all been found to be major determinants of food choices [33-36], and these structural interventions are likely to be complementary.

**Strengths and Limitations**

This study’s strength lies in its randomized design and the process evaluation used to investigate the intervention’s pathway to impact. The qualitative components helped to provide context-specific information about the usability of the app and the acceptability of the swap suggestions. The SwapSHOP app had undergone extensive testing before this trial, given that it is an enhanced version of a previously tested app (SaltSwap) that was specifically designed to help people with hypertension choose lower-salt foods when grocery shopping in combination with face-to-face advice from a health care professional [20,37]. Furthermore, beta testing helped refine the app in line with the intended user feedback. SwapSHOP incorporated several behavioral elements with proven evidence to support dietary change [9,14,15], allowing people to set goals to swap to foods.

**Comparison With Prior Work**

Previous studies have shown the potential of instore swaps to support healthier choices, but there is very little evidence from interventions involving habitual shopping in physical stores. A quasi-experimental study evaluated the impact of the Change4Life Smart Swaps campaign to promote changes to lower fat or sugar foods when grocery shopping, showing that a higher percentage of participants in the intervention group reported choosing healthier options at the end of the study [29]. A total of 3 randomized controlled trials used a smartphone app to help reduce salt intake by promoting lower-salt swaps at the point of choice, showing changes in purchased salt in the intended direction but without evidence of changes in salt intakes [20,23,30]. Our previous study conducted in primary care settings provided individually tailored regular feedback on food shopping and offered lower-SFA swaps to patients with raised low-density lipoprotein cholesterol. The study showed modest but nonsignificant reductions in SFA consumption and SFA in purchased foods [18]. Other randomized controlled trials conducted in online retail environments, both real and experimental platforms, have shown that offering healthier swaps at the point of choice helps improve the nutritional quality of food shopping [16,17,19].

**Discussion**

**Principal Findings**

Participants in this study were willing and able to use the SwapSHOP app as intended, and the research methods ran as planned, with high levels of adherence and follow-up recorded. This provided preliminary evidence of effectiveness to support dietary change to lower the intake of sugars and SFAs. There was little evidence that this general population sample was motivated to reduce salt intake using the SwapSHOP app. Further improvements to the app, especially enhancing the coverage of product data and the specificity of the swap algorithm, are needed to provide a higher quality user experience.

**Table 4. Secondary outcomes—exploratory effectiveness measures.a,b**

<table>
<thead>
<tr>
<th></th>
<th>Baseline, mean (SD)</th>
<th>Follow-up, mean (SD)</th>
<th>Change, mean (95% CI)</th>
<th>Change adjustedb, mean (95% CI)</th>
<th>Between-group differenceb, intervention vs control</th>
<th>Mean (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purchased sugar (g/100 g)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sugar group (n=34)</td>
<td>5.13 (2.63)</td>
<td>4.13 (2.34)</td>
<td>−1.00 (−1.97 to −0.03)</td>
<td>−0.86 (−1.67 to −0.05)</td>
<td>−0.68 (−1.94 to 0.58)</td>
<td>−0.68</td>
<td>0.28</td>
</tr>
<tr>
<td>Control (n=10)</td>
<td>4.26 (1.49)</td>
<td>4.58 (1.86)</td>
<td>0.32 (−1.47 to 2.11)</td>
<td>−0.18 (−1.14 to 0.79)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Purchased SFAc (g/100 g)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFA group (n=28)</td>
<td>2.13 (1.18)</td>
<td>1.58 (0.98)</td>
<td>−0.56 (−1.02 to −0.10)</td>
<td>−0.55 (−0.89 to −0.22)</td>
<td>−1.05 (−1.83 to −0.27)</td>
<td>−1.05</td>
<td>0.009</td>
</tr>
<tr>
<td>Control (n=12)</td>
<td>2.10 (0.84)</td>
<td>2.61 (1.22)</td>
<td>0.52 (−0.19 to 1.22)</td>
<td>0.50 (−0.20 to 1.20)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

aLinear regression adjusted for baseline values.

bThe number of participants recruited for the salt group did not reach the target sample size. Data from the participants in this group were not analyzed further.

SFA: saturated fat.
lower the nutrients of concern, identify major sources of SFA and sugar in their shopping, and provide feedback on achievement of their goal as well as on the overall nutrient reduction achieved through making swaps. A systematic review of interventions using apps to support dietary behavioral improvements suggested greater benefits of multicomponent interventions compared with single-component interventions [38]. Unlike SwapSHOP, the 2 other existing smartphone apps provided information on healthier swaps but did not offer the option of self-monitoring the nutritional quality of food shopping and had limitations regarding how information is presented to consumers or the limited comprehensiveness of the databases covering the UK food market [21,22].

Although the SwapSHOP app included the major UK grocery retailers, participants particularly reported issues related to the product database, the limited range of UK grocery stores, and the improvements needed in the swapping function to create a list of healthier options that are similar to the original product and to accommodate dietary preferences. These aspects must be addressed in future versions of the app to maximize usability and acceptability. Purchased products that were recorded manually by the participants or had missing nutrient information were excluded from the analysis of changes in total purchases, limiting the robustness of the exploratory effectiveness outcome results, although this should not differ by trial arm.

Another limitation is that the study recruited a small, self-selected sample of people who were motivated to take action to improve their diet quality. A large proportion of participants reported receiving higher education and living in less deprived geographical areas than the national average [39]. A lower socioeconomic status is related to poorer dietary quality and health outcomes; hence, the observed results may not represent a wider population with lower adherence to dietary recommendations [40]. Ethnicity is also related to food choice [41]. Although this study mostly included people of White backgrounds, the app database included a wide range of products available in the UK market, thus covering different dietary preferences that apply to other ethnic groups.

Conclusions
SwapSHOP is a behaviorally informed smartphone app that allows users to scan barcodes of grocery products from major UK supermarkets, providing tailored nutritional information and suggesting personalized swaps to support dietary change. This study provided evidence of feasibility as a stand-alone intervention to support motivated individuals wanting to reduce their SFA or sugar intake as well as preliminary evidence of effectiveness to support healthier food purchases. Given the low cost and scalability of this intervention, after further refinement of the app technology and expanded market coverage, a definitive trial is warranted to assess the potential of this tool to improve health outcomes.

Acknowledgments
This study was funded by the National Institute for Health Research (NIHR) Applied Research Collaborations Oxford. SAJ, MN, and AH were funded by the NIHR Oxford Biomedical Research Centre (BRC) Obesity, Diet and Lifestyle Theme. CP, GH, and SAJ were funded by the NIHR Applied Research Collaborations Oxford. CL was funded by the Engineering and Physical Sciences Research Council and the NIHR Oxford BRC. SPR was supported by a British Heart Foundation Clinical Research Training Fellowship (FS/16/34/32211). CP is currently funded by grant RYC2020-028818-I, MCIN/AEI/10.13039/501100011033 and “ESF Investing in your future” (Ministry of Science and Innovation, Spain). The funders had no role in designing the study, data collection, analysis, interpretation of data, writing the report, or the decision to submit the manuscript for publication. The authors would like to thank Peter Scarborough and Richard Harrington for their help and support in using the FoodDB database of food products included in the SwapSHOP app. The University of Oxford is the owner of the SwapSHOP app.

Data Availability
The data sets generated during and/or analyzed during this study are available on reasonable request. Access to the study data and the statistical code will be reviewed and granted upon request by the Nuffield Department of Primary Care PRimsDISC committee (primdisc@phc.ox.ac.uk)

Authors’ Contributions
All authors contributed to the concept and design of the study. CP, AH, CL, MN, GH, and SPR developed the study protocols and materials, performed the experiments, and collected the study data. CP and CL analyzed and interpreted the data. CP wrote the manuscript and generated the tables and figures. All authors discussed the results and implications and commented on the final version of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File, 692 KB - mhealth_v12i1e45854_app1.docx]
References


**Abbreviations**

**SFA:** saturated fat

©Carmen Piernas, Charlotte Lee, Alice Hobson, Georgina Harmer, Sarah Payne Riches, Michaela Noreik, Susan A Jebb. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Effectiveness of Metaverse Space–Based Exercise Video Distribution in Young Adults: Randomized Controlled Trial

Rami Mizuta¹, MSc, PT; Noriaki Maeda¹, PT, PhD; Tsubasa Tashiro¹*, PT, PhD; Yuta Suzuki², PT, PhD; Sayo Kuroda¹, BA, PT; Ayano Ishida¹, BA, PT; Sakura Oda¹, BA, PT; Tomoya Watanabe¹, BA, PT; Yuki Tamura¹, BA, PT; Makoto Komiya¹, PT, PhD; Yukio Urabe¹*, PT, PhD

¹Department of Sport Rehabilitation, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan
²Department of Physical Therapy, Faculty of Rehabilitation, Kyushu Nutrition Welfare University, Fukuoka, Japan
*these authors contributed equally

Corresponding Author:
Yukio Urabe, PT, PhD
Department of Sport Rehabilitation
Graduate School of Biomedical and Health Sciences
Hiroshima University
1-2-3 Kasumi, Minami-ku
Hiroshima, 734-8553
Japan
Phone: 81 082 257 5405
Email: yurabe@hiroshima-u.ac.jp

Abstract

Background: In response to the serious lack of physical activity among young adults, recent attempts have been made to encourage young people to exercise through exercise video distribution. However, merely distributing videos does not lead to improved physical activity levels. Metaverse space, which enables web-based interaction through avatars, allows users to watch exercise videos in the same space as other avatars.

Objective: This study explored whether exercise video distribution using metaverse space is effective in improving physical activity levels, along with mental health and locomotive function, among young people.

Methods: In this parallel-group randomized controlled trial participants were recruited using printed poster displays. A total of 48 young adults aged between 18 and 30 years were assigned to 3 groups of 16 each: the metaverse, YouTube, and control group. To encourage exercise, the metaverse group was given an exercise video each week with a load of around 4–8 metabolic equivalents of tasks (METs) for 8 videos delivered in the metaverse space. The YouTube group was sent a URL on YouTube every week to view exercise videos with the same content as the metaverse group. The control group was given no special instructions. The intervention period was 8 weeks. Pre- and postintervention physical activity, well-being, locomotive syndrome risk tests, and social capital were measured. Although this study was not blinded to the participants, the measurers did not know to which group the participants belonged. Mixed model repeated-measures analyses and a post hoc Wilcoxon signed rank sum test were performed to detect the effects of the intervention in all groups.

Results: The results of the mixed model repeated-measures analyses showed a significant interaction between groups and before and after the intervention for total physical activity (metaverse group: pre 737.1, SD 609.5 METs/week, post 1575.4, SD 1071.8 METs/week; YouTube group: pre 661.7, SD 710.7 METs/week, post 911.9, SD 1103.3 METs/week; and control group: pre 930.6, SD 665.1 METs/week, post 844.7, SD 701.8 METs/week; P=.04) but none for the indicators of well-being (P=.40), locomotive function scale (P=.17), and social capital (P=.23). A post hoc test showed a significant increase in physical activity in the metaverse group before and after the intervention (P=.006).

Conclusions: This study is the first to show that delivering exercise videos through metaverse space can increase physical activity in young adults by providing a gathering space for individuals similarly motivated for exercise practice. This way, the sense of isolation during exercise is reduced compared with merely distributing videos on YouTube. The use of metaverse space in health promotion is likely to spread, and this study provides a useful reference for its exploration.

Trial Registration: ClinicalTrials.gov NCT06019156; https://ichgcp.net/clinical-trials-registry/NCT06019156
Introduction

Physical inactivity among young people is a serious problem, with 40.3% of men and 66.1% of women in their 20s reporting not exercising even once a week, according to the National Health and Nutrition Survey, 2019 [1]. Surprisingly, a study of 100 college students aged between 18 and 23 years found that 65% of them were in the high-risk group for locomotive syndrome, a condition that reduces physical function and mobility such as standing and walking. A study of young Japanese adults aged between 18 and 20 years revealed that 45.9% of them had a prevalence of frailty-prefrailty [2]. As Japan’s aging society progresses, the number of people requiring nursing care also increases, and the reasons for this need are often locomotive disorders such as falls and joint diseases [3,4]. Therefore, establishing exercise habits at a young age and maintaining and improving locomotive function are important.

The spread of COVID-19 has led to a lack of exercise and a change to a more sedentary lifestyle, which has focused attention on exercise interventions over home-based web-based video delivery methods. This promotion of physical activity through exercise videos was intended to help address people’s lack of motivation to engage in physical activity when gymnasiums and recreation centers had to close and physical distance was required [5]. An exercise intervention study using YouTube reported that when a new exercise video of approximately 5 minutes was delivered once a week for 12 weeks, physical activity and frequency of muscle training increased in the exercise intervention group [6]. Hence, the distribution of exercise videos through YouTube, the most accessible tool for young people, has shown some effectiveness and has become commonplace even as of 2023, when COVID-19–related restrictions have eased.

However, this conventional promotion of physical activity through the distribution of exercise videos has elements that fail to encourage behavioral changes. Self-determination theory, as it relates to exercise implementation, begins with the fulfillment of 3 basic human psychological needs: competency (the experience of achievement and efficacy), autonomy (the experience of motivation), and relatedness (the experience of achievement and efficacy), autonomy (the experience of motivation), and relatedness (the experience of connection with others) [7]. However, traditional exercise video distribution using YouTube and other media has focused on individual video-watching, which fails to provide opportunities for connecting with other people. In fact, another study of an exercise intervention using YouTube reported improvements in mental health with no increase in physical activities [8]. Studies showed connection with others as an important factor in exercise implementation and that being in an exercise group was more effective in increasing physical activity than exercising alone [9,10]. Further, social encouragement by those around an individual and exercising together had a positive impact on the acquisition of exercise habits [11,12]. In addition, people who shared their step count information using a special app tended to exercise the next day when other people “liked” or commented on their step counts [13]. Moreover, the group that was provided with peer interaction and social support showed positive effects on locomotive functions, such as improved results in stand-up tests, compared with the group that lacked these supportive activities [14]. Thus, adding an element of social community when promoting exercise is more effective.

The metaverse is an internet-based 3D virtual world. For the generation of digital natives born in the 1990s and 2000s, it is expected that the metaverse will increasingly become a space where they spend part of their daily lives [15]. In recent years, the metaverse space has gained attention in the health care field [16-18], and a bibliometric analysis of virtual and augmented reality showed that the metaverse has been adapted for diagnosis, surgical treatment, and rehabilitation of pain, stroke, anxiety, depression, fear, cancer, and neurodegenerative diseases, with satisfactory results [16]. In a metaverse space, people can interact with and encourage others on the web through their avatars, which are their own alter egos, and create a community of people who gather in the same space. Social distance is known to interfere with mental health [19], and the metaverse’s ability to recreate a space for social interaction is expected to help improve the treatment of mental health symptoms such as anxiety, stress, and eating disorders [20-23]. Therefore, the use of the metaverse has the potential to improve mental health. Moreover, encouraging physical activity within the metaverse space has the potential to promote behavioral responses from users, such as motivating people to adopt a healthier lifestyle [24]. Therefore, exercise training using the metaverse is expected to develop as a method to promote physical activity among young people [16]; however, there is insufficient knowledge on whether physical activity can be promoted by distributing exercise videos in the metaverse space. This study verified whether exercise video distribution using an app that provides metaverse space is effective in improving physical activity levels, mental health, and locomotive function among young people. We hypothesized that the distribution of exercise videos in the metaverse space would increase physical activity and have positive effects related to mental health and locomotive function.

Methods

Study Design and Participants

This study was a parallel-group, randomized controlled trial. The participants were recruited among Hiroshima University students between August 10 and September 9, 2022. For recruitment, we displayed a poster explaining the study and indicated a contact person available for questions for those considering participation. Several people expressed their willingness to participate. Participants were offered an incentive of US $6.93. Written consent was obtained after we explained the study details face-to-face. All data were collected in person.
This study’s specific participation criteria were as follows: individuals (1) aged 18 and 30 years at the time of obtaining consent, (2) who owned a smartphone or computer and had internet access, and (3) who did not belong to an exercise community such as a sports club. The exclusion criteria were as follows: individuals (1) who had a history of a disease that prohibited exercise, (2) whose physical activity was >3000 metabolic equivalents of tasks (METs)/week because a person with >3000 METs/week is defined as active [25-27], and (3) who were (or possibly) pregnant.

The sample size was calculated using G*power 3.1.9.2 (version 3.1.9.7; Heinrich-Heine-University Düsseldorf). When the effect size was calculated using raw data from a previous study with a similar design, which used a web-based exercise intervention with a physical activity measure as the main outcome [28], a large effect size of 0.34 was obtained with a partial $\eta^2$ of 0.108. When setting $\alpha$ error probability to .05, power (1–$\beta$ error probability) to .8, number of groups to 3, and number of measurements to 2, the required sample size was 27 participants, with 9 participants per group. A total of 48 participants were required when the dropout rate was approximately 40% [8].

After completion of the baseline prequestionnaire and measurements, 3 participants who met the exclusion criteria and were physically active for >3000 METs/week were excluded. The participants were divided into 3 groups: “metaverse space-based exercise video distribution group (metaverse group),” “YouTube-based exercise video distribution group (YouTube group),” and “no video distribution group (control group)” in a 1:1:1 ratio. The allocation was performed by segregating blocks of size 3 to generate the sequence [29], and the allocation order was hidden until after each group was allocated. A total of 3 separate staff members performed each of the 3 tasks: generating the random allocation sequence, enrolling participants, and assigning participants to the interventions. We explained to all participants the nature of the 3 groups while obtaining their consent. After allocation, the metaverse and YouTube groups received an 8-week intervention in which the exercise videos were distributed, and a control group was established. The intervention period was from October 3 to November 27, 2022. Thereafter, a postquestionnaire survey and measurements were conducted for all 3 groups.

**Intervention**

**Procedure**

The recruitment ran for approximately a month, from August 10 to September 9, 2022, and the premeasurements were taken during a 2-week period from September 12 to 25, 2022. The 8-week intervention was conducted from October 3 to November 27, 2022. The postmeasurements were conducted over a period of 1 week (from December 5 to 11, 2022).

**Metaverse Group**

The participants in the metaverse group had access to Metaverse space (Spatial Systems Inc), where they could watch exercise videos whenever they wanted (Figure 1). The researcher created a metaverse space for watching exercise videos and set up new videos each week. The URL to access them was shared only with participants in the metaverse group. The unique feature of the metaverse was that users could enter a room to watch exercise videos with their avatars, which were their own alter egos and other avatars. The participants watched exercise videos, which were created by a physical therapist to increase their physical activity, for approximately 5 minutes [6,8]. Then, we instructed them to do the exercise while watching videos. The load setting was approximately 4-8 METs to train the trunk and the upper and lower extremities. A new video was released once a week [6], and the participants could watch 8 videos for 8 weeks. They were informed in advance that new videos would be updated every Monday and, as a reminder, were sent a SMS text message when a new video was available. A web-based survey regarding exercise video viewing frequency was administered, and adherence was tracked weekly.

**Figure 1.** The metaverse space in this study.
**YouTube Group**

The YouTube group was sent a message with the URL to access the exercise video every Monday. The YouTube group could watch videos at their preferred times by clicking on a link. The content and time of the video were identical to those of the metaverse group. As in the metaverse group, adherence to exercise implementation was confirmed through a web-based questionnaire, and video update reminders were sent every week.

**Control Group**

Participants were informed by email that they were assigned to the control group simultaneously with the start of the intervention in the Metaverse and YouTube groups. We asked the participants in the control group to spend 8 weeks as usual, without any special instructions or delivery of videos. This group was established as an indicator of the amount of physical activity that may be significantly affected by the infection status and government directives during the COVID-19 pandemic.

**Outcome Measurements**

**Basic Information**

The participants were asked and tested for age, height, weight, BMI, lifestyle, sex, living status, and frailty to characterize the basic information. The weight and frailty tests were measured by a physical therapist, and the other parameters were measured using a questionnaire.

**Lifestyle**

Breslow’s 7 health practice scores were used to assess lifestyle [30,31]. The 7 parameters were: smoking (presence or absence), alcohol intake (high or low: consumption of ≥5 or <5 alcoholic beverages), sleep (≥7 hours or <7 hours), breakfast (eating or not eating), eating between meals (eating or not eating), exercise (≥2 times, 30 minutes/week or <2 times, 30 minutes/week), and proper weight (BMI <25 kg/m² or ≥25 kg/m²).

**Living Status**

Regarding their living status, the participants were asked whether they lived alone or with others.

**Frailty Test**

The Japanese Cardiovascular Health Study criteria questionnaire was designed to indicate potential frailty, and its contents are as follows [32,33]: (1) have you lost 2 kg or more in the past 6 months? (2) Do you engage in moderate levels of physical exercise or sports aimed at improving health? (3) In the past 2 weeks, have you felt tired without reason? (4) Hand grip strength of <26 kg in men or <18 kg in women, and (5) gait speed of <1 m/second. Grip strength was measured using a hand dynamometer (TKK 5401; Takei). Each participant performed 2 trials, and the best value from both trials was used for the analysis. Walking speed was calculated by performing a 10-meter walk test and dividing the measured value by 10 to obtain the results per meter. All questions were rated 1 or 0, and higher scores indicated a higher possibility of frailty. Frailty, prefrailty, and robustness were defined by the number of applicable items: 3-5, 1-2, and 0.

**Primary and Secondary Outcomes**

The study assessed physical activity, well-being, locomotive syndrome risk tests, and social capital before and after the intervention for 8 weeks.

**Primary Outcome: Physical Activity**

The short form of the International Physical Activity Questionnaire was used to measure the effect of an intervention using exercise videos with different distribution methods (Metaverse and YouTube) on physical activity [34]. Total physical activity was measured as the average amount of vigorous physical activity, the amount of moderate activity, and the amount of walking (METs minutes/week). A MET is defined as the amount of energy required for a person to sit quietly. In the premeasurement, the participants answered for the most recent week, and in the postmeasurement, they responded for the week after the intervention was completed.

**Secondary Outcomes: Well-being, Locomotive Syndrome Risk Tests, and Social Capital**

**Well-Being**

To assess well-being and mental health, we used the Japanese version of the World Health Organization-Five Well-being Index [35]. Responses to each item were rated on a 6-point scale from 0 to 5 and consisted of 5 questions. The highest score is 25, with higher scores indicating better well-being.

**Locomotive Syndrome Risk Test**

We used the 25-question geriatric locomotive function scale, proposed by the Japanese Orthopedic Association [33,36], to measure levels of locomotive function based on normal daily activities and pain. All the items were answered on a scale of 0-4. The higher the score, the more impaired the motor function, which indicates a decline in mobility and interference with social life.

**Social Capital**

The social capital questions consisted of 11 items related to civic participation, social cohesion, and reciprocity, with a perfect score of 11 and higher scores indicating stronger social connections [37]. We included this question item because of the possibility that the sense of social participation could change in the metaverse group.

**Supplemental Outcomes: Impression of Each Intervention (Metaverse and YouTube Groups)**

We interviewed each of the metaverse and YouTube groups regarding their impressions of the intervention using exercise videos after the 8-week intervention. User experience is the experience of a user with the use of a product, system, or service. From the review of user experience, we selected the following list of questions: novelty, relatedness (connection with others), motivation, excitement, satisfaction, delight, comfort, attractiveness, expectation, and fulfillment [38]. The participants answered each question with a “good,” “rather good,” “rather poor,” or “poor” response.
Primary and secondary outcomes were analyzed by performing a mixed model repeated-measures ANOVA. This analysis was used to detect the effect of interventions between the metaverse, YouTube, and control groups. We assessed the main effect as well as group and time interactions on the outcome measure. A mixed model repeated-measures analysis is an intention-to-treat analysis with unbiased estimates that considers all available data from participants enrolled in the study [39]. The effect size for the interaction effect of a mixed model repeated-measures analysis was calculated using F statistics. As a posttest, the Wilcoxon signed rank sum test was performed after confirming nonnormality to compare the pre- and postintervention outcomes of each group. The z-statistic was used to calculate the effect size. Supplemental outcomes were analyzed by performing chi-square tests to compare the intervention impressions between the metaverse and YouTube groups. Note that to use the chi-square test, “rather good” was included in “good” and “rather poor” was included in “poor,” thus reorganizing responses into 2 groups. The effect size was calculated using Cramer V.

Statistical analyses were performed using SPSS (version 28.0; IBM Corp). The significance level was set at 5%.

Figure 2. Study flowchart. MET: metabolic equivalents of task.

This study was approved by the Ethical Committee for Clinical Research at Hiroshima University (C2022-0004) and registered with UMIN (UMIN000048046). We followed the guidelines of the Consolidated Standards for Reporting Trials [40]. There were no significant changes in the method used after the study’s initiation.

Participants

A total of 51 individuals expressed interest in the study; however, 3 were excluded because their physical activity was >3000 METs/week. All the participants in the metaverse and YouTube groups completed the intervention. Participants were also interviewed weekly during the 8-week intervention period to monitor their exercise implementation. The average frequency of exercise while watching the exercise video per week was 4.1 (SD 3.9) and 2.6 (SD 1.6) for the metaverse and YouTube groups, respectively (response rates: metaverse group 100% and YouTube group 99%). For the postmeasurement, a participant in the control group could not participate owing to illness.

Table 1 presents the participants’ demographic characteristics (mean 22.4, SD 2.4 years; 32/48, 67% women). Prefrailty accounted for 41/48 (85%) participants because we recruited those who were not members of a sports club and whose physical activity was <3000 METs/week.
Table 1. Participant’s demographics at baseline.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total</th>
<th>Metaverse group (n=16)</th>
<th>YouTube group (n=16)</th>
<th>Control group (n=16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>22.4 (2.4)</td>
<td>22.9 (1.4)</td>
<td>22.5 (2.0)</td>
<td>22.0 (3.4)</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>164.0 (7.9)</td>
<td>162.4 (6.6)</td>
<td>163.9 (6.2)</td>
<td>165.7 (10.3)</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>57.7 (11.6)</td>
<td>56.7 (9.8)</td>
<td>57.3 (10.2)</td>
<td>59.0 (14.8)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>21.3 (2.9)</td>
<td>21.4 (2.5)</td>
<td>21.2 (2.5)</td>
<td>21.3 (3.6)</td>
</tr>
<tr>
<td>Lifestyle, mean (SD)</td>
<td>4.6 (0.9)</td>
<td>4.6 (0.9)</td>
<td>4.4 (0.7)</td>
<td>4.7 (1.1)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16 (33)</td>
<td>6 (38)</td>
<td>6 (38)</td>
<td>5 (31)</td>
</tr>
<tr>
<td>Female</td>
<td>32 (67)</td>
<td>10 (63)</td>
<td>10 (63)</td>
<td>11 (69)</td>
</tr>
<tr>
<td>Living status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>31 (65)</td>
<td>11 (69)</td>
<td>10 (63)</td>
<td>10 (63)</td>
</tr>
<tr>
<td>With others</td>
<td>17 (35)</td>
<td>5 (31)</td>
<td>6 (38)</td>
<td>6 (38)</td>
</tr>
<tr>
<td>Frailty test, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>2 (4)</td>
<td>1 (6)</td>
<td>1 (6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Prefrailty</td>
<td>41 (85)</td>
<td>13 (81)</td>
<td>14 (88)</td>
<td>14 (88)</td>
</tr>
<tr>
<td>Robust</td>
<td>5 (10)</td>
<td>2 (13)</td>
<td>1 (6)</td>
<td>2 (13)</td>
</tr>
</tbody>
</table>

Physical Activity, Well-Being, Locomotive Syndrome Risk Test, and Social Capital

Table 2 shows the estimation results of the fixed effects from the mixed model repeated-measure analysis. Only total physical activity showed a significant interaction ($F_{2.45,042}=3.338$, $P=.04$; effect size=0.263). The main effects for the time showed a significant difference in vigorous activity ($F_{1.45,083}=6.921$, $P=.01$), locomotive function scale ($F_{1.44,895}=9.557$, $P=.003$), and social capital ($F_{1.44,650}=5.085$, $P=.03$).

Table 3 shows the results of the post hoc test comparison of the difference in total physical activity before and after the intervention in each group. In the metaverse group, a significant difference was detected between pre- and postintervention periods ($P=.006$; effect size=0.682).
<table>
<thead>
<tr>
<th>Variables</th>
<th>Metaverse group Mean (SD)</th>
<th>YouTube group Mean (SD)</th>
<th>Control group Mean (SD)</th>
<th>Main effect</th>
<th>Interaction effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Time effect</td>
<td>Group effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F test (df)</td>
<td>P value</td>
</tr>
<tr>
<td>Physical activity: total (METs\textsuperscript{a} minutes/week), mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td>5.153 (1, 45.051)</td>
<td>.03</td>
</tr>
<tr>
<td>Pre</td>
<td>737.1 (609.5)</td>
<td>661.7 (710.7)</td>
<td>930.6 (665.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>1575.4 (1071.8)</td>
<td>911.9 (1103.3)</td>
<td>844.7 (701.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity: vigorous (METs minutes/week), mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td>6.921 (1, 45.083)</td>
<td>.01</td>
</tr>
<tr>
<td>Pre</td>
<td>237.5 (423.4)</td>
<td>120 (231.9)</td>
<td>137.5 (253.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>745 (677.4)</td>
<td>345 (613.90)</td>
<td>162.7 (493.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity: moderate (METs minutes/week)</td>
<td></td>
<td></td>
<td></td>
<td>2.004 (1, 44.498)</td>
<td>.16</td>
</tr>
<tr>
<td>Pre</td>
<td>211 (270.3)</td>
<td>162.5 (319.2)</td>
<td>311.3 (458.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>425 (347.1)</td>
<td>235 (411.5)</td>
<td>317.6 (351.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity: walking (METs minutes/week)</td>
<td></td>
<td></td>
<td></td>
<td>0.013 (1, 45.119)</td>
<td>.91</td>
</tr>
<tr>
<td>Pre</td>
<td>288.6 (255.8)</td>
<td>379.2 (441.9)</td>
<td>481.8 (552.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>636.5 (686.8)</td>
<td>331.9 (368.1)</td>
<td>364.4 (284.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well-being</td>
<td></td>
<td></td>
<td></td>
<td>2.449 (1, 44.726)</td>
<td>.12</td>
</tr>
<tr>
<td>Pre</td>
<td>15.4 (3.9)</td>
<td>16.4 (4.1)</td>
<td>18.4 (4.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>17.6 (5.5)</td>
<td>17.3 (3.9)</td>
<td>18.5 (4.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locomotive function scale</td>
<td></td>
<td></td>
<td></td>
<td>9.557 (1, 44.895)</td>
<td>.003</td>
</tr>
<tr>
<td>Pre</td>
<td>3.9 (3.4)</td>
<td>3.8 (3.3)</td>
<td>3.5 (3.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>1.1 (1.9)</td>
<td>2.6 (2.9)</td>
<td>2.9 (3.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social capital</td>
<td></td>
<td></td>
<td></td>
<td>5.085 (1, 44.650)</td>
<td>.03</td>
</tr>
<tr>
<td>Pre</td>
<td>5.9 (1.9)</td>
<td>4.9 (1.8)</td>
<td>6.2 (1.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>7.1 (2.4)</td>
<td>5.7 (2.0)</td>
<td>6.2 (1.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}MET: metabolic equivalents of task.
Table 3. Differences between pre- and postintervention outcome measures (N=48).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Physical activity: total (metabolic equivalents of tasks minutes/week)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Metaverse group</td>
</tr>
<tr>
<td>Pre, mean (SD)</td>
<td>737.1 (609.5)</td>
</tr>
<tr>
<td>Post, mean (SD)</td>
<td>1575.4 (1071.8)</td>
</tr>
<tr>
<td>( P ) value</td>
<td>0.006</td>
</tr>
<tr>
<td>Effect size</td>
<td>0.682</td>
</tr>
</tbody>
</table>

**Differences observed between the 2 groups in novelty (**\( P = 0.003; \) effect size=0.529), relatedness (**\( P = 0.03; \) effect size=0.387), excitement (**\( P = 0.01; \) effect size=0.433), delight (**\( P = 0.002; \) effect size=0.539), attractiveness (**\( P = 0.006; \) effect size=0.487), and expectation (**\( P < 0.001; \) effect size=0.626). The results did not reveal significant differences in motivation, satisfaction, comfort, or fulfillment.

Figure 3. Impressions of intervention in the Metaverse and YouTube groups. *\( P < 0.05 \).
**Discussion**

This study examined whether an exercise intervention using exercise videos in the metaverse space has a positive impact on physical activity. The main results show a significant interaction between groups and time (pre- and postintervention) in total physical activity, with post hoc analysis showing a significant increase in total physical activity in the metaverse group after compared with before the intervention and no significant change in the YouTube and control groups. This study is the first to show that using the metaverse space to deliver exercise videos can promote increased physical activity.

The current results support the hypothesis that the use of the metaverse increases physical activity. The exercise videos used in the metaverse, and the YouTube groups are identical, and the metaverse group differs from the YouTube group in that the exercise videos can be viewed in a shared space with other avatars in the metaverse space. Studies have noted that interacting with others is more effective than exercising alone at increasing physical activity and sustaining exercise [9-13]. Metaverse characteristics include the ability to form communities within virtual reality and the persistence of virtual reality regardless of individual user access [15]. In this study, no change in social capital was detected before and after the intervention; however, the percentage of those who had an impression of relatedness (connection with others) was also significantly higher in the metaverse group (9/16, 56%) than in the YouTube group (3/16, 19%)—indicating that more people felt connected to others when using the metaverse than when using YouTube. Additionally, strategies to increase physical activity include improving the physical environment, such as its construction, and the social environment, including social support [41]. The use of the metaverse resulted in the persistence of a purposeful social space for exercise that could be accessed at any time. Therefore, the existence of the metaverse space is an aspect of environmental improvement for exercise for participants who previously had no exercise habits, leading to increased exercise implementation and physical activity.

The reason indicators other than physical activity did not improve in the metaverse group is also an important area to consider. This study asked about well-being as an item that could be related to mental health. Exercise is known to promote the secretion of estradiol and serotonin and regulate cortisol levels, an internal change that reduces depression and other depressive symptoms [42-44]. Additionally, reducing loneliness and encouraging social interaction decreases the risk of depression and mental health problems [45]. Therefore, we expected to observe mental health improvements in the metaverse group; however, there were no significant differences among the 3 groups. The standard for suspicion of mental problems is a cutoff value of ≤ 11 points for well-being [35,46]. However, the mean values for participant well-being did not correspond to the cutoff values. Thus, mental health changes were unlikely to be observed even if the amount of physical activity increased because the group did not initially have low well-being.

The locomotive function scale, a measure of body pain and locomotor function, showed an improving trend in the metaverse group; however, the difference was not significant. A previous study found pain relief after an 8-week exercise program, including strength training of the trunk, in participants with low back pain that appeared to be caused by inactivity [47]. Several studies have reported muscle hypertrophy occurring within 6-8 weeks [48]. These facts indicate that if training is managed meticulously in the right way, the body can be changed in a period of 8 weeks. However, it is possible that the unmonitored, 8-week intervention period did not result in sufficient changes in motor function, even though the amount of physical activity increased. To improve motor function, training to improve muscle strength and endurance should be continued at least 3 times per week [49].

As for the differences in impressions of the intervention between the metaverse and YouTube groups, many factors such as novelty, relatedness, excitement, delight, attractiveness, and expectation were more positive in the metaverse group than in the YouTube group. Regarding novelty, a previous study stated that sufficient novelty for the target audience leads to successful health support using the app [50]. Even for young adults who were relatively familiar with and used apps and social networking services on a regular basis, the metaverse seemed to be a novelty, which may have supported the implementation of the exercise in this study. Delight is an important factor in many areas of life [51]. Studies have reported that delight is the most important predictor of commitment and participation in general youth and elite sports [52,53]. This amplification of delight using a metaverse space to distribute videos is meaningful in promoting exercise. Interestingly, the metaverse group found the intervention more attractive. A previous study that used apps to promote healthy lifestyles, including regular exercise, found that a combination of different services, such as health-related quizzes and the ability to record health status, led to successful interventions [54]. Exercise video-delivered studies have also shown that simply delivering videos does not increase physical activity [8]. The essence of using the metaverse may have increased the attractiveness of exercise video distribution in this study.

This study examined the effectiveness of exercise video distribution using metaverse space by comparing the previously used YouTube and a control population. This study’s strength lies in making the first attempt to distribute exercise videos using metaverse space. Furthermore, it shows the potential of using metaverse space to improve the physical activity level of young adults.

However, this study has some limitations. First, selection bias may have occurred. We recruited participants who were not members of an athletic club and who did not exercise heavily. However, it is possible that the target population was composed of students who were interested in exercising, as suggested by their interest in participating in this study. Additionally, we recruited participants from a single university. The risk of selection bias may be reduced by conducting a study on a larger scale and involving more institutions. However, the 3 groups were randomly assigned in this study, and the required, precalculated sample size was secured. Second, it was difficult
to accurately determine the time and frequency of interactions between participants in the metaverse space. If a system can be implemented to monitor the time of entry into the metaverse space, the frequency of interactions, and so on, the mechanism for promoting physical activity through the use of the metaverse space could be made more visible. Third, participants could receive the intervention at the time of their choosing; however, the start time of the intervention was not monitored. The effectiveness of the intervention could be moderated or altered by the time of the intervention, and future researchers should accurately monitor the start time of the intervention. We note that the main outcome, the calculation of physical activity, did not depend on the starting time of the intervention. Final, there was room for improvement in the comfort of the interventions in the metaverse group. This study used metaverse space in the manner in which the videos were arranged. Metaverse space (Spatial, Spatial Systems, Inc) has the advantage of being a service that anyone can use free of charge; however, if industry and academia collaborate to devise a system more suited to the distribution of exercise videos and the promotion of exercise, it could have a more positive effect on physical activity and other aspects.

In the near future, services allowing access to the metaverse will likely continue to expand. Therefore, the number of opportunities for the general public to experience the metaverse will further increase. The fact that 100% of the metaverse group in this study had high expectations is a hopeful part of the development of using the metaverse to improve physical activity. This research provides a basis for establishing further methods of using the metaverse to promote exercise.

Acknowledgments

We thank all the participants. We also thank Editage for the English language editing. This work was supported by the Japan Science and Technology Agency through the establishment of university fellowships toward the creation of science and technology innovation (JPMJFS2129).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of impressions on intervention between the metaverse and YouTube groups.

[PDF File (Adobe PDF File), 160 KB - mhealth_v12i1e46397_app1.pdf ]

Multimedia Appendix 2

CONSORT-eHealth (V 1.6.1).

[PDF File (Adobe PDF File), 1312 KB - mhealth_v12i1e46397_app2.pdf ]

References


Abbreviations

MET: metabolic equivalents of task

https://mhealth.jmir.org/2024/11/e46397
Effectiveness of Metaverse Space–Based Exercise Video Distribution in Young Adults: Randomized Controlled Trial


Please cite as:
Effectiveness of Metaverse Space–Based Exercise Video Distribution in Young Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e46397
URL: https://mhealth.jmir.org/2024/1/e46397
doi:10.2196/46397
PMID:38227355

©Rami Mizuta, Noriaki Maeda, Tsubasa Tashiro, Yuta Suzuki, Sayo Kuroda, Ayano Ishida, Sakura Oda, Tomoya Watanabe, Yuki Tamura, Makoto Komiya, Yukio Urabe. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Review

Functionality and Quality of Asthma mHealth Apps and Their Consistency With International Guidelines: Structured Search and Evaluation

Billy Robinson¹, BMedSci, BMBS, MMed; Eleni Proimos², BMedSc, MD; Daniel Zou³, BBioMedSci, MD; Enying Gong⁴, BA, MPH, PhD; Brian Oldenburg⁵,⁶, PhD; Katharine See¹, MBBS

¹Department of Respiratory Medicine, Northern Health, Epping, Australia
²Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia
³Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia
⁴School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
⁵Baker Heart and Diabetes Institute, Melbourne, Australia
⁶Academic and Research Collaborative in Health, LaTrobe University, Melbourne, Australia

Corresponding Author:
Billy Robinson, BMedSci, BMBS, MMed
Department of Respiratory Medicine
Northern Health
185 Cooper St
Epping, 3076
Australia
Phone: 61 38405 8000
Email: billymed1994@gmail.com

Abstract

Background: Asthma is a chronic respiratory disorder requiring long-term pharmacotherapy and judicious patient self-management. Few studies have systematically evaluated asthma mobile health (mHealth) apps for quality and functionality; however, none have systematically assessed these apps for their content alignment with international best practice guidelines.

Objective: This review aims to conduct a systematic search and evaluation of current mHealth apps in the Australian marketplace for their functionality, quality, and consistency with best practice guidelines.

Methods: The most recent Global Initiative for Asthma (GINA) guidelines were reviewed to identify key recommendations that could be feasibly incorporated into an mHealth app. We developed a checklist based on these recommendations and a modified version of a previously developed framework. App stores were reviewed to identify potential mHealth apps based on predefined criteria. Evaluation of suitable apps included the assessment of technical information, an app quality assessment using the validated Mobile App Rating Scale (MARS) framework, and an app functionality assessment using the Intercontinental Medical Statistics Institute for Health Informatics (IMS) Functionality Scoring System. Finally, the mHealth apps were assessed for their content alignment with the GINA guidelines using the checklist we developed.

Results: Of the 422 apps initially identified, 53 were suitable for further analysis based on inclusion and exclusion criteria. The mean number of behavioral change techniques for a single app was 3.26 (SD 2.27). The mean MARS score for all the reviewed apps was 3.05 (SD 0.54). Of 53 apps, 27 (51%) achieved a total MARS score of ≥3. On average, the reviewed apps achieved 5.1 (SD 2.79) functionalities on the 11-point IMS functionality scale. The median number of functionalities identified was 5 (IQR 2-7). Overall, 10 (22%) of the 45 apps with reviewer consensus in this domain provided general knowledge regarding asthma. Of 53 apps, skill training in peak flow meters, inhaler devices, recognizing or responding to exacerbations, and nonpharmacological asthma management were identified in 8 (17%), 12 (25%), 11 (28%), and 14 (31%) apps, respectively; 19 (37%) apps could track or record “asthma symptoms,” which was the most commonly recorded metric. The most frequently identified prompt was for taking preventive medications, available in 9 (20%) apps. Five (10%) apps provided an area for patients to store or enter their asthma action plan.

Conclusions: This study used a unique checklist developed based on the GINA guidelines to evaluate the content alignment of asthma apps. Good-quality asthma apps aligned with international best practice asthma guidelines are lacking. Future app
development should target the currently lacking key features identified in this study, including the use of asthma action plans and the deployment of behavioral change techniques to engage and re-engage with users. This study has implications for clinicians navigating the ever-expanding mHealth app market for chronic diseases.

**Trial Registration:** PROSPERO CRD42021269894; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=269894

**International Registered Report Identifier (IRRID):** RR2-10.2196/33103

*(JMIR Mhealth Uhealth 2024;12:e47295)* doi:10.2196/47295

**KEYWORDS**

asthma; mobile health; mHealth; app; mobile; chronic disease; systematic review; smartphone; review methodology; respiratory; compliance; guideline; guidelines; review of apps; evaluation; quality; best practices; apps; mobile phone

**Introduction**

**Background**

Asthma is a chronic respiratory disorder that is clinically defined as a combination of typical episodic respiratory symptoms, such as wheezing, shortness of breath, cough, chest tightness, and significant variable reversible airflow limitation [1]. When the frequency or severity of these symptoms increases compared with the baseline respiratory status, it represents an “asthma exacerbation” or “flare-up” [2]. Judicious self-monitoring and management of regular asthma medications, symptoms, and exacerbations are key to allowing patients with asthma to live with a high quality of life and prevent hospitalizations or death [3]. The Global Initiative for Asthma (GINA) regularly releases updated best practice asthma guidelines based on reviews of scientific literature by an international panel of experts [4]. Many local asthma management guidelines have been derived from these international guidelines. In addition to pharmacotherapy, the guidelines advise that patient education on medication adherence, exacerbation recognition, and management is key to self-management [4].

Asthma is a significant chronic health issue worldwide, affecting 1% to 18% of the global population [4]. Australia is no exception, with asthma affecting millions and accounting for 34% of Australia’s respiratory disease burden and 2.5% of the total disease burden [5]. Asthma leads to numerous emergency department visits and urgent health care visits [5]. Furthermore, those living with asthma report a poorer quality of life and are less likely to rate their health status as excellent or very good [5]. When observing the total cost that asthma has on the Australian health system, it is evident that hospital-related costs outweigh non–hospital-related costs (Aus $205 million/year [approximately US $150 million] vs Aus $163 million/year [approximately US $120 million]) [5]. Theoretically, reducing exacerbations would reduce the requirement for hospitalizations; unplanned primary care presentations; and indirect costs, such as work absenteeism, and thus assist in reducing these costs.

With the increasing availability of smartphones, mobile health (mHealth) apps have become accessible to a large percentage of the population and represent a potential medium through which patients can improve their ability to self-manage asthma. Deloitte’s recent review of Australia’s telecommunication status found that 89% of the Australian population uses smartphones [6]. These apps are already available for download and use; however, it is imperative that a review of their quality, functionality, and alignment with evidence-based best practices is conducted to inform both users and health professionals. These apps represent an opportunity for new ways to empower patients to track asthma symptoms, learn about their condition, and undertake practical self-management strategies. The established Mobile App Rating Scale (MARS) is generally used to assess the usability and overall quality of mHealth apps [7,8]. Although systematic evaluations of asthma mobile apps have been conducted in the past, many of these studies did not assess the apps’ functionality or quality using a validated rating scale, such as MARS [9-11]. Furthermore, to our knowledge, none of these prior evaluations assessed all available apps systematically for the presence and quality of information they provide compared with available best practice management guidelines, such as the GINA guidelines [9-11].

This systematic search and evaluation review assessed the functionality and quality of free and paid asthma mHealth apps targeted toward adults with asthma available from the Apple App Store (iOS) and Google Play Store (Android), as well as their consistency with recommendations from the GINA guidelines, making it the first review of its kind.

**Objective**

The objective of this review was to conduct a systematic search of available English-language mHealth apps targeted toward adults with asthma in Australia, to evaluate their overall quality and functionality and to assess the consistency and quality of the content and information they provide in alignment with current best practice guidelines for asthma management.

**Methods**

**Overview**

The GINA guidelines were reviewed by 2 medical professionals to identify and establish a consensus of key recommendations from the guidelines that could feasibly be incorporated into an app for asthma management. The mobile apps in the selected app stores were identified and screened based on the selection criteria. Finally, we assessed the quality, functionality, and alignment of the apps with the guidelines identified in the first step of the screened mHealth apps. An in-depth description of the research protocol was published the previous year [12].

**Study Setting**

Given the primary residences of the researchers involved in this review, this study was conducted by medical practitioners,
medical students, and digital health researchers using apps available in Australia. The mHealth apps presented in English on Australian mobile app stores were assessed. Some mHealth apps identified in this review may not be available in regions outside Australia. Similarly, apps available in other regions may not be available in Australia. However, given that most of the apps identified in this review are also available in other English-speaking regions such as the United Kingdom and the United States, the results are largely generalizable to these regions. Given that the researchers were adult medical practitioners who did not manage pediatric patients, only those mHealth apps targeted toward adults with asthma were evaluated.

Wherever possible, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for systematic reviews were followed [13]. Given that this was a review of mobile apps instead of journal articles, some items in the PRISMA checklist were not relevant to this review. The checklist is shown in Multimedia Appendix 1.

Review of the GINA Guidelines and Checklist Creation

To assess the usability and overall quality of the app, we used the established MARS [7,8]. A review of the available literature using the CINAHL, MEDLINE, Embase, and PubMed databases revealed that 1 research group had developed an asthma app assessment framework yet to be derived and validated into an instrument [14]. For the reasons outlined in our published research protocol, we decided to combine aspects of the framework by Guan et al [12] with our own checklist derived from the GINA guidelines. Two reviewers, BR and KS, independently assessed the 2020 GINA guidelines for identifiable recommendations that could be incorporated into an mHealth app. Following this, the reviewers examined each other’s identified recommendations to see whether a consensus could be reached on the recommendations from the GINA guidelines that could be incorporated into an mHealth app. The identified recommendations from each author and those where a consensus was reached, which represent the final identified recommendations, are shown in Table 1.

A final checklist modified from the framework by Guan et al [14] (Table 1) was developed to include recommendations we identified from the above process while excluding the information we gathered through the MARS framework. To determine app consistency with the GINA guidelines, participants were assessed for the presence or absence of features identified through this process. This is further discussed in subsequent sections.
### Identification, Screening, and Selection of Mobile Apps for Review

This review included both free and paid apps from the 2 most popular app stores in Australia across the iOS and Android operating systems: the Apple App Store (Apple) and Google Play Store (Google). Our published protocol outlines the steps taken for quality assurance [12]. Our approach for identifying these apps followed the approach used in similar studies [9-11]. Before commencing the initial search for apps, the reviewers ensured that the operating systems on the chosen smartphones were up to date. Each reviewer used different phones to assess the apps, but all updated the Android operating system (OS) to the Android 11 OS (Google) when reviewing apps from the Google Play Store. In the search bar of each store, we input the term *asthma*. Two reviewers (BR and DZ) independently searched both app stores on August 10, 2021, from Melbourne, Australia. After obtaining the results for this search term, each reviewer stored the information on an Excel (Microsoft Corp) spreadsheet (Multimedia Appendix 2). The reviewers then compared their results to ensure that they captured all the available apps.

For further evaluation of all the apps identified above, BR and DZ individually reviewed the app title, description, and attached photos and determined whether the app met all inclusion criteria and no exclusion criteria. An identified asthma app was included in the evaluation stage if all the following applied: (1) its primary role was related to asthma, (2) it was targeted to those with asthma, (3) it could be run on mobile phones, and (4) it was in English. Apps were excluded if any of the following criteria were not met.

### Table 1. Recommendations identified from the Global Initiative for Asthma guidelines that could be incorporated into a mobile health app.

<table>
<thead>
<tr>
<th>Reviewer 1</th>
<th>Reviewer 2</th>
<th>Consensus reached</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assess symptom control (eg, ACQ³)</td>
<td>Support for assessing symptom control for a 4-week period</td>
<td>Support for assessing symptom control for a 4-week period considering the frequency of asthma symptoms, night waking because of asthma, frequency of SABA² use, and any activity limitation because of asthma; uses recognized screening, symptom control or numerical asthma control tools, and tracks peak flow measurement</td>
</tr>
<tr>
<td>Ability to self-track symptoms with or without peak flow</td>
<td>__c</td>
<td>Encourages patients to track symptoms and peak flow measurements</td>
</tr>
<tr>
<td>Risk factors for future exacerbations</td>
<td>Helps users identify the future risk of exacerbations</td>
<td>Helps users identify the risk of future exacerbations</td>
</tr>
<tr>
<td>Screens for comorbidities and education regarding managing them</td>
<td>Screens for comorbidities and assists patients with managing them</td>
<td>Screens for relevant comorbidities and educates patients on the management of these comorbidities</td>
</tr>
<tr>
<td>Inhaler technique with or without video</td>
<td>Provides education on appropriate inhaler techniques</td>
<td>Provides education on appropriate inhaler techniques with visual aids</td>
</tr>
<tr>
<td>Ability to record action plan</td>
<td>Provides an area for patients to keep and refer to their written action plan</td>
<td>Provides an area for patients to keep and refer to their written action plan</td>
</tr>
<tr>
<td>Reminder to engage with primary care</td>
<td>Reminds users to see their HCP⁴ for management and review of asthma</td>
<td>Provides reminders to users to see their HCP for management and review of asthma</td>
</tr>
<tr>
<td>—</td>
<td>Specifically provides suggestion to see an HCP if a patient is using only a SABA.</td>
<td>Specifically provides suggestion to see an HCP if a patient is using a SABA alone</td>
</tr>
<tr>
<td>Medication adherence</td>
<td>Prompts users to adhere to controller medications even when symptoms are infrequent</td>
<td>Prompts users to adhere to controller medications even when symptoms are infrequent</td>
</tr>
<tr>
<td>General asthma education</td>
<td>Provides knowledge on general asthma information, management of asthma, modifiable risk factors and strategies to address them, and when to see an HCP</td>
<td>Provides knowledge on general asthma information, management of asthma, modifiable risk factors and strategies to address them, when to see an HCP, and identification and management of comorbidities</td>
</tr>
<tr>
<td>Help with activating action plan</td>
<td>Provides advice on when to refer to a patient’s asthma action plan based on self-monitoring of symptoms or PEF⁵</td>
<td>Provides advice on when to refer to a patient’s asthma action plan based on self-monitoring of symptoms or PEF</td>
</tr>
<tr>
<td>—</td>
<td>Prompts patient to see the primary HCP if features of asthma exacerbation (symptoms and SABA use) are identified using the app</td>
<td>Prompts patient to see the primary HCP if features of asthma exacerbation (symptoms and SABA use) are identified using the app</td>
</tr>
</tbody>
</table>

aACQ: Asthma Control Questionnaire.
bSABA: short-acting β-agonist.
cRecommendation identified by one reviewer but not the other.
dHCP: health care provider.
ePEF: peak expiratory flow.
applied: (1) they were not primarily related to asthma, (2) they were primarily targeted toward health care professionals (as stated in the product description), (3) they were not in English, and (4) they were targeted toward pediatric patients. This information was entered into an Excel (Microsoft Corp) spreadsheet for record-keeping (Multimedia Appendix 3). For further evaluation, all apps identified as meeting the above criteria were downloaded by a third reviewer (EP) who identified apps that did not install or function properly after downloading, eliminating them from the review. Finally, the last round of screening was conducted by the reviewer EP. In this round, duplications (i.e., apps available on both stores), inaccessible apps, and “lite” version apps, where a pro version was available, were removed from the review. This process was comparable with similar reviews that examined the quality of mobile apps for diabetes self-management [15].

**App Evaluation and Data Extraction**

**Reviewer Training**

A day-long training session was conducted before the initial data extraction. This training session was similar to the one performed by Gong et al [15] for their diabetes app review [15]. A step-by-step reference guide was created by the primary researchers to inform reviewers regarding how to complete the various frameworks and checklists involved in the study. This is provided in Multimedia Appendix 4 [3,4,16,17].

**App Evaluation and Data Extraction Overview**

An internet database was established on Qualtrics (Qualtrics International Inc) for data extraction. A total of 3 reviewers were involved in data collection. A web-based random team generator was used for all apps identified for further evaluation during the screening process so that each app was randomly allocated to 2 assessors (BR, DZ, or EP). The 2 assessors assigned to the app independently reviewed the in-store app description, downloaded the app, and used it for a minimum of 20 minutes to become familiar with all its functions [12]. The reviewers subsequently conducted the evaluations and entered the data into the Qualtrics database. Each reviewer performed this process individually without communicating their results to one another. There are 4 key aspects of the app evaluation and data extraction process, as summarized in the checklist provided in Multimedia Appendix 5 and the step-by-step guide to data collection in Multimedia Appendix 4.

**Technical Information About the App**

The first step in data collection involved gathering basic technical information about the app. The decision of which technical information to include was based on the MARS checklist and previous app review studies [7,9,15]. This was derived from publicly available information in the in-store app descriptions and in-app information sections. If required, the app developer’s website was used. The technical information collected included the app name, date of release, date of update, developer, developer affiliations, price, rating, number of ratings, platform or platforms, size of the app, and number of downloads. A checklist for this section is provided in Multimedia Appendix 5.

**App Quality Assessment**

The app quality assessment was completed using the MARS tool to objectively determine the quality of the apps selected. This scale has 4 separate domains that are assessed to evaluate mobile app quality. These domains are engagement, functionality, esthetics, and information quality [7]. A total of 19 items, each with a 5-point scale regarding quality in the 4 domains mentioned above, make up the MARS score [7]. This framework is presented in Multimedia Appendix 5. Reviewers completed this tool by entering the information into the Qualtrics (Qualtrics International Inc) checklist for each app. Once this was completed, the mean score for that domain and the overall MARS score were calculated for every app. Following the objective MARS section, there are several subjective questions to evaluate user satisfaction and the perceived impact of the app on the user’s knowledge, attitude, motivation to change, likelihood of change, and awareness of the importance of changing their asthma self-management [7]. These questions were answered by reviewers based on their experience using the app and their knowledge gained through the training sessions and clinical practice. Once the data were collected, the mean total MARS value and SD were calculated for each app.

The MARS tool assesses the presence or absence of 19 behavioral change techniques (BCTs). Although 93 types of BCTs are known, only the techniques outlined in MARS were assessed. This approach aligns with previous research, with a scope comparable with our review [7,10,15]. These techniques are outlined in Multimedia Appendix 4. To capture the presence of BCTs, an app was considered to have a BCT present, even if only 1 reviewer identified it. The median number of BCTs and the corresponding IQR of the apps was calculated using Stata (StataCorp) statistical software.

**App Functionality**

App functionality refers to what the app can do for a user and is an important marker of whether an app offers much benefit to users and the overall quality of the app. Although the MARS framework examines the overall quality of a mobile app, it focuses on the performance, ease of use, gestural design navigation, and navigation of the app [7]. Therefore, the Intercontinental Medical Statistics Institute for Health Informatics (IMS) Functionality Scoring System, henceforth known as the IMS functionality score, was used. This score was developed by the above institute and is based on 7 functionality criteria and 4 subcategories in the record functionality criterion. The IMS functionality score focuses on the scope of functions, including the ability of the app to inform, instruct, record, display, guide, remind or alert, and communicate information [18]. Each app was assessed for having or not having these functions and then given a total functionality score between 0 and 11 [18]. To capture the presence of all functionalities, an app was considered to have a functionality present even if one reviewer identified it. The mean, median, and IQR were calculated. The reviewers assessed each app for these functions and entered the data into Qualtrics (Qualtrics International Inc) database. This scoring system is presented in Multimedia Appendix 5.
Presence of App Features Consistent With Asthma Guidelines

As discussed above, key recommendations that could feasibly be incorporated into an asthma mHealth app were identified from the GINA guidelines. These recommendations, summarized in Table 1, were used to develop a more extensive checklist provided in Multimedia Appendix 5. The main functions of the app that we were interested in assessing in our checklist included asthma information, self-management skill training (including peak flow use, inhaler technique, and nonpharmacological strategies), monitoring of asthma symptoms, risk evaluation, and prompting (medication reminders, referring to action plan reminders, and suggestions for seeking health advice). Each of the selected apps was assessed using this checklist, and the data were entered into the Qualtrics (Qualtrics International Inc) database. To ensure consistency, an app was only assessed for the presence of the above function if both reviewers reached a consensus that the said feature was present. For single-reviewed those apps, the sole reviewer’s analysis was used to determine whether the app did or did not have the examined feature.

Quality Assurance, Data Management, and Data Analysis

Training was provided to all the researchers, and a handbook for reviewers was provided to ensure the quality of this research. Selected apps were allocated to reviewers using a web-based random allocation software, and 2 different major app databases were searched to reduce selection bias. The apps were independently reviewed by 2 reviewers to reduce the likelihood of bias affecting the results. A protocol was published to reduce publication bias and enhance the transparency of this study [12]. During app evaluation, all data were entered into either an Excel (Microsoft Corp) spreadsheet during the screening process or into the web-based Qualtrics (Qualtrics International Inc) database. These were stored on a cloud-based system that only the researcher team could access. Once the evaluation was completed, all data were downloaded for subsequent analysis. This analysis comprised a descriptive analysis and calculation of the mean and SD or median and IQR.

All data analyses were performed using Stata statistical software version 14 (StataCorp). Visual figures were generated using Excel (version 16; Office 365; Microsoft Corp).

Results

Identification, Screening, and Selection of Mobile Apps for Review

The process and results of identifying, screening, and selecting mobile apps are shown in Figure 1. A total of 174 unique apps from the Apple Store and 248 unique apps from the Google Play Store were identified. These 422 apps were assessed by 2 reviewers (BR and DZ). In total, 94 apps met all inclusion criteria and no exclusion criteria, although there was a discrepancy between the reviewers’ opinions regarding the eligibility of 39 apps. A third reviewer (EP) identified 17 of these 39 apps as suitable for further review, which resulted in a total of 111 apps suitable for further assessment. A total of 40 apps were removed for reasons outlined in Figure 1. When a more recently updated app was available on 1 platform compared with the other, the older version was excluded from the review. When duplicate apps were available on both platforms and had been updated on the same date, the app from the Apple App Store was retained, whereas the app from the Google Play Store was excluded from the review. This was done to ensure consistency between reviewers and prevent skewing of results by assessing the same app twice. This resulted in 71 apps that were suitable for a complete assessment. Given the delay between the identification of apps and analysis, 18 of the above 71 apps were no longer assessable for the reasons outlined in Figure 1. A total of 4 apps were downloaded by 1 reviewer but not the other before they were removed from the market. This resulted in 49 apps assessed by 2 reviewers and 4 apps assessed by 1 reviewer (53 apps and 102 total reviews).
Technical Information About the App

The technical information for each reviewed app can be found in Multimedia Appendix 2. Of the 53 apps assessed, 29 (55%) were from the Apple App Store, and 24 (45%) were from the Google Play Store. A total of 19 (36%) apps downloaded from the Apple App Store were also available on Google Play Store. As outlined above, apps available on both marketplaces were only downloaded from the Apple App Store and assessed on the iOS platform. The apps’ last date of update ranged from February 2016 to April 2022. A total of 26 (49%) apps were updated from January 2020. The mean app size was 46.33 MB, and the median app size was 27 MB (IQR 9.2 MB–47.38 MB). App developers were primarily technical companies (28 apps), health care or pharmaceutical companies (4 apps), or a combination of both (4 apps). Six apps were created by private individuals, 3 were created from research or clinical institutions and the remaining 8 were created from developers from a variety of other backgrounds. The number of app downloads ranged from 10 to >10,000. Of 53 apps, 24 (45%) apps had a published user rating, and the median rating was 4 out of 5 (IQR 2.9–4.9). The number of people who provided a rating ranged from 0 to 523, with a median of 2.5 ratings per app (IQR 1–20). Of 53 apps, 42 (79%) apps were completely free, 5 (9%) apps required...
users to pay to download, and 6 (11%) of the above free apps had the in-app ability to upgrade for a cost.

**Presence of Behavior Change Techniques**

The total number of each BCT identified across the 53 apps assessed is demonstrated in Table 2. The most frequent BCT observed was self-monitoring or tracking, which was identified in 38 (72%) of the 53 apps. The next commonly identified BCTs were information or education, seen in 33 (62%) apps, and advice, tips, strategies, or skills training, seen in 26 (49%) apps. The average number of BCTs in a single app from those reviewed was 3.26 (SD 2.27). The median number of BCTs for the apps reviewed was 3 (IQR 1-4).

**Table 2.** Assessed behavioral change techniques and the number of apps found to use these techniques (n=53).

<table>
<thead>
<tr>
<th>Behavioral change technique</th>
<th>Apps with this technique, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information or education</td>
<td>33 (62)</td>
</tr>
<tr>
<td>Self-monitoring or tracking</td>
<td>38 (72)</td>
</tr>
<tr>
<td>Advice or tips or strategies or skills training</td>
<td>26 (49)</td>
</tr>
<tr>
<td>Assessment</td>
<td>17 (32)</td>
</tr>
<tr>
<td>Feedback</td>
<td>12 (23)</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>9 (17)</td>
</tr>
<tr>
<td>Model or demonstrate behavior</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Rewards and self-rewards</td>
<td>6 (11)</td>
</tr>
<tr>
<td>Goal setting</td>
<td>5 (9)</td>
</tr>
<tr>
<td>Provide social support</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Perceived risks</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Model or demonstrate behavior</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Action planning</td>
<td>11 (21)</td>
</tr>
<tr>
<td>Motivation</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Motivational readiness</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Mindfulness or meditation</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Perceived benefit</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

**App Quality (MARS)**

The mean MARS score for all reviewed apps was 3.05 (SD 0.54). Of the 53 apps, 27 (51%) achieved a total MARS score of ≥3. A score of 3 on the MARS tool correlates to an “acceptable” quality app, <3 is inadequate or poor quality, and >3 represents a good or exceptional app [7]. Functionality was the highest rated MARS category with a mean score of 3.85 (SD 0.52), followed by esthetics with a mean score of 3.21 (SD 0.6). The information category had an average score of 2.78 (SD 0.83), and engagement had a mean score of 2.77 (SD 0.59). Table 3 shows the mean score for each of the 19 items on the MARS tool. Notably, the apps reviewed had higher scores in the gestural design, app description accuracy, and ease of use domains and lower scores in the evidence base, credibility, and entertainment domains. The mean scores for the quality and quantity of information were 3.21 (SD 1.95) and 2.68 (SD 1.63), respectively.

The final component of MARS allows reviewers to complete a subjective assessment of their opinions on the app. For double-reviewed apps, the mean value for the level of agreement for each domain was first calculated. Therefore, the number of apps will not be a whole number. This is summarized in Table 4 and demonstrates that there were few apps that reviewers would recommend to others, use >2 times in a 12-month period, or pay for. Only 8% (n=4) of the apps were rated >3 stars by the reviewers. For >50% of the apps, reviewers either disagreed or strongly disagreed that the app would impact the user’s knowledge, attitudes, and intentions to change or change the rate of asthma exacerbations. For <20% of the apps, reviewers either strongly agreed or agreed that the app would improve the user’s knowledge, attitudes, awareness, or intention to change behaviors to improve asthma self-management. However, reviewers either strongly agreed or agreed that 29% (n=15) and 27% (n=14) of apps would encourage users to seek further help in asthma management and reduce asthma exacerbations, respectively.
Table 3. Mean score for each category in the Mobile App Rating Scale (MARS) tool for the 53 assessed apps. Each category is assessed on a 5-point scale.

<table>
<thead>
<tr>
<th>MARS category</th>
<th>Mean score for category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entertainment</td>
<td>2</td>
</tr>
<tr>
<td>Interest</td>
<td>2</td>
</tr>
<tr>
<td>Customization</td>
<td>2</td>
</tr>
<tr>
<td>Interactivity</td>
<td>2</td>
</tr>
<tr>
<td>Target group</td>
<td>3</td>
</tr>
<tr>
<td>Performance</td>
<td>3</td>
</tr>
<tr>
<td>Ease of use</td>
<td>3</td>
</tr>
<tr>
<td>Navigation</td>
<td>3</td>
</tr>
<tr>
<td>Gestural design</td>
<td>4</td>
</tr>
<tr>
<td>Layout</td>
<td>3</td>
</tr>
<tr>
<td>Graphics</td>
<td>3</td>
</tr>
<tr>
<td>Visual appeal</td>
<td>2</td>
</tr>
<tr>
<td>Accuracy of app description</td>
<td>4</td>
</tr>
<tr>
<td>Goals</td>
<td>3</td>
</tr>
<tr>
<td>Quality of information</td>
<td>3</td>
</tr>
<tr>
<td>Quantity of information</td>
<td>2</td>
</tr>
<tr>
<td>Visual information</td>
<td>2</td>
</tr>
<tr>
<td>Credibility</td>
<td>1</td>
</tr>
<tr>
<td>Evidence base</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 4. Results from the subjective assessment section of the Mobile App Rating Scale framework (n=53).

<table>
<thead>
<tr>
<th>Would you recommend this app to people who might benefit from it?</th>
<th>Apps, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all, I would not recommend this app to anyone.</td>
<td>12 (23)</td>
</tr>
<tr>
<td>There are very few people I would recommend this app to.</td>
<td>18 (35)</td>
</tr>
<tr>
<td>Maybe, there are several people whom I would recommend it to.</td>
<td>18 (35)</td>
</tr>
<tr>
<td>There are many people I would recommend this app to.</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Definitely, I would recommend this app to everyone.</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How many times do you think you would use this app in the next 12 months?</th>
<th>Apps, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4 (8)</td>
</tr>
<tr>
<td>1-2</td>
<td>28 (52)</td>
</tr>
<tr>
<td>3-10</td>
<td>7 (13)</td>
</tr>
<tr>
<td>11-50</td>
<td>14 (27)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Would you pay for this app?</th>
<th>Apps, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>34 (64)</td>
</tr>
<tr>
<td>Yes</td>
<td>18 (35)</td>
</tr>
<tr>
<td>Maybe</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What is your overall star rating for the app?</th>
<th>Apps, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (one of the worst apps I have used)</td>
<td>10 (19)</td>
</tr>
<tr>
<td>2</td>
<td>20 (38)</td>
</tr>
<tr>
<td>3 (average)</td>
<td>18 (35)</td>
</tr>
<tr>
<td>4</td>
<td>3 (6)</td>
</tr>
<tr>
<td>5 (one of the best apps I have used)</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strongly disagree or disagree that the app will improve</th>
<th>Apps, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness</td>
<td>40 (75)</td>
</tr>
<tr>
<td>Knowledge</td>
<td>37 (69)</td>
</tr>
<tr>
<td>Attitudes</td>
<td>32 (60)</td>
</tr>
<tr>
<td>Intention to change</td>
<td>32 (60)</td>
</tr>
<tr>
<td>Help seeking</td>
<td>27 (50)</td>
</tr>
<tr>
<td>Behavior change (reduce asthma exacerbations)</td>
<td>29 (54)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strongly agree or agree that the app will improve</th>
<th>Apps, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness</td>
<td>8 (15)</td>
</tr>
<tr>
<td>Knowledge</td>
<td>10 (19)</td>
</tr>
<tr>
<td>Attitudes</td>
<td>7 (13)</td>
</tr>
<tr>
<td>Intention to change</td>
<td>6 (12)</td>
</tr>
<tr>
<td>Help seeking</td>
<td>15 (29)</td>
</tr>
<tr>
<td>Behavior change (reduce asthma exacerbations)</td>
<td>14 (27)</td>
</tr>
</tbody>
</table>

**IMS Functionality Score**

Out of a potential 11 functionalities, an average IMS functionality score of 5.1 (SD 2.79) was achieved by the reviewed apps. The median number of functionalities identified was 5 (IQR 2-7). A total of 3 apps had 11 functionalities, although most apps had the ability to capture user-entered data (n=37, 70%), provide information in a variety of formats (n=35, 66%), and provide instructions to the user (n=33, 62%). Out of 53 apps, 7 (13%) apps had the ability to send alerts or propose behavioral interventions based on the collected data. The total number of apps that met these IMS functionality criteria is summarized in Table 5.
Table 5. Total number of apps meeting each of the Intercontinental Medical Statistics Institute for Health Informatics (IMS) functionality criteria (n=53).

<table>
<thead>
<tr>
<th>IMS functionality</th>
<th>Apps containing this functionality, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inform</td>
<td>35 (66)</td>
</tr>
<tr>
<td>Instruct</td>
<td>33 (62)</td>
</tr>
<tr>
<td>Record</td>
<td>37 (70)</td>
</tr>
<tr>
<td>Display</td>
<td>30 (57)</td>
</tr>
<tr>
<td>Guide</td>
<td>11 (21)</td>
</tr>
<tr>
<td>Remind or alert</td>
<td>27 (51)</td>
</tr>
<tr>
<td>Communicate</td>
<td>17 (32)</td>
</tr>
<tr>
<td>Collect data</td>
<td>37 (70)</td>
</tr>
<tr>
<td>Share data</td>
<td>26 (49)</td>
</tr>
<tr>
<td>Evaluate data</td>
<td>12 (23)</td>
</tr>
<tr>
<td>Intervene</td>
<td>7 (13)</td>
</tr>
</tbody>
</table>

Presence of App Features Consistent With Asthma Guidelines

Knowledge

Table 6 outlines the results of evaluating the presence of information or knowledge identified as important based on the GINA guidelines. This details the number of apps where reviewer consensus was achieved, the number of apps that provided knowledge or did not provide knowledge in the subcategories, whether knowledge was individualized, and whether knowledge was based on evidence.

Table 6. Number of apps providing knowledge in asthma in the various relevant domains identified as important based on the Global Initiative for Asthma guidelines (n=53).

<table>
<thead>
<tr>
<th>Reviewer consensus achieved, n (%)</th>
<th>Apps that provided knowledge in this domain, n (%)</th>
<th>Apps that did not provide knowledge in this domain, n (%)</th>
<th>Apps that provided individualized knowledge, n (%)</th>
<th>Apps that provided evidence-based knowledge, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General asthma knowledge</td>
<td>45 (85)</td>
<td>10 (22)</td>
<td>35 (78)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asthma medications</td>
<td>46 (87)</td>
<td>13 (28)</td>
<td>33 (72)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Exacerbation management</td>
<td>44 (83)</td>
<td>14 (32)</td>
<td>30 (68)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asthma risk factors and triggers</td>
<td>46 (87)</td>
<td>11 (24)</td>
<td>35 (76)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Skill Training

The number of apps that provide skills training in peak flowmeter use, inhaler device use, recognizing and responding to asthma exacerbations, and nonpharmacological management to reduce asthma exacerbations are summarized in Table 7.
Table 7. Number of apps which provide the specific skill training in the areas identified as important in the Global Initiative for Asthma guidelines (n=53).

<table>
<thead>
<tr>
<th>Apps with reviewer consensus, n (%)</th>
<th>Apps that provide skill training, n (%)</th>
<th>Apps which provide personalized skill training, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>App provides general skill training in peak flowmeter use</strong></td>
<td>46 (87)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Describes why and when to use peak flowmeter</td>
<td>0 (0)</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Describes operational criteria for peak flowmeter</td>
<td>0 (0)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Demonstrates the use of peak flow meter through photos or videos</td>
<td>0 (0)</td>
<td>5 (11)</td>
</tr>
<tr>
<td><strong>App provides general skill training in inhaler device use</strong></td>
<td>49 (93)</td>
<td>12 (25)</td>
</tr>
<tr>
<td>Describes how to use a spacer</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Demonstrates how to use a spacer through videos or photos</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Demonstrates how to care for a spacer</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Describes how to use common inhaler devices</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Demonstrates how to use common inhaler devices through videos or photos</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>App provides general skill training in recognizing and responding to exacerbations</strong></td>
<td>39 (74)</td>
<td>11 (28)</td>
</tr>
<tr>
<td>Encourages patients to monitor for signs of asthma exacerbation</td>
<td>0 (0)</td>
<td>9 (23)</td>
</tr>
<tr>
<td>Provide an area for asthma action plan</td>
<td>0 (0)</td>
<td>5 (13)</td>
</tr>
<tr>
<td>Specifically guides patients to use their asthma action plan</td>
<td>0 (0)</td>
<td>3 (8)</td>
</tr>
<tr>
<td>Provide information on how to use an asthma action plan</td>
<td>0 (0)</td>
<td>7 (18)</td>
</tr>
<tr>
<td>Prompts patient to see health care provider when required</td>
<td>0 (0)</td>
<td>7 (18)</td>
</tr>
<tr>
<td><strong>App provides general skill training in nonpharmacological management strategies to reduce asthma exacerbations</strong></td>
<td>45 (85)</td>
<td>14 (31)</td>
</tr>
<tr>
<td>Helps identify triggers that make symptoms worse</td>
<td>0 (0)</td>
<td>9 (20)</td>
</tr>
<tr>
<td>Advises avoidance of environmental smoke exposure</td>
<td>0 (0)</td>
<td>13 (29)</td>
</tr>
<tr>
<td>Advises avoidance of medications that can worsen asthma</td>
<td>0 (0)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Advises avoidance of occupation exposures</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Advises on the avoidance of allergen exposure</td>
<td>0 (0)</td>
<td>13 (29)</td>
</tr>
<tr>
<td>Advises on avoidance of indoor or outdoor pollution</td>
<td>0 (0)</td>
<td>12 (27)</td>
</tr>
<tr>
<td>Advises on avoidance of emotional stress</td>
<td>0 (0)</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Advises on regular moderate physical activity</td>
<td>0 (0)</td>
<td>6 (13)</td>
</tr>
</tbody>
</table>

**App’s Ability to Track and Display Health Information**

Table 8 demonstrates the results of the assessment of whether apps had the ability to track and display different aspects of a user’s key asthma information. All information is only related to apps where reviewer consensus was achieved or those that were reviewed by a single researcher. This demonstrates that most apps did not support tracking of all relevant asthma data, and for those that did, manual data input was the predominant entry method.
Table 8. Number of apps which allowed tracking and displaying of the specified asthma information identified as important from the Global Initiative for Asthma guidelines and by what means this information could be input into the app (n=53).

<table>
<thead>
<tr>
<th>App’s ability to track and display users’ asthma information</th>
<th>Apps with reviewer consensus, n (%)</th>
<th>Apps that allow tracking and recording of data, n (%)</th>
<th>Apps with manual data input, n (%)</th>
<th>Apps that allow data entry through external sensors or devices, n (%)</th>
<th>Apps with the ability to create tables or graphs demonstrating trends or analysis of entered data, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma symptoms</td>
<td>51 (96)</td>
<td>19 (37)</td>
<td>19 (37)</td>
<td>1 (2)</td>
<td>14 (27)</td>
</tr>
<tr>
<td>Night waking because of asthma</td>
<td>49 (92)</td>
<td>4 (8)</td>
<td>4 (8)</td>
<td>0 (0)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Activity limitation because of asthma</td>
<td>47 (89)</td>
<td>3 (6)</td>
<td>3 (6)</td>
<td>0 (0)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Peak flow meter values</td>
<td>50 (94)</td>
<td>20 (40)</td>
<td>17 (34)</td>
<td>4 (8)</td>
<td>20 (40)</td>
</tr>
<tr>
<td>SABA use</td>
<td>46 (87)</td>
<td>14 (30)</td>
<td>14 (30)</td>
<td>3 (7)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Preventive medication adherence</td>
<td>48 (91)</td>
<td>11 (23)</td>
<td>11 (23)</td>
<td>3 (6)</td>
<td>8 (17)</td>
</tr>
</tbody>
</table>

*SABA: short-acting β-agonist.*

App’s Ability to Provide Prompts or Reminders

Table 9 demonstrates the results of the assessment of whether the apps could provide reminders or prompts on the areas deemed relevant from the review of the GINA guidelines, as outlined in the Methods section. Overall, there were few apps that provided reminders or prompts to users, with only 9 (20%) out of 53 apps providing a reminder to use a person’s daily preventive medication. Few apps prompted users to assess the severity and frequency of their asthma symptoms or to seek health advice based on the data they entered.

Table 9. Number of apps that provided reminders and prompts on the specified asthma features chosen based on Global Initiative for Asthma guidelines (n=53).

<table>
<thead>
<tr>
<th>App provides reminders or prompts on asthma features</th>
<th>Apps with reviewer consensus, n (%)</th>
<th>Apps that provide reminders or prompts, n (%)</th>
<th>Apps where reminders or prompts can be individualized, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessing asthma symptoms for the last month</td>
<td>50 (94)</td>
<td>2 (4)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Appointment with physicians</td>
<td>47 (89)</td>
<td>1 (2)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Performing peak flow test</td>
<td>50 (94)</td>
<td>5 (10)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Preventive medication adherence</td>
<td>45 (85)</td>
<td>9 (20)</td>
<td>8 (18)</td>
</tr>
<tr>
<td>Checking the date of expiry and dosage of inhalers</td>
<td>51 (96)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Warning of changing health data (ie, very frequent exacerbations)</td>
<td>53 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Seeking urgent health advice based on the health data the user inputs into the app</td>
<td>52 (98)</td>
<td>2 (4)</td>
<td>2 (4)</td>
</tr>
</tbody>
</table>

Other App Information

Table 10 summarizes the other important features assessed during this review. Only 1 app allowed the user to make an appointment with a physician. Of import, only 5 (10%) of the apps were identified as having an area where users could keep a record of their asthma action plan. Most of these (4 apps) allowed the user to type in their action plan, and 1 app allowed users to upload a photo of their action plan. Only 1 app was identified using recognized asthma screening tools to assess patient’s current asthma symptoms and severity.
Table 10. Summary of number of apps containing further features and content identified as important (n=53).

<table>
<thead>
<tr>
<th>Feature Description</th>
<th>Reviewer consensus, n (%)</th>
<th>Apps with this feature, n (%)</th>
<th>Apps without this feature, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make appointment with physicians</td>
<td>53 (100)</td>
<td>1 (2)</td>
<td>52 (98)</td>
</tr>
<tr>
<td>For web-based consultation</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>53 (100)</td>
</tr>
<tr>
<td>For face-to-face consultation</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>52 (98)</td>
</tr>
<tr>
<td>App provides an area for patients to keep record of their asthma action plan</td>
<td>52 (98)</td>
<td>5 (10)</td>
<td>47 (90)</td>
</tr>
<tr>
<td>Can type in action plan</td>
<td>0 (0)</td>
<td>4 (8)</td>
<td>48 (92)</td>
</tr>
<tr>
<td>Can upload a photo of action plan</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>51 (98)</td>
</tr>
<tr>
<td>The app includes social forums or blogs that promote peer-support and communication among asthma patients</td>
<td>50 (94)</td>
<td>1 (2)</td>
<td>49 (98)</td>
</tr>
<tr>
<td>The users could send recorded data to others</td>
<td>44 (83)</td>
<td>11 (25)</td>
<td>33 (75)</td>
</tr>
<tr>
<td>To physicians</td>
<td>0 (0)</td>
<td>11 (25)</td>
<td>33 (75)</td>
</tr>
<tr>
<td>To each other</td>
<td>0 (0)</td>
<td>9 (21)</td>
<td>33 (75)</td>
</tr>
<tr>
<td>The app could help the users to evaluate the risk of having future asthma exacerbations</td>
<td>44 (83)</td>
<td>2 (5)</td>
<td>42 (96)</td>
</tr>
<tr>
<td>Using a validated scoring system</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>44 (100)</td>
</tr>
<tr>
<td>Yes, but not using a validated scoring system</td>
<td>0 (0)</td>
<td>2 (5)</td>
<td>42 (96)</td>
</tr>
<tr>
<td>The app could guide the users to find out the closest pharmacy, hospital, or clinic</td>
<td>51 (96)</td>
<td>0 (0)</td>
<td>51 (96)</td>
</tr>
<tr>
<td>The app uses recognized screening, symptom control and numerical asthma control tools</td>
<td>46 (87)</td>
<td>1 (2)</td>
<td>45 (98)</td>
</tr>
<tr>
<td>Yes, all 3</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>46 (100)</td>
</tr>
<tr>
<td>Yes, screening tool</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>46 (100)</td>
</tr>
<tr>
<td>Yes, symptom control tool</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>46 (100)</td>
</tr>
<tr>
<td>Yes, numerical asthma control tools</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>45 (98)</td>
</tr>
<tr>
<td>The app allows users to connect to wearable technology</td>
<td>53 (100)</td>
<td>8 (15)</td>
<td>45 (85)</td>
</tr>
<tr>
<td>Smartwatch</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>53 (100)</td>
</tr>
<tr>
<td>Activity sensor (eg, Fitbit)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>53 (100)</td>
</tr>
<tr>
<td>Smart peak flowmeter</td>
<td>0 (0)</td>
<td>4 (8)</td>
<td>49 (93)</td>
</tr>
<tr>
<td>Handheld spirometer</td>
<td>0 (0)</td>
<td>3 (6)</td>
<td>50 (94)</td>
</tr>
<tr>
<td>Oximeter</td>
<td>0 (0)</td>
<td>2 (4)</td>
<td>51 (96)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>6 (11)</td>
<td>47 (89)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

This review aimed to evaluate the quality and functionality of asthma apps and their consistency with international best practice guidelines. We conducted a comprehensive review of 422 asthma apps available on the Australian App Store and Google Play Store, of which we selected 53 apps that met our inclusion and exclusion criteria for detailed analysis. The most common reason for app exclusion was that they were not primarily related to asthma. Most apps were developed by technical companies rather than health care facilities and clinical or research institutes. This lack of involvement of practicing experts in the field of health care is concerning, as these apps are primarily related to the management of chronic health conditions. Apps that involved a clinician during the designing phase demonstrated a greater ability to facilitate behavioral change than those that did not [19]. We believe that in the future, asthma mHealth apps should be developed in consultation with health care professionals and organizations to ensure that they meet an appropriate standard.

Our review revealed that most asthma apps do not use key BCTs that can promote behavioral changes through feedback, goal setting, and rewards. Instead, the most commonly used BCTs were self-monitoring or tracking, education, and advice. This is consistent with other studies and demonstrates a potential area where asthma apps could be improved in the future [19,20]. Functionalities describe what an app can do for its user and are an important marker of whether an app offers much benefit to users. Our review demonstrated that although basic functionalities, such as informing, reading, collecting, instructing, and displaying, were prominent, more complex functionalities were lacking. The ability to evaluate and intervene based on the app information entered is not present
in most apps. Creating apps with this functionality could, for example, guide patients to see their health care provider based on the data they are entering, such as excessive short-acting β-agonist use or asthma symptoms. This feature was demonstrated in the Asthmahub app from NHS Wales, one of the higher-rated apps on assessment.

Our evaluation using the MARS tool showed that the asthma mHealth apps performed poorly in the information and engagement domains compared with the functionality and esthetics categories. These findings are consistent with previous asthma app reviews that showed poorer results in these categories [9]. Only half of the apps achieved an acceptable standard, and even fewer achieved a dual rating average >4, indicating a “good-quality” app. These apps were Kiss my Asthma, Asthmahub, and AioCare Patient. Acceptable or good ratings for subcategories such as app credibility, evidence base, and entertainment were particularly lacking in the apps we assessed. In contrast to this lack of evidence-based content identified through app assessment with the MARS framework, assessment with our checklist found that the knowledge presented in apps was largely based on evidence. “Evidence-based” in the MARS checklist refers to whether an app has undergone a clinical trial. By contrast, in our checklist we refer to “evidence-based” as that the information presented in the app was factual and in alignment with our clinical knowledge of asthma and the GINA guidelines. This explains the contrasting results between the MARS framework and our checklist for “evidence-based” knowledge. Although information was often factual and based on guidelines, the apps had not undergone clinical trials. The subjective star rating provided by reviewers was rarely >3, whereas the mean user rating was 3.56 out of 5, indicating a discrepancy between the perspectives of the reviewers and the user. This discrepancy in user ratings and reviewers’ perspectives has been demonstrated in similar app reviews in the past [9]. We propose that this could be because of our health care background bias when assessing the apps, even though we were attempting to assess apps from the perspective of a patient with asthma, or because we were approaching the review with a critical lens following an objective app assessment, a different mindset from the usual user.

Our review demonstrated that asthma mHealth apps do not contain key features consistent with international best practice guidelines for asthma management. Few apps contained important information regarding asthma and asthma management, and even when provided, they were not individualized to the user. Personalization is a key part in the management of asthma, with an individual’s triggers, baseline respiratory function, and inhaler device being common things we assess for, educate on, and consider when managing asthma in the community. Furthermore, personalized therapeutic management is a key aspect of asthma management, as outlined in the GINA guidelines [4]. There is also an expectation from previous research that personalized technological interventions may lead to better health outcomes, although this has not yet been specifically demonstrated for asthma mHealth apps [21]. The NHS Wales AsthmaHub app provides one solution to this problem by starting with the creation of a profile in which the user answers questions regarding asthma triggers, control, and medications. This allows the app to provide information targeted toward the user. Skill training in a field can be achieved by apps through written and visual information to teach users how to use peak flow meters, inhalers, and nonpharmacological strategies to manage asthma and asthma exacerbation. These are skills that the GINA guidelines advocate patients to learn and become proficient in, with the support of their health care provider [4]. Training in the use of inhaler devices, spacers, and peak flow meters, all of which in our experience, patients can inadvertently misuse, is lacking in most apps and is something that we feel is key to be included in an asthma mHealth app. Although not essential for all inhaler types, such as dry-powder inhalers, spacers are considered an essential part of using metered-dose inhaler preventer and reliever medications. Therefore, we believe it is reasonable that a high-quality asthma mHealth app should include advice to the users regarding the use of spacers, particularly given the prevalence of salbutamol or albuterol metered-dose inhaler use among those with asthma. No apps could individualize this training to the user, and given the magnitude of devices on the market and the difference in how they work, this represents a deficiency. In total, 40% (n=20) of the apps could record and display peak flow meter values and asthma symptoms. Peak flow is representative of worsening airway obstruction or asthma. If this value is reducing, it can indicate that a person’s asthma may be on the precipice of or in the middle of an exacerbation, and as such, it can be a valuable metric for patients and clinicians to monitor. Other key metrics that could be tracked were lacking, such as the amount of short-acting β-agonists and adherence to preventive medications. It is important for these metrics to be recorded, as excessive SABA use can indicate an asthma exacerbation or poorly controlled asthma and should prompt a review by a health care professional. Subtler signs of poor asthma control, such as activity limitation or night waking because of asthma, were rarely recorded and represented missed opportunities. Prompting and reminding users to do something is a basic functionality that is largely underutilized by apps. Chronic asthma management involves remembering to consume daily inhaler medications, assessing symptoms, and regular interactions with a health care provider [4]. All these tasks lend themselves to a reminder from an app to assist with asthma management, and their absence from most apps is a missed opportunity.

Asthma action plans are a key aspect of contemporary patient self-management of asthma [4]. The ability to quickly refer to this plan on a user’s digital device should be easy to include; however, most apps lack this feature. We see this as a significant deficiency and missed opportunity. Few apps use validated asthma symptom scoring systems for users to assess their symptoms and risk of exacerbation. Wearable technology and external sensors are a growing medium through which data related to a patient’s asthma status can be captured, yet only few apps use this technology. When external sensors are used, they are often smart spirometers (such as in AioCare patient) or peak flow meters, both of which carry extra costs and may not be palatable to all users. No apps were found to be linked to smartwatches or activity sensors, the use of which is becoming more prominent [22]. With the availability of pulse...
oximetry in smartwatches, this may be a new method for asthma apps to collect crucial data in the future.

Overall, we determined the quality of apps to be average at best, with many lacking features consistent with international best practice guidelines. Three apps achieved a MARS rating of >4. These were Kissmyasthma, Asthmahub, and AioCare patient. Their alignment of these apps with international best practice guidelines was mixed, with AioCare patients not having many of the functions deemed important by the guidelines. The Kissmyasthma and Asthmahub apps had great amounts of information related to asthma, consistent with the guidelines. Asthmahub stood alone in having these features and many features to support skill training and track and record information; however, its prompting or reminding functions were minimal. Notably, both these apps that ranked higher in MARS and our checklist were developed by health services or medical research centers (NHS Wales, the University of Sydney, The University of Melbourne, and the Woolcock Institute of Medical Research). All 11 functionalities from the IMS scale were identified in Asthmahub and AioCare patients, and Kissmyasthma had 6 functionalities, which were above the mean identified. A total of 6 to 8 BCTs were identified in these apps, which, although greater than the median number of BCTs, still did not include several potential BCTs.

Comparison With Previous Work

Prior studies that have conducted these assessments have primarily focused on evaluating the quality and functionality of apps using the MARS framework [9-11]. From a review of the literature for the past 5 years, only 2 prior studies were found to have conducted some sort of assessment of the alignment of apps with asthma self-management principles. Both these studies only looked at free apps, eliminating a number of apps from the review [11,23]. Data collection for both reviews occurred years ago [11,23]. In the rapidly developing marketplace of mobile apps, a number of new apps have been released during this time. Our review examined both free and paid apps, and provided an updated assessment, given that our data collection took place in 2022. Furthermore, Househ et al [23] did not assess apps from the Apple App Store, focusing only on the Google Play Store, and therefore, did not fully assess the breadth of available English-language apps in the marketplace. These authors evaluated whether apps included or did not include information consistent with GINA guidelines as per a checklist created by 1 author [23]. This was limited to asthma information and education and did not include further features, such as the ability of an app to track information, provide asthma skill training, or personalize information. This review also did not examine the overall app quality using the validated MARS framework [23].

Our review benefits from having 2 independent clinicians review the guidelines to establish all GINA self-management recommendations that can be feasibly incorporated into an mHealth app and review app quality using the MARS framework. Furthermore, we examined not only the presence of information, but also the presence or absence of the ability to track asthma symptoms and provide reminders and skill training, as well as all features derived from the GINA guidelines provided in Multimedia Appendix 5. As part of their app review, Tan et al [11] established a framework for assessing the alignment of mHealth apps with the theoretical principles of self-management of allergic rhinitis or asthma [11]. A total of 6 asthma self-management principles were identified based on a literature review and author consensus [11]. Our review has taken this a step further, specifically deriving self-management principles from the international best practice GINA guidelines and creating a more extensive checklist based on these principles. The inclusion of paid apps, the creation of an asthma self-management principle checklist derived from international best practice guidelines, and the up-to-date nature makes our study unique.

Limitations

This study has several limitations. First, we limited it to English-language apps available from 2 app stores in Australia. Although these stores make up a significant portion of the market, other stores do exist, such as the Blackberry store, which were not included in the review. It is also expected that some apps available on the Australian marketplace may not be available on international marketplaces, whereas some apps available internationally may not be available on the Australian marketplace, and thus not included in the review. Although the operating systems for phones used by reviewers were updated on the same day to ensure that the same OS was on each smartphone, the phones themselves were different models. This may have affected the user or reviewer experience to an unknown degree, reducing standardization in this study. A total of 4 apps were reviewed by 1 researcher only as they were removed from the market or were not available to the second reviewer by the time they tried to assess it. This may skew some results, although previous studies have only single-reviewed apps; therefore, the fact that the most apps in this study are double-reviewed is a strength of this research.

Furthermore, we only assessed the presence or absence of those BCTs embedded in the MARS checklist. This limited the number of BCTs that we assessed; however, reviewing the presence or absence of all 93 BCTs was outside the scope of this review and our protocol and requires further research. This evaluation was researcher-based, and thus does not reflect the real assessment from a patient’s perspective. Future research should also include people with asthma to determine their responses to the quality and functionality of these apps.

Conclusions

Currently, there is a lack of high-quality asthma apps aligned with international best practice asthma guidelines. Most apps do not provide patients with important asthma information, skills training in key aspects of asthma management, recording and tracking relevant data, or reminders regarding asthma control. The lack of interaction with smart technology or use of asthma action plans are significant flaws that merit attention in future apps. Few BCTs or in-depth functionalities have been deployed to engage and re-engage with users or generate meaningful behavioral modifications. Again, we see that app designers are typically skilled in providing an esthetically pleasing functional app but lack skills in providing engagement and information, as assessed in the MARS tool. Future app
development should target the key features identified in this study as currently lacking. Furthermore, not only are high-quality asthma mHealth apps lacking, there are minimal robust clinical trials examining the use of these apps and their effect on patient outcomes. Further research in this area will be valuable to determine the true clinical utility of these apps.

A wide spectrum of technological quality, accuracy, and breadth of health information was seen among available apps. Given this spectrum of poor-to high-quality apps in an unregulated market, it is unlikely that future guidelines or health professionals will be able to make generic recommendations to patients regarding asthma mHealth app use and instead will need to make targeted recommendations about specific apps. Guidelines that incorporate reviews such as this review that identify apps known to be of high quality, such as Asthmahub, AioCare patient, and Kissmyasthma, will be an important future resource for general clinicians navigating the ever-expanding mHealth app market for chronic diseases.

Acknowledgments
Funding was received from the Northern Health Respiratory Department (Victoria, Australia) for the costs associated with submitting the protocol and manuscript and purchasing mobile health apps. No funding or sponsorship was received from the mobile app developers or third parties.

Authors' Contributions
BO and KS conceived this study. BR developed the study protocol by adapting a protocol previously created by EG for a diabetes mobile health app review project. KS and BR reviewed the Global Initiative for Asthma guidelines to establish key recommendations that could be feasibly incorporated into a mobile health app. BO and KS provided valuable guidance and advice for this research. BR, DZ, and EP reviewed the apps. BR was the primary author of this manuscript. EG and BR also provided training to all the reviewers of the project. All authors contributed to the refinement of the study protocol and approval of the final manuscript. The Respiratory Research Team at the Northern Hospital assisted in providing general advice to the authors.

Conflicts of Interest
The Northern Health Respiratory Department is providing funding for this study. The head of this department, KS, is a key author of this paper. The department does not have competing or fiduciary interests that could be affected by the results of this study. KS is not responsible for directly reviewing the apps included in the study or for the data analysis.

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

Multimedia Appendix 2
App technical information.

Multimedia Appendix 3
Data from initial app review process to identify apps that met inclusion but not exclusion criteria.

Multimedia Appendix 4
Stepwise guide on the key steps of data extraction and evaluation for reviewers to follow.

Multimedia Appendix 5
Data extraction forms (technical information, Mobile App Rating Scale, Intercontinental Medical Statistics functionality score, and asthma assessment checklist).

References


Abbreviations

BCT: behavioral change technique
GINA: Global Initiative for Asthma
IMS: Intercontinental Medical Statistics Institute for Health Informatics
MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 14.03.23; peer-reviewed by Y Zhang, R Almeida, JA Fonseca, PM Matricardi; comments to author 18.07.23; revised version received 21.10.23; accepted 21.11.23; published 10.01.24.

Please cite as:
Robinson B, Proimos E, Zou D, Gong E, Oldenburg B, See K
Functionality and Quality of Asthma mHealth Apps and Their Consistency With International Guidelines: Structured Search and Evaluation
JMIR Mhealth Uhealth 2024;12:e47295
URL: https://mhealth.jmir.org/2024/1/e47295
doi:10.2196/47295
PMID:38198204

©Billy Robinson, Eleni Proimos, Daniel Zou, Enying Gong, Brian Oldenburg, Katharine See. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
**Abstract**

**Background:** The field of eHealth is growing rapidly and chaotically. Health care professionals need guidance on reviewing and assessing health-related smartphone apps to propose appropriate ones to their patients. However, to date, no framework or evaluation tool fulfills this purpose.

**Objective:** Before developing a tool to help health care professionals assess and recommend apps to their patients, we aimed to create an overview of published criteria to describe and evaluate health apps.

**Methods:** We conducted a systematic review to identify existing criteria for eHealth smartphone app evaluation. Relevant databases and trial registers were queried for articles. Articles were included that (1) described tools, guidelines, dimensions, or criteria to evaluate apps, (2) were available in full text, and (3) were written in English, French, German, Italian, Portuguese, or Spanish. We proposed a conceptual framework for app evaluation based on the dimensions reported in the selected articles. This was revised iteratively in discussion rounds with international stakeholders. The conceptual framework was used to synthesize the reported evaluation criteria. The list of criteria was discussed and refined by the research team.

**Results:** Screening of 1258 articles yielded 128 (10.17%) that met the inclusion criteria. Of these 128 articles, 30 (23.4%) reported the use of self-developed criteria and described their development processes incompletely. Although 43 evaluation instruments were used only once, 6 were used in multiple studies. Most articles (83/128, 64.8%) did not report following theoretical guidelines; those that did noted 37 theoretical frameworks. On the basis of the selected articles, we proposed a conceptual framework to explore 6 app evaluation dimensions: context, stakeholder involvement, features and requirements, development processes, implementation, and evaluation. After standardizing the definitions, we identified 205 distinct criteria. Through consensus, the research team relabeled 12 of these and added 11 more—mainly related to ethical, legal, and social aspects—resulting in 216 evaluation criteria. No criteria had to be moved between dimensions.

**Conclusions:** This study provides a comprehensive overview of criteria currently used in clinical practice to describe and evaluate apps. This is necessary as no reviewed criteria sets were inclusive, and none included consistent definitions and terminology.
Although the resulting overview is impractical for use in clinical practice in its current form, it confirms the need to craft it into a purpose-built, theory-driven tool. Therefore, in a subsequent step, based on our current criteria set, we plan to construct an app evaluation tool with 2 parts: a short section (including 1-3 questions/dimension) to quickly disqualify clearly unsuitable apps and a longer one to investigate more likely candidates in closer detail. We will use a Delphi consensus-building process and develop a user manual to prepare for this undertaking.

**Trial Registration:** PROSPERO International Prospective Register of Systematic Reviews CRD42021227064; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021227064

*(JMIR Mhealth Uhealth 2024;12:e48625)* doi:10.2196/48625

**KEYWORDS**

telemedicine; smartphone; mobile apps; program evaluation; decision-making; systematic review; mobile phone

---

**Introduction**

**Background**

eHealth, that is, “the use of information and communication technology for health” [1], can support the delivery of interventions for self-management support and behavior change in patients with acute and chronic illnesses [2,3]. According to the World Health Organization (WHO) [4], self-care health interventions can be classified into individual agency (eg, promoting awareness about self-care), health information seeking (eg, education for informed health decision-making), social and community support (eg, peer mentorship and counseling), personal health tracking (eg, home-based records for health and diagnostic data), self-diagnosis of health conditions (eg, self-testing), self-management of health (eg, self-medication or treatment), individuals’ links to their health systems (eg, individuals sharing data with health care professionals [HCPs]), and individuals’ financial outlays for health (eg, expenses for prescription medicines). However, a recent evaluation of self-care interventions delivered via eHealth apps noted that only 20% of the 100 included apps used evidence-based information, whereas experienced HCPs considered only 32% to be useful and deemed 52% to be misleading and 11% to be dangerous [5].

Searching for a common characteristic linking the most effective apps, several systematic reviews and meta-analyses have found that those developed on firm theoretical foundations are more likely to be effective [6,7]. However, a systematic review of health-promoting smartphone apps found that only 55.6% of the included 27 studies described a theoretical basis for their smartphone app development [8]. A 2018 review found that only 8 (1.2%) of 681 smartphone apps to support medication adherence had documented evidence of their effectiveness. Such evidence is health care systems’ main consideration regarding certification and reimbursement [9]. Furthermore, although a user-centered design (sometimes also called human-centered design) and the involvement of patients and HCPs in the development of smartphone apps is known to provide insights into end users’ needs and helps ensure both relevant, reliable content and high quality [9,10], only 84 (12.3%) of the apps in this review had been developed in collaboration with HCPs. None reported patient involvement in their development processes [9].

Owing to the increasing availability of eHealth smartphone apps, it is vitally important and increasingly challenging for HCPs to identify, evaluate, and recommend relevant, trustworthy, and high-quality eHealth smartphone apps [11-14]. One tempting way for HCPs to form a first idea of an eHealth smartphone app’s quality is the star ratings and written reviews it receives on an app store [15]. However, this information is often subjective and distorted by individuals, comes from unverified or fraudulent sources, or provides no insights on an eHealth app’s quality [15-18]. HCPs also face a lack of reliable guidance on evaluating eHealth smartphone apps’ applicability to clinical practice [13,19,20]. Therefore, many are now struggling to describe and evaluate eHealth smartphone apps. A guideline regarding their characteristics and quality using standardized methods that will allow HCPs to propose reliable apps to their patients is needed [21,22].

Previous efforts to evaluate apps have generally focused on guidance for researchers [23-25]. Although the criteria were often overly complex or tailored to specific health areas, they also tended to be incomplete. Their underlying theories, scientific rationales, and development processes have rarely been described [20,25-27]. Furthermore, their unsuitable foci, nontransparent development processes, complexity, and often excessive time demands make them a poor fit for clinical practice. Finally, the existing instruments used a variety of criteria that only partially overlapped [20,23-27]. A clear description and evaluation of an app is important as, in rapidly evolving fields, even small changes or improvements to an app can have significant impacts on its use and usefulness [28]. To date, evaluation tools to help clinicians describe and evaluate eHealth apps, allowing them to recommend high-quality apps to their patients and share their thoughts using common terminology, are lacking [29,30].

**Objectives**

Therefore, the aim of this study was to obtain an overview of the evaluation criteria used in the literature. This process was conducted in three steps: (1) conducting a systematic review to identify existing criteria for evaluating eHealth apps, (2) developing a conceptual framework for the evaluation of eHealth smartphone apps, and (3) developing a comprehensive list of criteria for describing and evaluating eHealth smartphone apps. This was the foundational phase 1 of an overarching project to develop and pilot-test a theory-based tool to help HCPs evaluate the characteristics and quality of eHealth smartphone apps in a practical and standardized way (Figure 1). Phase 2 will involve
narrowing down, refining, and testing the evaluation criteria via 3 further steps: conducting a Delphi survey to narrow down the criteria (step 4), developing a user guide for the processes of description and evaluation (step 5), and pilot-testing the user guide and processes with HCPs (step 6). This paper focuses on reporting on the foundational phase 1 and outlining the proposed steps for phase 2.

Figure 1. Overview of the 2 phases and steps in the development of the eHealth smartphone app evaluation tool (the focus of this paper is framed on the left side).

Methods

Design

We used a 3-step descriptive, iterative, and developmental approach in phase 1. We first conducted a systematic review, then iteratively developed a new conceptual framework, and finished by compiling a comprehensive list of criteria for the evaluation of eHealth apps. The methodology for each of these steps is described in detail in the following sections. As this study did not deal with human participants or identifiable data, no ethics approval was needed.

Step 1: Systematic Review

To identify existing criteria for evaluating eHealth apps, we conducted a systematic review complying with the Cochrane Handbook for Systematic Reviews of Interventions [31]. The manuscript was written following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [32].

Protocol and Registration

Our review was registered in PROSPERO (CRD42021227064 [33]). No other protocol has been published.

Eligibility Criteria

Our systematic review included studies on any health condition fulfilling the following inclusion criteria: all had to be primary studies or reviews (1) explicitly describing tools, guidelines, dimensions, instruments, criteria or items, or development processes for tools to evaluate eHealth smartphone apps; (2) clearly describing the evaluation of interventions delivered via eHealth smartphone apps according to predefined (self-developed or existing) criteria; (3) having full-text versions available; and (4) being available in English, French, German, Italian, Portuguese, or Spanish. We incorporated criteria encompassing a broad spectrum of health care areas, including health promotion, prevention, and both physical and mental health. Recognizing the interconnected nature of physical and mental health and the diverse purposes and user groups for which eHealth is used, we integrated a wide array of objectives and stakeholders into our evaluation. Although these areas may exhibit distinct characteristics, it is conceivable that there are fundamental criteria that could be consistently applied in evaluating eHealth apps across different domains. These fundamental criteria may encompass aspects such as user-friendliness, data security, privacy, and usability, forming a shared foundation for evaluation to ensure that essential quality aspects are addressed. In summary, our systematic review encompasses these comprehensive topics to provide a thorough evaluation of eHealth quality criteria, which are applicable to diverse health care needs. Our aim was to encourage consistency and standardization in the evaluation process. Articles were excluded if they (1) described criteria to evaluate interventions delivered via eHealth websites and videos, among other media, but not smartphone apps; and (2) were study protocols, conference abstracts, editorials, or letters to the editor.
Information Sources
We queried the MEDLINE (OvidSP), CENTRAL (via Cochrane), CINAHL (EBSCOhost), and Web of Science databases. Supplementary searches were conducted on trial registries (ClinicalTrials.gov and WHO trial registry) and the reference lists of the included papers. The search was conducted on December 5, 2022. No time restrictions were imposed.

Search Strategy
We developed our MEDLINE search string based on the terms used in articles on (partially) similar topics [34-41] combined with key Medical Subject Heading and free-text terms (see the search strategy in Multimedia Appendix 1). For the other databases, we adapted the search string accordingly. We combined thematic blocks with various keywords related to eHealth, smartphone, application, evaluation, and tool. No filters were applied.

Selection Process
All identified titles and abstracts were independently screened for relevance by 2 reviewers (JR and TH). The full texts were assessed by the same reviewers using the criteria described previously. The reasons for full-text exclusion were reported. In one case of disagreement, an independent third researcher (SDG) contributed to help reach a consensus.

Data Collection Process and Data Items
In total, 2 reviewers (JR and TH) independently extracted the data and cross-checked their results. We extracted information on the author, year, country, research question or study aim, design, operating system, population or specific condition, main intended intervention purpose, name of the tool, and framework or theoretical guidance. The intended purpose of each eHealth app–delivered intervention was categorized according to the WHO classification for self-care health interventions [4]: individual agency, health information seeking, social and community support, personal health tracking, self-diagnosis of health conditions, self-management of health, individuals’ links to their health systems, and individuals’ financial outlays for health. Specific eHealth app quality evaluation dimensions or criteria were extracted and tabulated in a separate table.

Study Assessment
The included studies were assessed using the Appraisal of Guidelines for Research and Evaluation–II (AGREE-II) instrument [42], which is widely used to evaluate guideline development processes. As many of the included studies did not have a priori intend to develop an evaluation guideline, this instrument might not have been the best option for all studies. However, as we were mainly interested in the justification and development of the dimensions or criteria used in the studies, this instrument provided us with the best support for evaluating these aspects. The AGREE-II instrument consists of 23 items classified into 6 domains (3 items on scope and purpose; 3 on stakeholder involvement; 8 on rigor of development; 3 on clarity of presentation; 4 on applicability; and 2 on editorial independence, ie, whether funding body and competing interests were reported). We concluded our AGREE-II assessment by rating the degree to which each included study described each domain. For this, we used a 4-descriptor scale: accurately (all AGREE-II items fulfilled), partially (two-thirds of all AGREE-II items fulfilled), hardly (one-third of all AGREE-II items fulfilled), and not at all (0 AGREE-II items fulfilled).

Step 2: Development of a Conceptual Framework
The original dimensions of the frameworks reported in the selected articles were listed in a table. Similar descriptions of dimensions were merged. The first draft of the proposed conceptual framework and graphical representations was reviewed and discussed with various stakeholders (researchers, clinicians, designers, and software developers with backgrounds in nursing, medicine, ethics, and informatics). During these discussions, the participants recommended that we distinguish between technical dimensions (eg, design, usability, security, safety, and privacy) and those that focused on content (eg, evidence base and scientific evaluation). It was also recommended that the dimensions be presented as a linear, step-by-step process. During these meetings, the first author (JR) took notes and recorded the proposed changes until consensus was reached on the next version.

The second draft of the framework was discussed with 18 international volunteers (patient representatives, researchers, clinicians, and technology developers) from diverse backgrounds in health care (eg, psychology, nursing, and pharmacy) who were participating in a public webinar on quality evaluation of eHealth technology. As it was a public webinar, only limited data on participant demographics were collected (Table S1 in Multimedia Appendix 2). A survey via AhaSlides (AhaSlides Pte Ltd) to rate the importance and clarity of the dimensions and subgroups (1=not important; 5=very important) and open-ended questions to add missing dimensions or subgroups were used to engage with the participants. In addition, the participants were engaged to add comments orally that the first author (JR) put in writing. The quantitative data were analyzed descriptively (eg, frequency and mean), whereas the qualitative data were analyzed using the mind-mapping technique. The participants found the technology dimension too large (ie, covering too many subtopics) and partially unclear. Therefore, it was recommended to split this dimension into technological concerns (eg, technical requirements, security, safety, and privacy) and functional requirements (eg, the user-centeredness and usability of the design). In addition, they understood eHealth evaluation as a cyclical process and recommended presenting the conceptual framework as a continuous circuit as opposed to the initially linear process recommended previously.

Their feedback was used to draft a third version of the framework, which was presented and discussed with 34 researchers, clinicians, and technology developers (Table S2 in Multimedia Appendix 2; only 1 person overlaps with the volunteers from the webinar) mainly with a background in pharmacy who participated in the Next Chapter in Patient Care Conference in April 2022 in Pärnu, Estonia. This time, a survey via Mentimeter (Mentimeter AB) was used to rate the importance and clarity of the dimensions and subgroups (1=not important; 5=very important), and open-ended questions were used to add missing dimensions or subgroups. In addition, the participants were engaged to add comments orally while the
first author (JR) took notes. Quantitative data were analyzed descriptively, whereas qualitative data were analyzed using the mind-mapping technique. The Next Chapter in Patient Care participants recommended highlighting the overarching nature of the ethical, legal, and social aspects, which must be considered in all eHealth smartphone app evaluation dimensions. This resulted in a fourth version reflecting the general character of relevant ethical, legal, and social considerations.

Subsequent rounds of discussion and feedback with the research team focused on the scope and relationships between the dimensions. The participants highlighted the legal, ethical, and social aspects to be treated as part of the context. In addition, they agreed that stakeholder involvement should be seen as another overarching aspect that is important in all dimensions of eHealth app evaluation. This discussion resulted in the final version of our new conceptual framework for evaluating eHealth apps: the eHealth Smartphone App Evaluation (eHAPPI) framework.

Step 3: Development of a Comprehensive Criteria List

The eHAPPI framework was then used to synthesize all the eHealth smartphone app evaluation criteria identified in the selected studies. The redundant criteria were combined. The classification of the criteria according to the eHAPPI dimensions and suggestions for changes and regarding additional or irrelevant criteria were discussed and refined by the research team according to consensus.

Results

Step 1: Systematic Review

Study Selection

The results of our study selection process are presented in the PRISMA flow diagram (Figure 2). The search strategy described previously yielded 1021 nonduplicate titles. After screening of the titles and abstracts as well as full-text assessment for eligibility, our final analysis included 128 articles that met all the inclusion criteria (Multimedia Appendix 3 [43]).

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

Study Characteristics

General Description

A detailed description of the characteristics of the included studies can be found in Multimedia Appendix 4. The years of publication ranged from 2013 to 2022. All but 1.6% (2/128) of the articles were written in English—1 was in French [44] and the other in Spanish [45]. Four-fifths of the studies (101/128, 78.9%) were conducted in Western Europe, North America, or Australia. The studies used a variety of designs, mainly cross-sectional reviews of existing apps (42/128, 32.8%); reports on various app development and evaluation approaches (29/128, 22.7%); quantitative, qualitative cross-sectional, or longitudinal user testing of a single app (27/128, 21.1%); or different forms of reviews (15/128, 11.7%). The 128 included studies involved apps covering 30 topics, such as mental health, health promotion, or support for specific physical conditions (eg, heart disease and diabetes). However, only 50.8% (65/128) of the studies provided enough detail to categorize them according to
the WHO classification for self-care interventions for health and well-being [4].

Evaluation Tools
In total, the included studies used 142 distinct tools to evaluate eHealth apps. Although 76.6% (98/128) of the included studies used a single evaluation tool, 23.4% (30/128) used multiple tools and scales. Almost one-quarter of the studies (30/128, 23.4%) used an evaluation tool with investigator-developed criteria and then provided only scantily described development processes or theoretical backgrounds for those criteria. The most frequently used tool was the Mobile App Rating Scale (33/128, 25.8%) followed by its adapted versions (8/128, 6.3%) and the System Usability Scale (22/128, 17.2%). Less frequently used tools were the Post-Study System Usability Questionnaire (2/128, 1.6%), the Questionnaire for User Interaction Satisfaction (2/128, 1.6%), and the quality of experience survey (2/128, 1.6%). A total of 43 other tools were used in only 0.8% (1/128) of the studies each. In total, 10.2% (13/128) of the studies used qualitative methods (eg, interviews and focus groups) to generate the app evaluation criteria. In some cases (4/128, 3.1%), the origin of the criteria was unclear, or similar names were used for different tools.

Theoretical Frameworks
Most studies (83/128, 64.8%) did not clearly report a theoretical underpinning. The 32% (41/128) that did used 59 different frameworks, including various non–eHealth-specific behavioral, social, or implementation theories (10/59, 17%), the technology acceptance model (7/59, 12%), heuristic evaluation (5/59, 8%), models of the International Organization for Standardization (3/59, 5%), the (extended) Unified Theory of Acceptance and Use of Technology (3/59, 5%), or user-centered design (2/59, 3%). In total, 29 frameworks were used in only 2% (1/59) of the studies each. Of the 59 frameworks used in the included studies, 16 (27%) guided the development and 43 (73%) guided the evaluation of eHealth smartphone apps. In 1.6% (2/128) of the studies, different frameworks were used for development and evaluation.

Study Assessment
On average, the studies described 8.6 (SD 2.4; range 3-15) of the AGREE-II instrument’s 23 items [42]. Few studies described the scientific or theoretical basis and development processes of the app evaluation criteria that they applied. The most completely described or justified domains were related to scope and purpose (116/128, 90.6% described it accurately; 11/128, 8.6% described it partially; and 1/128, 0.8% hardly described it), editorial independence (91/128, 71.1% described it accurately; 30/128, 23.4% described it partially; and 7/128, 5.5% did not describe it at all), and stakeholder involvement (13/128, 10.2% described it accurately; 42/128, 32.8% described it partially; 70/128, 54.7% hardly described it; and 3/128, 2.3% did not describe it at all). The least fulfilled domains were applicability (1/128, 0.8% described it accurately; 2/128, 1.6% described it partially; 22/128, 17.2% hardly described it; and 103/128, 80.5% did not describe it at all), rigor of development (13/128, 10.2% described it partially; 53/128, 41.4% hardly described it; and 62/128, 48.4% did not describe it at all), and clarity of presentation (14/128, 10.9% described it accurately; 32/128, 25% described it partially; 52/128, 40.6% hardly described it; and 30/128, 23.4% did not describe it at all).

Step 2: Development of a Conceptual Framework
The full list of frameworks and original eHealth evaluation dimensions identified in the selected studies can be found in Multimedia Appendix 5 [43,45-47]. Some dimensions were included in only a few frameworks, and no framework included all possible dimensions.

The condensed dimensions were presented graphically and refined iteratively with the stakeholders until consensus was reached and no further adaptions were needed. The final eHAPPI conceptual framework (Figure 3) consists of six interrelated dimensions: (1) context, (2) stakeholder involvement, (3) development processes, (4) implementation, (5) evaluation, and (6) features and requirements.

A detailed definition of each dimension, including the subgroups, is presented in Textbox 1. Context describes a set of unique factors and conditions in which the app will be implemented [48]. This is an overarching dimension that depends on and, in turn, influences the other domains. Stakeholder involvement is essential in all aspects of eHealth. It involves the active engagement of relevant partners in all processes of the app life cycle, from conceptualization to sustainable implementation (eg, with end users, HCPs, researchers, and health insurers) [49,50].

Several subgroups were defined to further outline and structure the framework: basic information (concerning the app) and ethical, legal, and social aspects were seen as relevant subgroups of the context dimension. The featuresandrequirements dimension was assigned 4 subgroups: evidence-based content; functionality; usability, privacy, and security; and performance. Similarly, the development process dimension was divided into cocreation/user-centered design and characteristics of the development team, and adoption (integration into daily life) and maintenance were seen as relevant subgroups of the implementation dimension. Finally, the evaluation dimension included only 1 subgroup: scientific evaluation.
Figure 3. The final eHealth Smartphone App Evaluation framework.
Textbox 1. Overview of framework dimensions, subgroups, and definitions.

Context
- **Context** describes the set of unique factors and conditions in which the app is implemented. During the implementation process, the app, implementation, and context interact, influencing, modifying, facilitating, or hindering one another. As an overarching concept, context encompasses a physical location and roles, relationships, and interactions at various levels (ethical, legal, social, geographical, epidemiological, and political) [48,51].
- **Basic information** describes general information concerning the app (eg, name, URL, or available languages).
- **Ethical, legal, and social aspects** describe activities to help evaluators consider ethical, legislative, and social implications of an app’s development and implementation. These aspects cannot be completely separated from each other as they overlap somewhat [28].

Ethics describes activities to understand and explore the moral life, wherein morality includes beliefs, norms of behaviors, principles, and rules that guide individual and institutional behavior. Morality is a widely shared set of norms that result from a certain social consensus. Ethical aspects include the prevailing moral norms and values during the development and implementation of apps. Although moral values are influenced by cultural, sociopolitical, religious, and economic differences, many ethical reflections are shared by all countries and societies. Regarding apps, important ethical topics include benefit-harm balance, autonomy, respect for persons, and justice and equity and legislation (when using the app, ethical challenges may arise that were not considered in existing legislation and regulations) [28].

Legal aspects include rules and regulations that must be considered when developing and implementing apps. Rules and regulations are designed to protect the rights and interests of the patients and other members of society (eg, legislation on patients’ rights; data protection laws; or the regulations, rights, and obligations of health care professionals [HCPs] in general). Important legal topics are, for example, autonomy of the patient (eg, legal requirements), privacy of the patient (eg, no use of “unnecessary” patient information), equality in health care, ethical aspects (eg, impact on achievement of human rights), authorization and safety (eg, laws and rules regarding safety), ownership and liability, and regulation of the market [28].

Social aspects describe human-centered activities that approach end users and their social networks as reference points in an app’s development and implementation. These include groups of patients or individuals who might require special consideration (eg, vulnerable populations, people living in remote communities, people with learning difficulties, older people, ethnic minority individuals, and immigrants). Patients’, individuals’, and relatives’ perspectives should be considered when developing and implementing apps. Some social groups may be important for a particular app (eg, ethnic minority individuals and individuals with disabilities), which should be specified. Important social topics include patients’ perspectives (eg, expectations and wishes), social group aspects (accessibility), and communication aspects (eg, explanation of treatment choices) [28].

Stakeholder involvement
- **Stakeholder involvement** describes the active engagement of relevant partners in all processes of the smartphone app life cycle, from conceptualization to sustainable implementation (eg, end users, HCPs, researchers, and health insurers [49,50]).

Note: for end-user involvement in the development process, see the Cocreation/user-centered design section under the Development process subheading in this textbox.

Development process
- **Development process** describes all activities performed regarding the app’s creation, such as activities related to its aim, target population, guiding principles, applied approaches, stakeholder contributions, required changes for subgroups, or continuing uncertainties [52-54].

- Cocreation of technologies or user-centered design: cocreation describes a stepwise process in which partners (eg, patients, caregivers, and HCPs) are actively involved in the strategic design and planning of the app’s development and implementation at every stage of the process [55].

User-centered design is a multidisciplinary design approach using cyclic iteration and evaluation. As it actively involves users to improve the developers’ understanding of their requirements and wishes, this approach is seen as essential regarding product benefits and usability [56].

- **Characteristics of the development team** describes information about the multidisciplinary group responsible for the creation of the working, validated app. Important characteristics of the development team include the name of the app provider (developer and manufacturer of technology and content), contact details, organization attributes, and funding, as well as any conflicts of interest [57].

Evaluation
- **Evaluation** describes the assessment of the app’s efficacy, effectiveness, cost-effectiveness, safety, implementation, and impact [58].

- **Scientific evaluation** is the systematic assessment of the app’s efficacy, cost-effectiveness, safety, implementation, and impact through observation, measurement, and experimentation in a scientific study. Such an evaluation is essential to reliably measure an app’s effects and outcomes as a basis for decision-making [58-60].

Implementation
- **Implementation** describes the uptake and sustainable integration of evidence-based innovations such as apps into routine use [61].

- **Adoption** (ie, integration into daily life) deals with the app’s uptake (ie, activities focused on using the app in everyday life). This includes the user’s reaction to the app with respect to desired activities and interactions, such as downloads, clicks, and data entries. User engagement, that is, the user’s investment in learning about and participating in the app (eg, time and memory load), is an important prerequisite for the app’s long-term integration into daily life [62-64].
• Maintenance covers activities that ensure that an app is sustained in an acceptable and safe condition to perform its specified functions. Maintenance also includes repair as a stand-alone action to restore deteriorated or damaged parts. The activities needed to ensure the required performance of the app during its lifetime include updates, performance evaluations, and the planning and execution of necessary repairs [65].

Features and requirements

- Features and requirements describe the app and the characteristics or attributes it must provide to meet the users’ needs (ie, its information content, functionality, usability, security, privacy, and performance). The indications, performance features, and app options may differ for different generations or versions of an app. Important features and requirements topics include app description, claimed benefits, care level of use, reference values, or cutoff points used [28, 66]. A clear description of the app is important as, in rapidly evolving fields, even small changes or improvements to an app can have large impacts on its use and benefits. The description should enable those who are not familiar with the app to understand what it does, how it works, and how it can be used..

- Evidence-based content describes information, subject matter, and data content made available by the app based on the best available scientific knowledge and clinical expertise [67].

- Functionality describes the properties or functional requirements that affect the app’s use (ie, its features, components, and usefulness). In the context of self-monitoring, there are specific requirements related to design issues, particularly those concerning wearable technologies, as well as requirements regarding the behavioral aspects that are part of the design. These requirements need to be considered when evaluating a product or system for self-monitoring. The functionality information is intended to summarize the app’s overall suitability for use in a particular situation [38,88,89].

- Usability, privacy, and security are nonfunctional requirements (ie, they deal with how the system should operate in terms of usability, security, privacy, design, modularity, modifiability, reliability, availability, portability, and operability). Usability describes the extent to which an app can be used by its target users to achieve its intended goals effectively, efficiently, and satisfactorily. Ideally, an app should be easy to use, easy to learn, and easy to understand [70].

- Privacy describes the protection and control over personal data during the app’s processing operations [71].

- Security describes the provision of safeguards that serve the security of the individual or community. For example, this helps prevent or avoid poverty, hardship, theft, or espionage. Security is a major component of a stable, relatively predictable environment in which people can pursue their goals without interference, harm, or the fear of them. This includes protecting the app and associated data from events that could cause loss or severe damage, such as fire, burglary, theft, or vandalism [72,73].

- Performance describes whether an app works quickly and without errors and does not cause problems. Important performance topics include reliability and scalability (ie, whether an app still works properly when the number of users increases [74,75]).

Step 3: Development of a Comprehensive Criteria List

In total, 205 criteria for describing and evaluating eHealth smartphone apps were reported in the selected articles. None of the articles included all the possible criteria. The use of terminology differed within the publications or was attributed different meanings. For example, several were named usability but referred to different aspects of that criterion, such as ease of use, usefulness, or speed. Others, such as research-backed, scientific references, information accuracy, and information quality, were named differently but clearly referred to a single criterion—in this case, evidence-based information.

Using the new eHAPPI framework, it is clear that most studies reported criteria that focused on usability, evidence-based content, functionality, or scientific evaluation. However, only 1 reported criterion dealt with context, and only 5 dealt with development processes.

The research team’s discussions emphasized that some criteria are objective (eg, average rating in the app store and purpose of the app). In contrast, others are more subjective (eg, matching the needs of the target population and intention to use). In addition, several were dependent on the tested app’s purpose or content (eg, whether an app community exists and features to support behavior change).

The research team agreed on how all the criteria were classified into dimensions and how most were formulated. However, to improve understanding or align with known formulations, they suggested revisions to the wording of 12 criteria [28]. On the basis of the research team’s recommendations (Multimedia Appendix 6 [25,45-47,76-85]), 11 new criteria were added (Multimedia Appendix 7 [28]). These additions were mainly to the ethical, legal, and social aspects section. No criteria were deemed completely irrelevant, so none were removed. Finally, the research team agreed to classify these added criteria into the existing dimensions.

The final list contained 216 criteria for describing and evaluating eHealth smartphone apps (Multimedia Appendix 8). Although this list was comprehensive, our discussions highlighted that it was not practical for use by HCPs in clinical practice. The research team agreed that, in the next phase, a short version with only 1 to 3 items per dimension would be useful to make a quick initial decision (ie, acting as an algorithm to gauge whether an app should be given further consideration). Only if an app passed this pretest would it undergo a more thorough evaluation using detailed criteria and offering nuanced results. In addition, the research team recommended a specific algorithm with thresholds that could be adjusted depending on each tested app’s purpose and context.

The research team recommended that the next step for the project’s second phase be a Delphi process to condense the list and develop the proposed decision support tool. This process will have two aims: (1) to provide a means to reach a consensus and (2) to develop a useful and feasible (ie, practical for use in...
clinical practice) tool to describe and evaluate eHealth smartphone apps.

Moreover, participants expressed concerns that the necessary information to complete the tool may be difficult to find. Therefore, the development of a user guide for HCPs on how to apply the tool and where to typically find the required information was also proposed for the next phase. Finally, the eHealth smartphone app evaluation tool and user guide will need to be pilot-tested with HCPs.

**Discussion**

**Principal Findings**

Although the eHealth field is rapidly expanding and evolving, there is no consensus on how the quality of eHealth apps should be defined and evaluated by HCPs [13,19,20,86]. In this paper, we described how we developed a comprehensive list of criteria to evaluate eHealth apps. We used a meticulous methodological approach to derive the list, consisting of a systematic literature review and iterative rounds with stakeholders from various backgrounds to compose a comprehensive framework—the eHAPPI—and use it to synthesize all the criteria identified in the selected studies. We found 6 overarching dimensions (ie, context, stakeholder involvement, features and requirements, development processes, implementation, and evaluation) of eHealth app evaluation and 205 criteria in the literature. A research team discussion resulted in 11 additional criteria, bringing the total to 216. Using this comprehensive list, HCPs would be able to evaluate eHealth apps designed for diverse health care needs. No original studies included all the dimensions or all the criteria. Most of the selected studies (83/128, 64.8%) did not describe an underlying framework or theoretical guidance on how their criteria were developed. In addition, a general lack of common terminology among the included studies further complicated efforts toward comparison and direct application in clinical practice.

Although the proposed complete list of criteria was comprehensive, applicable to diverse health care needs, and applied consistent terminology, its length and depth limited the feasibility of using it in clinical practice. As a stand-alone reference, this list would be better suited for comprehensive evaluations of a broad range of eHealth apps by a regulatory body. Therefore, we proposed a 2-step approach to developing these criteria for use by HCPs. The first step would be defining a few critical initial decision criteria. A more detailed description and assessment would be useful only for positive evaluations using these first criteria. Using such an algorithm would allow the most obviously inferior apps to be quickly excluded from the process.

However, our progress in this direction was slowed by a lack of consensus regarding which criteria were essential and which would simply be nice to have [23]. As a compromise, we agreed that developing the proposed algorithm would require more expertise than we had and further expert discussion. Therefore, for the second phase of this project, we proposed a Delphi process [87] to guide the further development and fine-tuning of the final evaluation tool.

Although we did not reach a consensus on which criteria qualify as essential, our group discussions provided insights into which qualities to consider in describing and evaluating apps as well as how an optimal tool might be structured. For example, given the dynamic and rapidly evolving use of apps in clinical practice [29,88], flexibility is a significant concern [23,25]. Therefore, the research team recommended an algorithm not only whose cutoff criteria can be modified to fit each evaluated app’s purpose and context but also whose overall functionalities can evolve alongside the surrounding technology [89].

Another point of discussion concerned the difficulty we encountered in finding the information to complete this evaluation tool. HCPs who are less familiar with eHealth apps may have difficulty gathering even basic data, such as the name of an app’s developer or its latest update [13,21,22]. More technologically savvy individuals may have trouble finding information on that app’s scientific basis, how or whether its development processes included stakeholder involvement, or which strategies were used in its implementation. Therefore, we suggest that the proposed tool include a user guide describing why such criteria are important and where to find and how to judge the required information. This echoes a recommendation by the European Network to Advance Best Practices and Technology on Medication Adherence in their Cooperation in Science and Technology Action (CA19132), which facilitates the use of a web-based repository of information on medication adherence technology [90,91]. Although eHealth app developers clearly need to provide relevant details in a clear and easily accessible way, health educators also need to include eHealth evaluation in HCP education and training curricula.

The previous discussion provides the foundation for conducting phase 2 of this study. This phase has three goals: (1) to conduct a Delphi survey to narrow down the number of criteria and develop an algorithm for initial decision-making, (2) to develop a user guide, and (3) to pilot-test the resulting iteration.

**Limitations**

This study has several limitations. Most notably, at this point in the project, although our list of eHealth app description and evaluation criteria is comprehensive, it remains a preliminary version. That is, despite discussions with various interdisciplinary experts, phase 1 did not produce an in-depth consensus on the essential criteria for a richly detailed but broadly feasible means of evaluating eHealth apps. This drove the decision to design a 2-phase project. In phase 2, which will be a Delphi study [87], we aim to develop a criteria-sorting algorithm. With this in place, the phase will culminate in a version of an eHealth app evaluation tool for pilot-testing.

In addition, all the included studies were assessed using the AGREE-II instrument [42], which was specifically designed to assess clinical practice guideline development reports. Considering the high level of heterogeneity across many of the study characteristics, direct comparability using a single tool was limited. However, although other instruments would have been more suitable in many cases, using various instruments would have yielded equally varied results. As we were primarily interested in the rationales and development processes that
supported the dimensions and criteria, the AGREE-II scales provided a consistent assessment of these aspects.

Finally, the participants in the discussion rounds for the development of the conceptual framework were primarily health care researchers and professionals. There were very few technology developers or industry representatives present, and only 2 patient representatives participated. This may have resulted in a limited consideration of the patient perspective and an increased risk of interventions exacerbating existing health care inequalities [92]. Therefore, we included in the context dimension of the eHAPPI framework subgroups focusing on ethical and social aspects. These subgroups aim to underscore the necessity of addressing the risk of intervention-generated inequalities.

Comparison With Prior Work

Although 23 of the existing tools were explicitly intended for HCPs [14,43,46,47,76-86,93-100], none of these were complete; rather, they were too focused on specific conditions, or their theoretical justifications or development processes were not described. Such omissions make it difficult for HCPs in clinical practice to comprehensively but feasibly describe and evaluate eHealth apps in a standardized way to guide the recommendation of relevant, reliable, and high-quality apps to their patients [11,13,14]. Other studies supplemented the dimensions and criteria for describing and evaluating apps. However, it remains unclear which criteria are essential and how detailed they need to be. Recently, there has been much discussion about how to define and evaluate eHealth quality and what criteria are needed for an app to be used in the health care system [23,89,101]. Future findings from the planned phase 2 will likely provide a basis for further discussion on this topic among app developers or providers, HCPs, patients, researchers, and policy makers.

Our first comprehensive list of criteria as a result of phase 1 provides an excellent basis for the next steps in phase 2 to develop a new eHealth app evaluation tool.

The need for a tool to describe and evaluate eHealth apps and help HCPs and their patients navigate the digital health ecosystem is pressing [23]. Our path to a proposed resolution has been quite complex as this field is also complex. After listing the criteria identified via a literature review, we developed them through expert discussions, revealing important improvement areas. In particular, compared with recommendations from the Health Technology Assessment Core Model [28], the criteria concerning the ethical, legal, and social aspects of eHealth apps were deemed incomplete. Therefore, in addition to adapting many criteria, we added several.

Contribution of This Study

This study focuses on addressing the rapidly growing and somewhat chaotic field of eHealth, particularly the challenges faced by HCPs when it comes to evaluating and recommending health-related smartphone apps to their patients. This study’s contribution lies in its comprehensive methodology for gathering and categorizing existing criteria for evaluating health apps, which is essential as no single framework or evaluation tool effectively serves this purpose. The methodology involved a systematic review of the literature, which resulted in the identification of 216 distinct evaluation criteria organized within a conceptual framework comprising 6 app evaluation dimensions. These dimensions encompass various aspects, including the app’s context, stakeholder involvement in its development, features, development processes, implementation, and evaluation. This study highlights the need for a more purpose-built, theory-driven tool to help HCPs assess and recommend apps effectively and outlines plans to create a 2-part app evaluation tool based on the gathered criteria, which will expedite the process of disqualifying unsuitable apps and scrutinizing potential candidates more closely. This study serves as a crucial foundational step toward developing a practical tool that can guide HCPs in evaluating and recommending health-related apps.

Conclusions

Developing a tool comprehensive enough for HCPs to reliably describe and evaluate the full range of eHealth apps yet short enough to be feasible for daily clinical practice is a daunting challenge. After our literature review yielded a list of criteria too bulky for routine use, there was a lack of consensus either on terminology or on relevance to define and evaluate app quality. In this report of phase 1, we provided our initial comprehensive overview of 216 relevant criteria used in the selected studies to describe, evaluate, and recommend eHealth apps. To condense this list to a more manageable size, in phase 2, we will formulate and apply a robust consensus-building process to generate a list of criteria ranked by importance, followed by the creation of an algorithm to produce short- and long-form evaluations to match the characteristics of the apps to be evaluated. In addition, the development of a user guide and pilot-testing of the tool are planned. As a basis for informed guidance and decision-making, such a tool will help HCPs reliably describe and evaluate eHealth apps for their patients.

Acknowledgments

This work was conducted mainly by a PhD student as part of her PhD project funded by the University of Basel. The funders had no role in the study design; collection, analysis, and interpretation of the data; writing of the report; or decision to submit the paper for publication. The authors thank all the participants in the various discussion rounds for their reflections and input for the development of the conceptual framework and criteria. Special thanks also go to the participants of the Research Round Table Meeting seminar for their critical feedback in developing the study methods and manuscript, as well as to Chris Shultis for his language editing.

https://mhealth.jmir.org/2024/11/e48625

JmIR Mhealth Uhealth 2024 | vol. 12 | e48625 | p.357
(page number not for citation purposes)
Authors' Contributions

The study was conceptualized by JR, SDG, and ADD. The literature search strategy was developed by JR, SDG, and FD. Articles were located, identified, and evaluated by JR and TH. Data were extracted and checked by these same authors. The initial dimension and criteria list was composed and discussed by all the authors. The manuscript was drafted and edited by JR with close revision and feedback from SDG as well as review and feedback from all other coauthors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE search strategy.
[PDF File (Adobe PDF File), 153 KB - mhealth_v12i1e48625_app1.pdf ]

Multimedia Appendix 2
Demographics of participants.
[PDF File (Adobe PDF File), 184 KB - mhealth_v12i1e48625_app2.pdf ]

Multimedia Appendix 3
Included articles.
[PDF File (Adobe PDF File), 470 KB - mhealth_v12i1e48625_app3.pdf ]

Multimedia Appendix 4
Characteristics of the included studies.
[PDF File (Adobe PDF File), 289 KB - mhealth_v12i1e48625_app4.pdf ]

Multimedia Appendix 5
Original frameworks and dimensions.
[PDF File (Adobe PDF File), 282 KB - mhealth_v12i1e48625_app5.pdf ]

Multimedia Appendix 6
Revised criteria.
[DOCX File, 129 KB - mhealth_v12i1e48625_app6.docx ]

Multimedia Appendix 7
Additional criteria.
[DOCX File, 45 KB - mhealth_v12i1e48625_app7.docx ]

Multimedia Appendix 8
Complete criteria list.
[PDF File (Adobe PDF File), 376 KB - mhealth_v12i1e48625_app8.pdf ]

References


Abbreviations

AGREE-II: Appraisal of Guidelines for Research and Evaluation–II

eHAPPI: eHealth Smartphone App Evaluation

HCP: health care professional

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

WHO: World Health Organization

©Janette Ribaut, Annette DeVito Dabbs, Fabienne Dobbels, Alexandra Teynor, Elisabeth Veronica Mess, Theresa Hoffmann, Sabina De Geest. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Original Paper

Evaluation of Chinese HIV Mobile Apps by Researchers and Patients With HIV: Quality Evaluation Study

Peng Liu¹,², MS; Lingmeng Wang¹, BS; Fuzhi Wang¹,², PhD

¹School of Health Management, Bengbu Medical College, Bengbu, China
²Innovation Team of Health Information Management and Application Research, Bengbu Medical College, Bengbu, China

Corresponding Author:
Fuzhi Wang, PhD
School of Health Management
Bengbu Medical College
2600 Donghai Road
Bengbu, 233030
China
Phone: 86 0552 3173195
Email: wfz.bbmc@foxmail.com

Abstract

Background: Against the backdrop of globalization, China remains one of the most heavily burdened countries in Asia with regard to AIDS. However, many high-risk groups and patients affected by AIDS may be less likely to actively seek care from medical institutions because of fear of experiencing shame or discrimination. Mobile apps provide a promising avenue for supporting the prevention, diagnosis, and treatment of AIDS. However, a comprehensive systematic evaluation of these mobile apps’ functionality and quality has not been conducted yet.

Objective: This study aims to identify the available mobile apps for AIDS in China, assess and discuss the functional features and quality of these Chinese AIDS mobile apps, and offer decision support for patients and clinical practitioners in accessing high-quality AIDS mobile apps. Furthermore, based on the evaluation results, recommendations for improvement will be provided.

Methods: A systematic search was conducted on the Qimai app data platform, the Aladdin WeChat applet data platform, and WeChat to identify mobile apps related to AIDS. A snowball sampling method was used to supplement the potentially overlooked apps. The selected mobile apps underwent a rigorous screening process based on unified criteria. Subsequently, assessments were independently undertaken by 3 separate researchers and 2 patients with HIV, using both the Mobile App Rating Scale (MARS) and the User Mobile App Rating Scale (uMARS). Quantitative interpretations of the data were facilitated by the MedCalc statistical software (version 20.217, MedCalc Software).

Results: A total of 2901 AIDS mobile apps were included in the study, with 2897 identified through information retrieval and an additional 4 added via snowball sampling. After a rigorous selection process, 21 apps were determined to be usable. Among them, the Hong Feng Wan app achieved the highest combined average score, calculated based on the MARS (3.96, SD 0.33) and uMARS (4.47, SD 0.26). Overall, there was no significant correlation between MARS and uMARS ($r_{\text{app quality total score}}=0.41; P=0.07; r_{\text{subjective quality}}=0.39; P=0.08$). A notable issue was the widespread lack of user privacy protection, with only 24% (5/21) of the apps offering this feature.

Conclusions: The number of available Chinese AIDS mobile apps is limited, with WeChat applets dominating the market. Nonetheless, the performance of WeChat mini-apps is generally inferior to that of independent apps, and there may be significant discrepancies between assessments conducted by researchers and those provided by genuine end users, emphasizing the necessity of involving real users in the development and evaluation of HIV mobile apps. In addition, developers of these Chinese HIV mobile apps need to devote attention to improving privacy protection mechanisms, in addition to considering the evaluations of researchers and real users. This will help attract more users and increase user loyalty.

(JMIR Mhealth Uhealth 2024;12:e52573) doi:10.2196/52573

KEYWORDS
HIV; mobile app; evaluation; mobile phone
Introduction

Background

AIDS is a chronic infectious disease caused by HIV infection, resulting in high mortality rates. Since the first reported case of AIDS, it has rapidly spread worldwide, becoming a major public health concern. China, influenced by globalization, is no exception. According to the 2021 National Notifiable Infectious Diseases Summary released by the Chinese Center for Disease Control and Prevention, AIDS has the highest mortality rate and death toll, with 19,623 reported deaths. Available data indicate that as of 2022, the number of confirmed AIDS cases in China has exceeded 1.05 million and is projected to surpass 1.6 million infections by 2023. AIDS remains a significant infectious disease affecting public health and socioeconomic development in China [1].

For individuals who are at a high risk of HIV infection but have not yet been infected, pre-exposure prophylaxis (PrEP) is a highly effective method of preventing HIV through daily medication. However, the current use of and adherence to PrEP remain low. However, mobile apps have shown effectiveness in promoting PrEP use and adherence among high-risk populations [2]. Muessig et al [3] highlighted the advantages of internet and mobile-based interventions, which facilitate wider dissemination of PrEP at a lower cost compared with conventional methods. The MyChoices app developed by Biello KB demonstrated feasibility and potential in improving HIV PrEP use rates among gay and bisexual men in the United States [4]. The prevalence of HIV infections remains disproportionately high among gay, bisexual, and other men who have sex with men (GBMSM) [5]. With the widespread use and convenience of smartphones, sexual networking apps have replaced traditional dating websites as the primary online social platform for GBMSM. As such, apps have become crucial venues for sexual health research [6]. Targeted video and text-based sexually transmitted infection and HIV information provided through mobile apps has proven effective in reducing new infection rates among GBMSM [5].

For people living with HIV, the highly active antiretroviral therapy developed by scientist David Ho represents the most effective treatment method. Highly active antiretroviral therapy can suppress HIV replication and prevent the emergence of drug-resistant viruses. However, strict adherence is required, necessitating patients to adhere to prescribed medication schedules [7]. Despite simplified treatment regimens, adherence remains challenging for some people living with HIV [8]. Evidence suggests that HIV case management can improve treatment adherence and quality of life and reduce risky sexual behavior [9,10]. However, case managers face significant workloads and limited policy and funding support [11,12].

Mobile health (mHealth) apps have demonstrated potential in assisting people living with HIV with effective self-management and delivering personalized interventions. Schnall et al [13,14] identified ideal features for an HIV app, including reminders, health information delivery, medication logs, communications, settings, and search functions. However, comprehensive apps meeting these criteria are scarce in the current market, and there is a lack of rigorously evaluated mHealth apps specifically designed for people living with HIV [13,14]. Yang et al [15] found that most Chinese GBMSM apps, selected based on relaxed inclusion criteria, primarily focused on dating and lacked HIV prevention and health information.

For both high-risk populations and individuals infected with HIV, HIV testing plays a vital role in combating HIV. Early HIV testing allows individuals to learn about their infection status promptly, facilitating timely access to antiretroviral therapy and significantly reducing mortality rates. However, fear of shame or discrimination may discourage many high-risk populations and patients from actively seeking care at health care facilities [16-19]. With the increasing prevalence of mobile phones, various mHealth interventions have been developed to diversify HIV self-testing (HIVST) approaches, including telephone hotlines, SMS text messaging–based interventions, and internet-based interventions. These interventions have shown potential in improving testing rates, particularly among hard-to-reach populations [20-26]. Although these platforms have achieved varying levels of success, the use of mobile apps has emerged as a highly popular trend because of their flexibility and scalability. For instance, the mLab app serves to enhance users’ understanding of their HIV test results while facilitating their access to pertinent HIV information and services [27]. Another noteworthy app is Aspect HIVST, which offers an acceptable means of uploading mobile HIVST results and demographic information to a centralized database [28]. In addition, ApiDé serves as a multilingual electronic tool (app) that assists health care providers in offering and explaining HIV screening to immigrants facing language barriers [29].

In conclusion, the use of mobile apps is becoming increasingly prevalent in HIV/AIDS prevention and control. These apps provide a convenient and accessible means for high-risk individuals and patients to access information, consultations, and support related to HIV/AIDS. They serve as effective adjunct measures in improving antiretroviral therapy adherence and reducing AIDS incidence rates, thereby mitigating the currently imbalanced resource allocation between patients and health care providers involved in antiviral treatment. However, research on Chinese AIDS mobile medical apps is relatively scarce and late, and empirical studies evaluating the effectiveness of HIV mobile medical apps remain sparse. Therefore, the effectiveness of Chinese HIV mobile apps in meeting user needs requires further assessment. To evaluate the efficacy of AIDS mobile apps in China, the Mobile App Rating Scale (MARS) and the User Mobile App Rating Scale (uMARS) scales, which are widely accepted and applied for uses such as assessing chronic disease management [30], COVID-19 tracking [31], psychiatric interventions [32], physical exercise among older people [33], and menstrual monitoring [34], were incorporated into this study as standardized assessment tools for the quality of mobile apps.

Objectives

This study aims to achieve the following objectives:

1. To identify a comprehensive list of available Chinese mobile apps for HIV/AIDS.

https://mh health.jmir.org/2024/1/e52573
JMI R MHEALTH AND UHEALTH
Liu et al
To evaluate the functional features and quality of Chinese HIV/AIDS mobile apps from 2 distinct standpoints: those of researchers and those of individuals diagnosed with HIV, and to provide decision support for AIDS-related groups and health care professionals in accessing high-quality AIDS mobile apps.

3. To conduct a thorough analysis of the evaluation results and to provide improvement recommendations based on the findings, with the ultimate goal of enhancing the quality of Chinese HIV/AIDS mobile apps.

Methods

Search Strategy
Considering the use status of Chinese AIDS apps, a systematic search was conducted on the Qimai app data platform, Aladdin WeChat applet data platform, and WeChat from February 18 to 19, 2023, adhering to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The specific search strategies are as follows.

Data Collection of Apps
The Qimai platform, a well-known domestic mobile app data analytics platform, was used to retrieve AIDS-related apps. Qimai provides comprehensive data on iOS and Android app markets (including Huawei, Baidu, Xiaomi, Vivo, etc), along with app store optimization and app store search marketing optimization service tools and SearchAds data reports. It also offers professional data analysis and optimization strategies. The platform features an intelligent keyword expansion and association tool, which proved valuable for obtaining the required data for this study. In this research, the Qimai Keyword Expansion Assistant was used to search for AIDS-related terms. We selected keywords with a relevance score exceeding 50% and a search index surpassing 4605 (typically indicating a higher search frequency) to identify relevant apps that met the criteria. The search was conducted on February 18, 2023.

WeChat Applet Data Collection
WeChat applets, launched in China in 2017, are app programs developed on the WeChat platform. These applets offer users the convenience of accessing and using various functionalities and services directly within the WeChat app without requiring any separate downloads or installations. Known for their lightweight, fast, and user-friendly nature, they have gained popularity. To collect WeChat applet data, searches were performed on both the Aladdin Index and WeChat. The Aladdin Index, developed by Beijing Aladdin Future Technology Co, Ltd, serves as a ranking platform for the WeChat applet, providing a comprehensive reference for applet developers’ operations nationwide. For this study, keywords such as HIV, AIDS, Human Immunodeficiency Virus, Acquired Immunodeficiency Syndrome, as well as their corresponding Chinese keywords with similar meanings, such as huo-de-xing-mian-yi-que-xian-zong-he-zheng, ai-zi, ai-zi-bing, and ai-si-bing, were used to search for the WeChat applet related to AIDS in both the Aladdin applet list and WeChat. The search was conducted on February 19, 2023.

Supplementary Method
In collaboration with the HIV prevention and control team at the local Centers for Disease Control and Prevention (CDC), a snowball sampling method was used to distribute electronic questionnaires to individuals living with AIDS during their follow-up visits conducted by the CDC staff. The questionnaires (without personal information and providing adequate privacy protection) aimed to inquire about the HIV/AIDS apps they had used. Participants were encouraged to provide feedback using the Wenjuanxing platform and to share the electronic questionnaire with other individuals living with AIDS they knew. This supplementary method aimed to identify AIDS-related apps that were widely used within the AIDS community but were not captured by the aforementioned search strategies. The feedback collection period was extended from February 20 to May 19.

The exclusion criteria for apps were as follows: (1) nonsimplified Chinese language, (2) games, (3) apps with excessively limited functionality, (4) mobile apps that did not primarily focus on HIV or AIDS-related content, (5) >50% of the content is inaccessible, (6) apps that had terminated their services, (7) apps restricted to internal use by specific personnel (apps had to be accessible and functional for evaluation purposes), and (8) apps predominantly intended for advertising and product sales.

Evaluation Tool
In this study, we selected the MARS and uMARS scales as the evaluation tools for assessing the quality of the apps. These scales were chosen for 2 main reasons. First, these scales have been extensively validated and have demonstrated good reliability and validity across different contexts. Second, in our study, the included apps were a mix of independent apps and WeChat applets, with the majority being WeChat applets rather than independent apps. However, the MARS and uMARS scales offer broader applicability in this regard.

Developed by Zelmer et al [35], the MARS scale provides a standardized set of criteria for evaluating mobile apps in terms of engagement, functionality, esthetics, and information quality. The MARS scale has been widely used in user research, mobile app evaluations, and related fields. Previous studies have indicated that the MARS scale demonstrates good validity and reliability, making it a reliable measurement tool [30-34]. The uMARS scale, developed by Manning et al [36], serves as an extension of the original MARS scale, specifically designed to prioritize the assessments of average users toward mobile apps.

Evaluation Process
The evaluation was conducted by 3 researchers. Rater 1 (PL) holds a master’s degree in medical informatics and has extensive experience in medical information analysis. Rater 2 (Bin Li) holds a master’s degree in computer science and has abundant software development experience. Rater 3 (FW), an esteemed scholar in medical informatics in China, possesses a master’s degree in computer science and a PhD in social medicine. Being an expert in app quality evaluation, rater 3 boasts rich experience and high authority in research on assessing mobile app quality. Both rater 1 and rater 2 were recruited as real patients with HIV.
The evaluations by the researchers and patients with HIV were conducted separately to avoid any interference. Each group followed a 3-step evaluation process. First, basic information for 21 apps, including their names, developers, platforms, and core functionalities, was collected from the SevenMa platform, Aladdin WeChat platform, and WeChat. Second, 2 WeChat applets and 2 independent apps were randomly selected for pilot evaluation. Before conducting the pilot evaluation, all 3 raters watched training videos on the MARS scale to better understand the purpose and significance of each item on the scale. Third, the 21 collected apps underwent a formal evaluation. To ensure consistent evaluation results, rater 1 and rater 2 independently assessed the same samples, completing the evaluation of the nth subdimension for all samples before moving on to evaluate the n+1th dimension. In cases where rater 1 and rater 2 had conflicting evaluations, we applied the Delphi method, which involved multiple rounds of consultation and feedback with experts, to achieve agreement regarding the discrepancies. Rater 3, an expert with relevant qualifications, reassessed the dimension that exhibited disagreement. The evaluation results from rater 3 were then shared with rater 1 and rater 2, along with appropriate explanations. This gave rater 1 and rater 2 the opportunity to revise their ratings and evaluations. If substantial differences persisted even after re-evaluation, the entire process was repeated until their evaluations converged or reached a satisfactory level of agreement.

Ethical Considerations
This study focused solely on mobile apps and did not involve the collection of any personal information or data. Therefore, there were no ethical concerns or disputes associated with this research.

Results
Overview
A total of 21 apps that met the criteria were obtained through screening, with 18 being WeChat applets, dominating in terms of quantity, whereas only 3 were independent apps (Figure 1). Representative AIDS mobile apps were also displayed, as detailed in Figure 2. The characteristics of these apps are shown in Tables 1 and 2.

The developers of these apps can be categorized into 3 groups: companies (6/21, 29%), government organizations (7/21, 33%), and nonprofit social organizations (8/21, 38%). In addition, of the 21 included apps, 18 (86%) provided HIV/AIDS knowledge dissemination, 12 (57%) offered appointment booking for testing, 11 (52%) had counseling services, 8 (38%) had features related to PrEP and postexposure prophylaxis (PEP), 2 (10%) had community networking functionalities, and only 1 (5%) had live streaming capabilities. Of the 21 apps included in the study, 16 (76%) were able to update their content within a year, whereas 5 (24%) had not been updated for >1 year. Only 6 apps had privacy protection features, all of which provided privacy policies or agreements, with 2 of the apps offering privacy password functionality.

Statistical Analysis
All analyses were performed using MedCalc software (version 20.217, MedCalc Software). Descriptive scores were derived from the MARS and uMARS scales. To evaluate the reliability of the raters’ assessments, Bland-Altman analysis was used to assess both interrater agreement and the range of variability in their scores. Visual representations were also used to visualize the differences in ratings between the 2 evaluators, providing a more concrete quantification of the discrepancies.
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection process for inclusion of the apps.

Figure 2. Example of a typical Chinese AIDS mobile app.
Table 1. General characteristics of the AIDS mobile apps (N=21).

<table>
<thead>
<tr>
<th>Features</th>
<th>Total apps, n (%)</th>
<th>Independent apps (n=3), n (%)</th>
<th>WeChat applets (n=18), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Affiliation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Company</td>
<td>6 (29)</td>
<td>3 (14)</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Government organization</td>
<td>7 (33)</td>
<td>0 (0)</td>
<td>7 (33)</td>
</tr>
<tr>
<td>Nonprofit social organization</td>
<td>8 (38)</td>
<td>0 (0)</td>
<td>8 (38)</td>
</tr>
<tr>
<td><strong>Content type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live streaming</td>
<td>1 (5)</td>
<td>0 (0)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Appointment for testing</td>
<td>15 (71)</td>
<td>3 (14)</td>
<td>12 (57)</td>
</tr>
<tr>
<td>Consultation</td>
<td>10 (48)</td>
<td>2 (10)</td>
<td>8 (38)</td>
</tr>
<tr>
<td>AIDS knowledge</td>
<td>18 (86)</td>
<td>2 (10)</td>
<td>16 (76)</td>
</tr>
<tr>
<td>PrEP\textsuperscript{a} or PEP\textsuperscript{b}</td>
<td>8 (38)</td>
<td>3 (14)</td>
<td>5 (24)</td>
</tr>
<tr>
<td>Web community</td>
<td>2 (10)</td>
<td>1 (5)</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Last updated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within 1 month</td>
<td>5 (24)</td>
<td>1 (5)</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Within 2 to 6 months</td>
<td>8 (38)</td>
<td>2 (10)</td>
<td>6 (29)</td>
</tr>
<tr>
<td>Within 7 to 12 months</td>
<td>3 (14)</td>
<td>0 (0)</td>
<td>3 (14)</td>
</tr>
<tr>
<td>&gt;1 year</td>
<td>5 (24)</td>
<td>0 (0)</td>
<td>5 (24)</td>
</tr>
<tr>
<td><strong>Privacy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Privacy policy or agreement</td>
<td>6 (29)</td>
<td>3 (14)</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Privacy password</td>
<td>2 (10)</td>
<td>2 (10)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PrEP: pre-exposure prophylaxis.

\textsuperscript{b}PEP: postexposure prophylaxis.
Table 2. Basic information about the AIDS mobile apps.

<table>
<thead>
<tr>
<th>App name</th>
<th>App OR WC</th>
<th>Knowledge sharing</th>
<th>Appointment testing</th>
<th>Live streaming</th>
<th>Consultation</th>
<th>Risk assessments</th>
<th>PrEPb and PEPc</th>
<th>Web-based community</th>
<th>Privacy protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chabei</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Danlan Happy Test</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Suzhou Red Ribbon</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Baiyin HIV test</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Linqu County CDC</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Wu Ai Fang Hua</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>E Ai Jian</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Ai Yi Jian</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Rong Ai Jian</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Liaocheng Dongchangen District Anti-AIDS Service Platform</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Zhecheng County AIDS consulting and test</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Qingai Health Services</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Ai Zhiku</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Ai Cheng Wang Shi</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Beijing AIDS Association</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Douai Check</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Nanyue Gaozhibao</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Red Ribbon Volunteer House</td>
<td>WC</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Hong Feng Wan</td>
<td>App</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Life4me+</td>
<td>App</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
<tr>
<td>Xiao Ai</td>
<td>App</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
</tr>
</tbody>
</table>

aWC: WeChat applet.
cPEP: postexposure prophylaxis.
dCDC: Centers for Disease Control and Prevention.

Functionality of the App
Most mobile apps (18/21, 86%) provided HIV/AIDS knowledge dissemination, and 71% (15/21) of the apps offered appointment booking for testing. Nearly half of the apps (10/21, 48%) provided counseling services, whereas 38% (8/21) of the apps offered features related to PrEP or PEP. Only 10% (2/21) of the apps had functionalities for live streaming and web-based community.

In addition to these functions related to function evaluation, we also included additional statistics on privacy protection settings. Among the 21 apps included in the study, only 6 (29%) provided privacy protection settings and 2 (10%) designed privacy passwords.
Quality of the App

Overview of App Composite Scores

The composite scores for the app quality total score in MARS and uMARS were obtained by averaging the scores for each app. The overall composite score for the 21 included apps was 3.43 (mean app quality total score of MARS=3.47, SD 0.37; mean app quality total score of uMARS=3.38, SD 0.53). Hong Feng Wan achieved the highest composite score, with a score of 4.22 (mean app quality total score of MARS=3.96; mean app quality total score of uMARS=4.47), followed by Wu Ai Fang Hua with a composite score of 3.9 (mean app quality total score of MARS=4.06; mean app quality total score of uMARS=3.74), and Chabei secured the third rank with a composite score of 3.76 (mean app quality total score of MARS=3.66; mean app quality total score of uMARS=3.86). Suzhou Red Ribbon obtained the lowest composite score, with a score of 2.68 (mean app quality total score of MARS=3.67; mean app quality total score of uMARS=1.68).

Comparative Analysis of MARS Score and uMARS Score

Figure 3 presents a correlation analysis between the scores of MARS and uMARS. No significant relationship was observed between the app total quality scores of MARS and uMARS (0.41; P=0.07) or between their subjective quality scores (0.39; P=0.08). The app most reflective of the disparity between MARS and uMARS is Suzhou Red Ribbon, which had the greatest discrepancy between its app total quality scores on MARS (3.67) and uMARS (1.68). Its score on uMARS (1.68) was the lowest in the ranking, with respective dimensional scores of engagement (1.8), functionality (1.5), information (1.75), and subjective quality (1). Conversely, for MARS, it ranked seventh for the app quality total score, with respective dimensional scores of engagement (2.9), functionality (4.38), aesthetics (3.5), information (3.92), and subjective quality (2.5). There was considerable variation disclosed in the rankings of the most highly rated apps between MARS and uMARS. The top 5 app total quality scores on MARS were Wu Ai Fang Hua (4.06), Hong Feng Wan (3.96), Ai Yi Jian (3.95), Qing Ai Health Services (3.77), and Ai Zhiku (3.76), whereas the top 5 on uMARS were Hong Feng Wan (4.47), Chabei (3.86), Wu Ai Fang Hua (3.74), Linqu County CDC (3.66), and Zhecheng County AIDS consulting and test (3.62).

Interpreting composite data from Figures 4 and 5 and Table 3 reveal that within the 4 MARS scale dimensions of the evaluated 21 apps, the functionality dimension achieved the highest score (4.28, SD 0.39).

The information (mean 3.82, SD 0.33), aesthetics (mean 3.4, SD 0.49), and engagement (mean 2.4, SD 0.53) dimensions sequentially trailed behind. In contrast, under the uMARS scale’s evaluation, the information dimension prevailed with the top score (mean 3.89, SD 0.74), followed by the functionality (mean 3.84, SD 0.59), aesthetics (mean 3.21, SD 0.48), and engagement (mean 2.56, SD 0.63) dimensions.

Figure 3. Results of correlation analysis between the Mobile App Rating Scale (MARS) and the User Mobile App Rating Scale (uMARS) scores.
Figure 4. Box plot of the Mobile App Rating Scale (MARS) score.

Figure 5. Box plot of the User Mobile App Rating Scale (uMARS) score.
### Table 3. The MARS\(^a\) and uMARS\(^b\) scales’ scores for apps. The top 5 apps with app quality total score for each dimension are italicized.

<table>
<thead>
<tr>
<th>App name</th>
<th>App quality ratings</th>
<th>MARS</th>
<th>uMARS</th>
<th>MARS</th>
<th>uMARS</th>
<th>MARS</th>
<th>uMARS</th>
<th>MARS</th>
<th>uMARS</th>
<th>MARS</th>
<th>uMARS</th>
<th>MARS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ai Cheng Wang Shi</td>
<td></td>
<td>2.00</td>
<td>2.10</td>
<td>4.25</td>
<td>4.00</td>
<td>2.33</td>
<td>3.33</td>
<td>3.50</td>
<td>3.88</td>
<td>3.02</td>
<td>3.33</td>
<td>1.88</td>
</tr>
<tr>
<td>Ai Yi Jian(^c)</td>
<td></td>
<td>3.00</td>
<td>2.80</td>
<td>4.63</td>
<td>4.00</td>
<td>4.00</td>
<td>3.17</td>
<td>4.17</td>
<td>4.13</td>
<td>3.95</td>
<td>3.52</td>
<td>3.38</td>
</tr>
<tr>
<td>Ai Zhiku(^c)</td>
<td></td>
<td>2.50</td>
<td>2.30</td>
<td>4.38</td>
<td>4.00</td>
<td>4.00</td>
<td>3.17</td>
<td>4.17</td>
<td>4.38</td>
<td>3.76</td>
<td>3.46</td>
<td>3.00</td>
</tr>
<tr>
<td>Baiyi HIV test</td>
<td></td>
<td>2.30</td>
<td>2.60</td>
<td>4.63</td>
<td>4.00</td>
<td>3.50</td>
<td>3.33</td>
<td>3.67</td>
<td>4.00</td>
<td>3.52</td>
<td>3.48</td>
<td>2.50</td>
</tr>
<tr>
<td>Beijing AIDS Association</td>
<td></td>
<td>2.40</td>
<td>2.20</td>
<td>3.50</td>
<td>4.00</td>
<td>3.33</td>
<td>3.17</td>
<td>3.65</td>
<td>4.25</td>
<td>3.22</td>
<td>3.40</td>
<td>2.50</td>
</tr>
<tr>
<td>Chabei(^d)</td>
<td></td>
<td>2.20</td>
<td>2.70</td>
<td>4.50</td>
<td>4.50</td>
<td>3.67</td>
<td>3.50</td>
<td>4.29</td>
<td>4.75</td>
<td>3.66</td>
<td>3.86</td>
<td>3.00</td>
</tr>
<tr>
<td>Danlan Happy Test</td>
<td></td>
<td>1.60</td>
<td>2.30</td>
<td>3.25</td>
<td>3.75</td>
<td>2.84</td>
<td>3.00</td>
<td>3.33</td>
<td>2.25</td>
<td>2.63</td>
<td>2.83</td>
<td>1.13</td>
</tr>
<tr>
<td>Dou Ai Jian</td>
<td></td>
<td>1.90</td>
<td>1.80</td>
<td>4.13</td>
<td>3.25</td>
<td>3.00</td>
<td>2.67</td>
<td>3.45</td>
<td>3.25</td>
<td>3.12</td>
<td>2.74</td>
<td>2.25</td>
</tr>
<tr>
<td>E Ai Jian</td>
<td></td>
<td>2.30</td>
<td>2.70</td>
<td>4.75</td>
<td>4.00</td>
<td>3.84</td>
<td>3.33</td>
<td>3.74</td>
<td>4.38</td>
<td>3.66</td>
<td>3.60</td>
<td>2.25</td>
</tr>
<tr>
<td>Red Ribbon Volunteer House</td>
<td></td>
<td>2.00</td>
<td>2.50</td>
<td>3.88</td>
<td>4.00</td>
<td>3.00</td>
<td>3.17</td>
<td>3.37</td>
<td>4.13</td>
<td>3.06</td>
<td>3.45</td>
<td>2.13</td>
</tr>
<tr>
<td>Liaocheng Dongchangfu District Anti-AIDS Service Platform</td>
<td></td>
<td>2.00</td>
<td>2.20</td>
<td>4.50</td>
<td>4.00</td>
<td>3.50</td>
<td>3.33</td>
<td>4.25</td>
<td>4.25</td>
<td>3.56</td>
<td>3.45</td>
<td>2.25</td>
</tr>
<tr>
<td>Linqu County CDC(^d)</td>
<td></td>
<td>2.10</td>
<td>2.60</td>
<td>4.63</td>
<td>4.00</td>
<td>3.17</td>
<td>3.67</td>
<td>3.80</td>
<td>4.38</td>
<td>3.42</td>
<td>3.66</td>
<td>2.50</td>
</tr>
<tr>
<td>Nanyue Gaozhibao</td>
<td></td>
<td>1.90</td>
<td>2.20</td>
<td>4.00</td>
<td>4.00</td>
<td>2.84</td>
<td>3.17</td>
<td>3.67</td>
<td>3.75</td>
<td>3.10</td>
<td>3.28</td>
<td>1.75</td>
</tr>
<tr>
<td>Qing Ai Health Services(^c)</td>
<td></td>
<td>2.90</td>
<td>3.30</td>
<td>4.50</td>
<td>4.50</td>
<td>3.84</td>
<td>3.17</td>
<td>3.83</td>
<td>3.88</td>
<td>3.77</td>
<td>3.59</td>
<td>2.63</td>
</tr>
<tr>
<td>Rong Ai Jian</td>
<td></td>
<td>2.30</td>
<td>2.40</td>
<td>4.63</td>
<td>3.88</td>
<td>3.84</td>
<td>3.33</td>
<td>3.33</td>
<td>3.38</td>
<td>3.52</td>
<td>3.25</td>
<td>2.13</td>
</tr>
<tr>
<td>Suzhou Red ribbon</td>
<td></td>
<td>2.90</td>
<td>1.80</td>
<td>4.38</td>
<td>1.50</td>
<td>3.50</td>
<td>1.67</td>
<td>3.92</td>
<td>1.75</td>
<td>3.67</td>
<td>1.68</td>
<td>2.50</td>
</tr>
<tr>
<td>Wu Ai Fang Hua(^c,d)</td>
<td></td>
<td>3.20</td>
<td>3.30</td>
<td>4.63</td>
<td>4.00</td>
<td>4.17</td>
<td>3.67</td>
<td>4.25</td>
<td>4.00</td>
<td>4.06</td>
<td>3.74</td>
<td>3.38</td>
</tr>
<tr>
<td>Zhecheng County AIDS consulting and test(^d)</td>
<td></td>
<td>2.40</td>
<td>2.60</td>
<td>4.00</td>
<td>4.00</td>
<td>3.17</td>
<td>3.50</td>
<td>4.09</td>
<td>4.38</td>
<td>3.41</td>
<td>3.62</td>
<td>1.88</td>
</tr>
<tr>
<td>Life4me+</td>
<td></td>
<td>1.80</td>
<td>2.00</td>
<td>4.13</td>
<td>3.50</td>
<td>2.67</td>
<td>2.67</td>
<td>3.64</td>
<td>3.63</td>
<td>3.06</td>
<td>2.95</td>
<td>2.25</td>
</tr>
<tr>
<td>Hong Feng Wan(^c,d)</td>
<td></td>
<td>3.60</td>
<td>4.70</td>
<td>4.25</td>
<td>4.25</td>
<td>3.67</td>
<td>4.17</td>
<td>4.33</td>
<td>4.75</td>
<td>3.96</td>
<td>4.47</td>
<td>4.13</td>
</tr>
<tr>
<td>Xiao Ai</td>
<td></td>
<td>3.10</td>
<td>2.70</td>
<td>4.38</td>
<td>4.00</td>
<td>3.50</td>
<td>3.33</td>
<td>3.88</td>
<td>4.25</td>
<td>3.71</td>
<td>3.57</td>
<td>2.38</td>
</tr>
</tbody>
</table>

\(^a\)MARS: Mobile App Rating Scale.  
\(^b\)uMARS: User Mobile App Rating Scale.  
\(^c\)The top 5 apps in terms of app quality total score in MARS.  
\(^d\)The top 5 apps in terms of app quality total score in uMARS.

**Comparative Analysis of Quality Between WeChat Mini-Apps and Independent Apps**

As shown in Table 4, insignificant disparities are discernible between WeChat mini-apps and stand-alone apps regarding 2 metric domains: app quality ($P_{\text{MARS}}=.70$; $P_{\text{uMARS}}=.54$) and subjective app quality ($P_{\text{MARS}}=.48$; $P_{\text{uMARS}}=.80$).
Table 4. Independent samples $t$ test for unequal variances.

<table>
<thead>
<tr>
<th>MARS$^a$ quality</th>
<th>MARS, mean (SD)</th>
<th>uMARS$^b$, mean (SD)</th>
<th>MARS, $t$ value (df)</th>
<th>uMARS, $t$ value (df)</th>
<th>MARS, $P$ value</th>
<th>uMARS, $P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>App quality</td>
<td>---</td>
<td>---</td>
<td>-0.448 (2.4)</td>
<td>0.73 (2.3)</td>
<td>.70</td>
<td>.54</td>
</tr>
<tr>
<td>WeChat applets</td>
<td>3.45 (0.37)</td>
<td>3.33 (0.50)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Independent apps</td>
<td>3.58 (0.47)</td>
<td>3.66 (0.76)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>App subjective quality</td>
<td>---</td>
<td>---</td>
<td>-0.852 (2.2)</td>
<td>0.289 (2.3)</td>
<td>.48</td>
<td>.80</td>
</tr>
<tr>
<td>WeChat applets</td>
<td>2.39 (0.57)</td>
<td>2.72 (0.58)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Independent apps</td>
<td>2.92 (1.05)</td>
<td>2.88 (0.90)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

$^a$MARS: Mobile App Rating Scale.

$^b$uMARS: User Mobile App Rating Scale.

$^c$Not available.

Internal Consistency and Reliability Testing for MARS and uMARS Scores

Similarly, we formed 4 Bland-Altman plots using the disparities and mean values between the scorings by reviewers on app quality (both MARS and uMARS) and app subjective quality (MARS and uMARS). The relatively limited range of the 95% limits of agreements suggest that the evaluators’ judgment outputs contain minor dissimilarities. A significant proportion of dots in all 4 Bland-Altman plots lie within the concordance interval (Figures 6-9), with their arithmetic means impressingly approaching 0 (Table 5).

This provides evidence of the high degree of internal uniformity and dependability in both MARS and uMARS scores.

Figure 6. Bland-Altman plot of the app quality (Mobile App Rating Scale).
Figure 7. Bland-Altman plot of the app subjective quality (Mobile App Rating Scale).

Figure 8. Bland-Altman plot of the app quality (User Mobile App Rating Scale).
Table 5. Summary of results of Bland-Altman analysis.

<table>
<thead>
<tr>
<th>Index</th>
<th>95% CI (LoA)</th>
<th>95% CI (arithmetic mean)</th>
<th>95% CI (LoA: upper limit)</th>
<th>95% CI (LoA: lower limit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>App quality (MARS&lt;sup&gt;b&lt;/sup&gt;)</td>
<td>−0.19 to 0.21</td>
<td>−0.04 to 0.06</td>
<td>0.13 to 0.30</td>
<td>−0.28 to −0.11</td>
</tr>
<tr>
<td>App subjective quality (MARS)</td>
<td>−0.89 to 1.03</td>
<td>−0.15 to 0.29</td>
<td>0.64 to 1.41</td>
<td>−1.27 to −0.50</td>
</tr>
<tr>
<td>App quality (uMARS&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>−0.38 to 0.64</td>
<td>0.01 to 0.25</td>
<td>0.44 to 0.85</td>
<td>−0.58 to −0.17</td>
</tr>
<tr>
<td>App subjective quality (uMARS)</td>
<td>−0.66 to 1.12</td>
<td>0.02 to 0.43</td>
<td>0.76 to 1.47</td>
<td>−1.02 to −0.30</td>
</tr>
</tbody>
</table>

<sup>a</sup>LoA: limits of agreement.  
<sup>b</sup>MARS: Mobile App Rating Scale.  
<sup>c</sup>uMARS: User Mobile App Rating Scale.

Discussion

Principal Findings

This study provides a comprehensive statistical summary of the functionalities and characteristics of 21 Chinese HIV-related mobile apps, along with their quality assessment using the MARS scale. The primary objectives behind the development of these mobile apps were the dissemination of HIV prevention knowledge (18/21, 86%), appointment booking for testing (15/21, 71%), and counseling services (10/21, 48%). These mobile apps aim to assist users in HIV prevention and treatment while improving the quality of communication and interaction between patients and health care providers.

Overview of App Functions

In recent years, the Chinese government has implemented a series of regulations and policies for HIV/AIDS prevention and has established an effective mechanism for HIV/AIDS control. These efforts have achieved certain progress in HIV/AIDS prevention. However, the burden of AIDS disease continues to increase, and the prevention and control situation remains severe. With the widespread adoption of smartphones and the rapid development of mobile internet, mobile apps have played an increasingly important role in disseminating HIV/AIDS knowledge and raising public awareness of prevention [1,37].

Among the 21 apps included in this study, 18 (86%) had HIV/AIDS knowledge dissemination functionality. This indicates that HIV/AIDS knowledge dissemination is an almost indispensable feature of Chinese HIV/AIDS mobile apps, which have become important tools for promoting HIV/AIDS knowledge in China.

In this study, 17 (81%) of the 21 mobile apps were identified, including 14 (67%) WeChat applets (such as Chabei, Danlan Happy Test, Baiyin HIV test, Suzhou Red Ribbon, Linqu County CDC, E Ai Jian, Ai Yi Jian, Rong Ai Jian, Liaocheng Dongchangfu District Anti-AIDS Service Platform, Zhecheng County AIDS consulting and test, Qing Ai Health Service, Ai Cheng Wang Shi, Beijing AIDS Association, and Dou Ai Jian) and 3 (14%) independent apps (Life4me+, Hong Feng Wan, and Xiao Ai). These apps provided various forms of health services, including appointment booking, counseling, risk assessment, and PrEP or PEP. A total of 2 mobile apps (Wu Ai Fang Hua and Hong Feng Wan) offered web-based community functionalities and scored excellently in terms of app quality and subjective quality, ranking among the top 2 positions. This finding is consistent with that of a previous study by Nour et al [38]. Therefore, we recommend considering the incorporation of web-based community functions when developing and designing mobile apps related to AIDS. This can encourage users to share their experiences, ask questions, provide
suggestions, and offer mutual assistance and support. Organizing regular web-based or offline activities can also promote face-to-face interactions among users, thereby enhancing their sense of belonging and engagement.

Privacy protection is another crucial topic that should be given sufficient attention in AIDS-related mobile apps. As social stigmatization and discrimination are still associated with AIDS, many users are concerned about their personal information being disclosed or misused, posing a threat to their privacy and security [39]. Therefore, unlike general mobile apps, ensuring user privacy protection in AIDS mobile apps is of utmost importance, aligning with previous research findings [40]. However, in this study, the 21 apps, only 5 (24%) mobile apps, including 2 (10%) WeChat applets (Chabei and Su Zhou Red Ribbon) and 3 (14%) apps (My Life+, Hong Feng Wan, and Xiao Ai), provided privacy policies or agreements. The average subjective quality score (mean 2.85, SD 0.77) was significantly higher than that of the mobile apps without privacy protection features (mean 2.34, SD 0.58). This observed phenomenon may be attributed to the fact that inadequate privacy protection design may lead to user wariness and lower intentions and frequencies of use regarding AIDS mobile apps [41]. Therefore, we strongly suggest that all AIDS-related mobile apps incorporate privacy policies or agreements to alleviate user concerns about privacy protection, enhance user trust, and promote willingness to use these apps. In addition to transparent privacy policies and user agreements, user education is also critical. Developers of AIDS mobile apps should provide users with relevant information and educational resources on privacy protection. This can be achieved through in-app prompts, tutorials, and frequently asked questions. Users should receive clear instructions on configuring their privacy settings, handling their personal data, and the steps to take in reporting privacy concerns or data breaches. These instructions should be easily accessible, comprehensible, and provide specific examples to guide users through each process.

Rating of App Quality

According to the results from MARS scores, among the 4 dimensions of app quality, both researchers and patients with HIV rated 21 mobile apps lowest on engagement (mean\textsubscript{MARS} 2.4, SD 0.53; mean\textsubscript{uMARS} 2.56, SD 0.63), especially the WeChat mini-apps, which had the lowest average scores (mean\textsubscript{MARS} 2.33, SD 0.43; mean\textsubscript{uMARS} 2.47, SD 0.42). The average engagement scores for the 7 mobile apps developed by the Chinese government were even lower (mean\textsubscript{MARS} 2.27, SD 0.33; mean\textsubscript{uMARS} 2.34, SD 0.30). This underlined the concern that the level of engagement would remain low, even for officially developed apps. An analysis of the respective scores for each item within the MARS and uMARS engagement portions revealed a predominant lack of interest and amusement within these mobile apps, thus leading to decreased user engagement. Many AIDS-related mobile apps merely provide basic information and functions without interactive and stimulating designs to attract user participation, lacking a strong appeal to users. In addition, the absence of personalized and customized features limits user engagement. A study has shown a significant correlation between user engagement and an increase in the adoption rate of mobile apps [42]. Therefore, it is recommended that developers not only provide high-quality HIV/AIDS prevention and treatment information but also focus on meeting users’ needs in terms of multidimensionality, functionality, and depth. In addition, attention should be paid to design in terms of amusement, entertainment, customization, interactivity, and other participatory aspects. This will help attract more user participation and enhance user stickiness.

This investigation revealed a lack of correlation between the MARS and uMARS scores. Specifically, (1) the app with the lowest overall uMARS scores across all dimensions surprisingly ranked seventh in terms of MARS scores. (2) A notable discrepancy was found in the functionality dimension ratings among patients with HIV and researchers, exposing the highest demographic variance in this attribute. (3) Moreover, the apps predominantly preferred by patients with HIV exhibited robust performance in functionality and information dimensions, with the latter appearing particularly predominant. However, the apps gravitating toward researchers demonstrated high competence in functionality, esthetics, and information, with functionality being the most superior. These data imply possible significant divergences distinct between the evaluations of researchers and genuine users of the apps. To maintain superior app quality and consumer satisfaction, rigorous surveillance of app quality should be sustained from both research and real-user perspectives. This bifocal assessment permits the accurate identification of genuine user requirements and researcher appraisal, providing valuable insights for the pinpoint and scientific enhancement of both app quality and user experience.

In this study, we observed that most Chinese AIDS-related mobile apps are WeChat applets. A reason for this is that the WeChat applet has advantages such as not requiring downloading or installation, having minimal resource consumption, and high user retention rates, which independent apps do not possess [43]. Another reason is that the HIV/AIDS population is relatively niche [44], which means that the market for AIDS mobile apps represents a low-frequency demand niche market. Independent apps often overlook this niche market owing to their high development costs, whereas WeChat applets, with their low development costs and the ability to cater to low-frequency demand niches through segmented scene construction, can effectively meet the needs of this market. Although WeChat mini-apps dominate the landscape of mobile apps for HIV/AIDS in China, their comprehensive performance lags behind that of stand-alone apps. The app quality score (mean\textsubscript{MARS} 3.45, SD 0.73 vs 3.58, SD 0.47; mean\textsubscript{uMARS} 3.33, SD 0.50 vs 3.66, SD 0.76) and subjective quality score (mean\textsubscript{MARS} 2.39, SD 0.57 vs 2.92, SD 1.05; mean\textsubscript{uMARS} 2.72, SD 0.58 vs 2.88, SD 0.90) for WeChat mini-apps are both lower compared with stand-alone apps. A granular analysis of the scores in different dimensions reveals the greatest discrepancy in the engagement dimension, with WeChat mini-apps scoring markedly lower (mean\textsubscript{MARS} 2.33, SD 0.43 vs 2.83, SD 0.93; mean\textsubscript{uMARS} 2.47, SD 0.42 vs 3.13, SD 1.40). This can be attributed to the constraints and limitations imposed by the WeChat platform, which prevent WeChat applets from providing the same user experience as independent apps in terms of...
interface design, interaction methods, and customization settings. Improvement in these aspects should be a key focus for the future development of AIDS-related WeChat applets.

Limitations
The mobile apps evaluated in this study represent a snapshot of the current status of Chinese HIV-related mobile apps during the research period. Over time, mobile apps may be removed or updated, so the list of included mobile apps meeting the criteria may change in the future. In addition, some mobile apps intended solely for internal organizational use or no longer available in the market were excluded.

Conclusions
This study presents a systematic introduction to the functionality and quality of the currently available Chinese mobile apps for AIDS, providing valuable decision-making support for AIDS-related groups and health care professionals in accessing high-quality AIDS mobile apps. Through our systematic search and evaluation of existing Chinese mobile apps for AIDS, it has been observed that because of the lower demand frequency of AIDS mobile apps in China, less costly WeChat mini-apps have become the primary mode of app, and the overall quality attains a merely average level. A significant implication of our research is identifying the potentially significant discrepancy between the assessments made by researchers and the authentic users of the apps. Consequently, the inclusion of genuine users during the assessment and refinement stages of HIV apps is crucial. The main purpose of developing these mobile apps is to spread HIV prevention knowledge and facilitate booking appointments for testing and counseling services. However, most of these apps lack privacy protection features. Unlike general mobile apps, privacy protection is especially crucial in AIDS-related mobile apps because it directly affects users’ willingness to use them [40]. Therefore, the introduction of legal and ethical frameworks for privacy protection as well as privacy protection technologies is essential. In addition, enhancing user education on privacy protection and ensuring informed consent is of utmost importance. Research related to privacy protection in Chinese AIDS mobile apps may be a vital and urgent topic to address in the future. It is our intent that the findings of our research may function as a road map and reference for the future development of HIV apps in China. Furthermore, we aim to provide crucial decision-making support for individuals living with HIV in their quest for superior HIV apps.

Acknowledgments
The authors sincerely acknowledge the editors and anonymous reviewers for their insights and comments to further improve the quality of the manuscript. The authors thank Wang Xiang, chief of the AIDS Prevention Section of Bengbu Centers for Disease Control and Prevention, for his great help in the investigation. Finally, the first author (PL) would like to personally express his gratitude to his son (Liu Daixin) and his family for their support all through this work. This work was supported by the Anhui Provincial University Philosophy and Social Science Major Project (grant 2023AH040283).

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
PL and FW initiated this study. PL and FW contributed to the study design. PL drafted the manuscript. LW conducted data sorting and screening. All authors revised the manuscript and approved the final version of the manuscript.

Conflicts of Interest
None declared.

References


23. Chan PS, Chidgey A, Lau J, Ip M, Lau JT, Wang Z. Effectiveness of a Novel HIV self-testing service with online real-time counseling support (HIVST-online) in increasing HIV testing rate and repeated HIV testing among men who have sex with


Abbreviations

CDC: Centers for Disease Control and Prevention
GBMSM: gay, bisexual, and other men who have sex with men
HIVST: HIV self-testing
MARS: Mobile App Rating Scale
mHealth: mobile health
PEP: postexposure prophylaxis
PrEP: pre-exposure prophylaxis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
uMARS: User Mobile App Rating Scale
Abstract

Background: With the rapid development of mobile health (mHealth) technology, many health apps have been introduced to the commercial market for people with back pain conditions. However, little is known about their content, quality, approaches to care for low back pain (LBP), and associated risks of use.

Objective: The aims of this research were to (1) identify apps for the self-management of LBP currently on the market and (2) assess their quality, intervention content, theoretical approaches, and risk-related approaches.

Methods: The UK iTunes and Google Play stores were initially searched for apps related to the self-management of LBP in May 2022. A repeat search in June 2023 was conducted to ensure that any relevant new apps developed in the last year were incorporated into the review. A total of 3 keywords recommended by the Cochrane Back and Neck Group were used to search apps “low back pain,” “back pain,” and “lumbago.” The quality of the apps was assessed by using the 5-point Mobile App Rating Scale (MARS).

Results: A total of 69 apps (25 iOS and 44 Android) met the inclusion criteria. These LBP self-management apps mainly provide recommendations on muscle stretching (n=51, 73.9%), muscle strengthening (n=42, 60.9%), core stability exercises (n=32, 46.4%), yoga (n=19, 27.5%), and information about LBP mechanisms (n=17, 24.6%). Most interventions (n=14, 78%) are consistent with the recommendations in the National Institute for Health and Care Excellence (NICE) guidelines. The mean (SD) MARS overall score of included apps was 2.4 (0.44) out of a possible 5 points. The functionality dimension was associated with the highest score (3.0), whereas the engagement and information dimension resulted in the lowest score (2.1). Regarding theoretical and risk-related approaches, 18 (26.1%) of the 69 apps reported the rate of intervention progression, 11 (15.9%) reported safety checks, only 1 (1.4%) reported personalization of care, and none reported the theoretical care model or the age group targeted.

Conclusions: mHealth apps are potentially promising alternatives to help people manage their LBP; however, most of the LBP self-management apps were of poor quality and did not report the theoretical approaches to care and their associated risks. Although nearly all apps reviewed included a component of care listed in the NICE guidelines, the model of care delivery or embrace of care principles such as the application of a biopsychosocial model was unclear.

Introduction

Low back pain (LBP) is a complex multifactorial disorder, often considered a combination of physical, psychological, and social dysfunction [1]. A multidisciplinary self-management intervention based on a biopsychosocial model holds significant potential to manage LBP [2,3] and has been demonstrated to be more effective than unimodal exercise therapy [4].
National Institute for Health and Care Excellence (NICE) guidelines recommend using self-management for LBP, which can be described as the patient’s proactive adoption of strategies to manage their symptoms and monitor their health and well-being [5]. Despite being a promising approach to managing LBP, it can be challenging for an individual to self-manage any long-term condition [6]. The adherence to self-management strategies is commonly poor, especially without support, feedback, and positive reinforcement [7,8]. A qualitative study also noted poor adherence to advice and exercises as a limiting factor to recovery from LBP [9].

With the increasing popularity of electronic products, digital health solutions such as smartphone apps can be used as an innovative way to support self-management for many conditions, including LBP and may provide a solution to some of the problems outlined above [6,10]. Mobile health (mHealth) apps for pain management may be beneficial to patients [11,12], helping monitor those with acute or chronic pain and providing them with information and support for pain management. However, while many mHealth apps have been introduced into the commercial marketplace to manage pain, most have not been regulated in a uniform or standardized way before being released to the market [12,13]. The involvement of health care professionals in their development and content has been lacking [14]. This has raised concerns about the quality of these mHealth apps and whether their content information is evidence-based [15,16]. It is therefore desirable to assess the quality of current apps in the UK market and whether their content aligns with guideline recommendations. Another concerning issue is the paucity of evidence on specific intervention approaches in developed mHealth apps that have been developed, including underpinning evidence and theory [17] and relative risk management [18]. This potentially impacts on safety and efficacy of health-related smartphone apps [19].

Since there is no unified framework for assessing the theoretical and risk-based approaches associated with LBP self-management applications, we developed a theoretical framework that considers the theoretical care model of the intervention, the personalization of care, and the rate of intervention progression, as well as a risk-related framework that includes the targeted age group and the provision of appropriate safety checks. Interventions based on specific theoretical frameworks are known to be more effective in health care [20]. The Medical Research Council (MRC) guidelines strongly recommend using theoretical approaches in designing complex interventions [21], thus its inclusion in the evaluation. A critical factor in mHealth apps is personalization, where the management or treatment of the disease is tailored to the patient’s situation and individual needs, which will make the user feel that it is relevant and meaningful to them [22,23], thus the inclusion of criteria in relation to personalization. Treatment methods such as exercise need to be both adaptable and progressive. This means starting with basic exercises and gradually advancing to more complex levels at a pace that patients with LBP can comfortably handle. Such a progression helps enhance their functional abilities and quality of life, while simultaneously equipping them with strategies to effectively manage their pain. [24]. Since the treatment paths for LBP vary according to different age groups, including children, adults, and the elderly [25], the target user age group must be defined. Finally, patient safety is an essential component of health care provision and is critical to primary care management, which will effectively reduce the clinical risks associated with LBP management [18,26].

Therefore, this app review, aimed to first, identify apps for the self-management of LBP currently on the market and second, to assess their quality (eg, functionality and design), intervention content (compliance with best practice guidelines), underlying theory (eg, theoretical care model), and risk-related approaches (eg, the age group targeted).

**Methods**

**Search Strategy**

Apps currently on the market for the self-management of LBP were identified, reviewed, and analyzed using a systematic approach. The UK official app stores for both Apple’s iOS and Google’s Android OS were used to search for mobile apps. These 2 operating systems currently dominate the marketplace of mobile medical apps [27]. We logged into Apple iTunes and Android Google Play stores in May 2022. A total of 3 keywords recommended by the Cochrane Back and Neck Group [28] were used to search for apps “low back pain,” “back pain,” and “lumbago.” A subsequent search in June 2023 was repeated to ensure that any relevant new apps in the last year were incorporated into the review.

**Study Selection**

Criteria for inclusion in the review were (1) apps were a self-contained product (ie, did not depend on an external device or add-ons), (2) apps offered at least 1 active treatment option for LBP (eg, unsupervised exercise program or patient education), (3) apps only designed for people with LBP, (4) apps created or updated in the last 5 years to ensure software functionality and ongoing technical support, and (5) apps developed in English. Exclusion criteria were (1) apps targeted at managing general chronic pain, (2) apps only focused on risk factors and diagnostic tests for LBP, (3) apps only focused on specific LBP pathologies (eg, lumbar disk herniation), (4) apps designed for clinicians, (5) general back fitness apps with no mention of physiotherapy or physical therapy or musculoskeletal (MSK) conditions, and (6) apps were not downloadable or had restrictions (eg, requiring an active access password). Apps that incurred a cost were also included; however, when both paid and free versions of an app were available, we reviewed only the paid version to ensure access to the full content. If the same app was available on iOS and Android, the iOS version was kept for inclusion and analysis.

An independent reviewer initially screened the eligible apps based on the apps’ names and descriptions and the screenshots provided. After the preliminary screening phase, the same independent reviewer downloaded apps that met the eligibility criteria for a second screening. Concerns regarding inclusion were discussed and resolved within the research group (AM and DS) until a final decision was reached.
**Data Extraction**

The selected apps were downloaded onto either a Samsung SM-N975F (Android version 7.1.2) or an iPhone 12 (iOS version 16.5) for a complete assessment of eligibility and characteristics. Relevant background information offered in the included apps was extracted, including the name, version, developer, update date, cost, and presence of in-app purchases. We extracted the age or content rating and consumer rating (5-star rating system) when available. In addition, we extracted if the apps contained advertisements and whether these adverts were relevant to their back pain. It was also noted if the apps were asynchronous or synchronous (available with support from a provider), and whether the apps tracked user engagement. The collection of personally identifiable information by apps and whether consent is stated were also checked. The category of management content, specific intervention component, theoretical care model, personalization of care, the rate of intervention progression, the age group targeted, and safety checks were also extracted.

**App Content Assessment**

Main categories and specific components of LBP app management content were identified and classified. Frequency analysis was performed to determine the number of apps providing these intervention contents. The recently updated 2020 NICE guideline for LBP was used to assess whether the included apps provided evidence-based interventions (categorized as “yes/no”) [29]. For this, we mapped app interventions to recommendations listed in the NICE guidelines. This guideline provides the most recent best practice recommendations for assessing and managing LBP and sciatica in people aged 16 years or older [5].

**App Quality Assessment**

The Mobile App Rating Scale (MARS) was used to assess the quality of included apps in this review. MARS is a brief tool with a 23-item questionnaire to classify and assess the quality of mHealth apps for researchers, professionals, and clinicians [30]. It assesses app quality across 4 domains: engagement, functionality, aesthetics, and information quality. All items are assessed on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent). The MARS has demonstrated excellent internal consistency and interrater reliability for evaluating the quality of mHealth apps [30]. To standardize the quality ratings, the assessor completed a MARS video training recommended by the developers of MARS [30]. A total of 10 apps were randomly selected for training until a consensus on the scores was reached.

**App Assessment of the Theoretical and Risk-Related Approaches**

A total of 5 aspects considered in our theoretical and risk-related approaches were assessed. These included the underpinning LBP management theory, tailoring of content, the intervention of progression approach, the age group the app targeted, and appropriate safety checks.

**Ethical Considerations**

This study does not involve human participants.

**Results**

**App Selection**

The searches performed in May 2022 yielded 392 apps from 2 platforms. Of these, 156 apps were identified from iTunes stores, and 236 apps were identified from Google Play stores. A total of 5 duplicates were removed, resulting in 387 apps identified for screening based on the titles and app descriptions. After initial screening, 319 apps were excluded. The eligible 68 apps were downloaded for a full evaluation and further 7 apps were excluded. The subsequent search in June 2023 found a total of 33 newly developed apps based on the initial search, of which 8 apps were newly developed. Finally, 69 apps were included in this review, of which 25 were iOS apps and 44 were Android apps. Figure 1 illustrates the selection procedure of smartphone apps for LBP.
Characteristics of Included Apps

Of the 69 apps included in this review, 25 (36%) were found on iTunes exclusively, 44 (64%) on Google Play exclusively, and 5 (7%) were found on both app stores. There are 5 (7%) of 69 apps that required payment, ranging in price from US $1.13 to US $22.87 (median US $8.17). The majority (n=64, 93%) of the apps included were free of charge. Of these, 8 offered in-app purchases ranging from US $3.80 to US $12.70. Android apps (n=41, 93%) were more often free to access full functionality than iOS apps (n=15, 60%). Of the 41 apps reviewed by customers on a 5-star rating system, the median customer rating in 19 apps from iTunes (4.3 stars) was higher than in 22 apps from Google Play (4.1 stars). With respect to age or content rating, most of the included iOS apps (n=19, 76%) were downloaded without any age limitation; 4 apps were restricted to those of 12 years or above and 3 apps were restricted to those of at least 17 years old. However, all Android apps were labeled as suitable for all age groups.

In terms of developers, there is a mix of some health care groups and other private companies. A total of 24 (35%) apps contained advertisements, 10 (42%) of which were for products or medical companies targeting MSK disorders, leaving 14 (58%) random advertisements. In addition, all apps were asynchronous which means that they failed to deliver continuously updated application data to users. A total of 9 (13%) apps collect personally identifiable information from users and only 2 (22%) asked for their consent for collection. It appeared that none of the apps tracked user engagement. The characteristics of each app are presented in Multimedia Appendix 1.

MARS Quality Assessment

The mean MARS total score obtained from 69 applications was 2.4 out of 5 (SD 0.44). Table 1 summarizes the MARS total scores for each app. All apps were initially assessed using MARS for engagement, functionality, aesthetics, and information. Mean scores for each subscale (out of 5) were calculated. Of the 5 categories, apps scored highest in the functionality (mean 3.0, SD 0.55) domain, followed by aesthetics (mean 2.6, SD 0.61) and engagement (mean 2.1, SD 0.58). The information domain received the lowest score (mean 2.1, SD 0.46). The MARS total score and domain score for each app are shown in Multimedia Appendix 2.
Table 1. Mobile app rating scale scores (N=69 apps).

<table>
<thead>
<tr>
<th>MARSa subscale</th>
<th>iOS, mean (SD)</th>
<th>Android, mean (SD)</th>
<th>Total, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement</td>
<td>2.4 (0.52)</td>
<td>1.9 (0.55)</td>
<td>2.1 (0.58)</td>
</tr>
<tr>
<td>Functionality</td>
<td>3.1 (0.57)</td>
<td>3.0 (0.49)</td>
<td>3.0 (0.55)</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>2.9 (0.57)</td>
<td>2.4 (0.47)</td>
<td>2.6 (0.61)</td>
</tr>
<tr>
<td>Information</td>
<td>2.3 (0.49)</td>
<td>1.9 (0.40)</td>
<td>2.1 (0.46)</td>
</tr>
<tr>
<td>MARS overall scoreb</td>
<td>2.6 (0.43)</td>
<td>2.3 (0.35)</td>
<td>2.4 (0.44)</td>
</tr>
</tbody>
</table>

*MARS: Mobile App Rating Scale.
bAverage of 4 objective subscales.

Intervention Contents for LBP

The LBP interventions embedded with the included apps were mainly classified into 3 categories (Table 2). Of these, most (n=47, 68.1%) of the apps offered only an exercise program, while 14.5% (n=10) apps provided patient education alone, and 13.0% (n=9) apps recommended patient education in combination with an exercise program. The remaining 2.9% (n=2) apps provided some psychological intervention for LBP in combination with an exercise program. Finally, only 1.4% (n=1) app-prescribed combinations of patient education, exercise, and psychological intervention.

Table 2. Number and percentage of category for low back pain interventions used in included apps.

<table>
<thead>
<tr>
<th>Main category</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient education + exercise program + psychological intervention</td>
<td>1 (1.4)</td>
</tr>
<tr>
<td>Exercise program + patient education</td>
<td>9 (13.0)</td>
</tr>
<tr>
<td>Exercise program + psychological intervention</td>
<td>2 (2.9)</td>
</tr>
<tr>
<td>Patient education only</td>
<td>10 (14.5)</td>
</tr>
<tr>
<td>Exercise program only</td>
<td>47 (68.1)</td>
</tr>
</tbody>
</table>

More specifically, of the 69 apps included in this review, most (n=51, 73.9%) apps recommended muscle stretching as a self-management strategy for LBP. A total of 42 (60.9%) apps suggested muscle strengthening, 32 (46.4%) apps offered core stability exercises, and 19 apps (27.5%) recommended using yoga to manage LBP. Also, there were 17 (24.6%) apps providing information about LBP mechanisms, followed by advice to use medication (n=9, 13%), staying active (n=9, 13%), postural therapy (n=8, 11.6%), cold and heat therapy (n=8, 11.6%), and aerobic exercise (n=8, 11.6%). Only some apps mentioned manual therapy (n=4, 5.8%), cognitive behavioral therapy (n=4, 5.8%), meditation (n=4, 5.8%), mindfulness (n=3, 4.3%), McKenzie exercise (n=3, 4.3%), electrotherapy (n=3, 4.3%), acupuncture (n=3, 4.3%), and lifestyle advice (n=2, 2.9%). Concerning app intervention content, 14 (78%) of 18 interventions for LBP were aligned with the NICE guidelines and 4 included interventions that were not endorsed by the NICE guideline: postural therapy, electrotherapy, cold and heat therapy, and acupuncture [29]. Details of the interventions offered for LBP managed in the included apps are summarized in Table 3.
Table 3. Number and percentage of specific component for low back pain interventions used in included apps.

<table>
<thead>
<tr>
<th>Specific components</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding LBP&lt;sup&gt;a&lt;/sup&gt; mechanisms</td>
<td>17 (24.6)</td>
</tr>
<tr>
<td>Staying active</td>
<td>9 (13.0)</td>
</tr>
<tr>
<td>Postural therapy</td>
<td>8 (11.6)</td>
</tr>
<tr>
<td>Lifestyle advice</td>
<td>2 (2.9)</td>
</tr>
<tr>
<td>Electrotherapy</td>
<td>3 (4.3)</td>
</tr>
<tr>
<td>Cold and heat therapy</td>
<td>8 (11.6)</td>
</tr>
<tr>
<td>Manual therapy</td>
<td>4 (5.8)</td>
</tr>
<tr>
<td>Medication use</td>
<td>9 (13.0)</td>
</tr>
<tr>
<td>Core stability exercise</td>
<td>32 (46.4)</td>
</tr>
<tr>
<td>Muscle strengthening</td>
<td>42 (60.9)</td>
</tr>
<tr>
<td>Muscle stretching</td>
<td>51 (73.9)</td>
</tr>
<tr>
<td>McKenzie exercise</td>
<td>3 (4.3)</td>
</tr>
<tr>
<td>Aerobic exercise</td>
<td>8 (11.6)</td>
</tr>
<tr>
<td>Yoga</td>
<td>19 (27.5)</td>
</tr>
<tr>
<td>Mindfulness</td>
<td>3 (4.3)</td>
</tr>
<tr>
<td>Meditation</td>
<td>4 (5.8)</td>
</tr>
<tr>
<td>CBT&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4 (5.8)</td>
</tr>
<tr>
<td>Acupuncture</td>
<td>3 (4.3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>LBP: low back pain.
<sup>b</sup>CBT: cognitive behavioral therapy.

**Theoretical and Risk-Related Approaches**

None of the 69 apps included in this review referred to or explained their theoretical care model. Only 1 app considered or incorporated a tailored approach to their intervention. A total of 18 (26%) of these apps provided an intervention program based on principles of gradual intervention progression. No app mentioned the age group for which their intervention content was appropriate and none set an age limit for their use. Even fewer, 11 (16%) apps offered safety checks for app users, including identifying red flags, signs, and symptoms of LBP that required medical attention or providing safety-netting advice if the back pain did not resolve or worsen (Multimedia Appendix 3).

**Discussion**

**Intervention Content in LBP Apps**

We attempted to benchmark the content of the included apps against the most recently published best practice guideline from NICE and found that most LBP self-management app components were consistent with those recommended by NICE. The NICE LBP guidelines [29] recommended the use of a group exercise program, including biomechanical, aerobic, mind-body, or a combination of approaches. Accordingly, muscle stretching, strengthening, core stability exercises, and yoga are the most common interventions in self-management apps. The findings of this review support a 2016 systematic review, which found that interventions in LBP selected by app developers were primarily based on clinical practice guidelines [31]. Similarly, the 2021 Cochrane review also reported moderate-certainty evidence suggesting that different types of exercise therapy are effective in treating LBP [32]. This means that most current LBP self-management apps offer evidence-based interventions. However, while the content of most apps appears evidence-based, they are not often delivered in the context of a complex intervention and as such do not reflect the current LBP care models. We found that most self-management apps rely on exercise interventions, and very few apps incorporate social and psychological interventions for managing LBP. This suggests that current self-management apps emphasize a more biological care model to manage LBP rather than considering the influence of psychological and social factors in the development and maintenance of pain [33]. Research has emphasized interrelationships among biological changes, psychological status, and the sociocultural context and as such, they all need to be considered to understand the impact of chronic pain and its subsequent management [34]. The biopsychosocial model has been widely accepted as a holistic approach to increase efficacy and outcomes in managing chronic LBP [2,35]. Also, a complex multidisciplinary approach with a biopsychosocial model has been recommended in the early stages of LBP to reduce the likelihood of chronicity following acute LBP [36]. Thus, digital self-management interventions for LBP should consider adopting this model.
Quality Assessment in Self-Management Apps

Generally, apps for the self-management of LBP are of poor quality as assessed by MARS. Functionality (mean 3.0, SD 0.55) was the domain that scored the highest on the MARS test, as described by other authors [31,37]. This implies that the apps are functioning well, easy to learn and navigate, and efficient. However, the quality assessment revealed that these apps had low scores on the domains of “engagement” (mean 2.1, SD 0.58) and “aesthetics” (mean 2.6, SD 0.61). This indicates that the features that make the app equally engaging and important to a wide user base may have been overlooked. This was partly because most apps did not consider using specific strategies to increase involvement and aesthetics from the user’s point of view (eg, entertainment, interactivity, customization, layout, and graphics). Involving patient users during the development of these apps might better identify their needs and characteristics, and increase adherence to improve self-management and health outcomes [38].

The lowest score on information (mean 2.1, SD 0.46) indicated weakness in the quality and trustworthiness of information presented in the included apps. This was evidenced through assessing credibility (MARS item 18) and evidence base (MARS item 19). In terms of MARS item 18, most apps were developed by either commercial businesses with a vested interest or a legitimate source without verification (eg, has no web page), yet few are developed by credible health organizations (eg, government or universities). The lack of health professional involvement is a consistent issue highlighted by Rizwana, who has expressed concern about the accuracy and trustworthiness of in-app information [39]. In addition, none of the apps available for LBP management were evaluated using a randomized controlled trial. This was evident in MARS item 19, which assesses whether the app has been trialed or tested in the published scientific literature, and therefore the effectiveness and safety of these apps remain unknown. These results align with previous reviews on mHealth apps directed at pain management, in which a lack of scientific basis of the outcomes was found to support the use of such apps [40,41]. A possible explanation could be that most of the apps are of commercial rather than scientific origin, which suggests that the need to promote cross-disciplinary collaboration between academic and commercial institutions might help develop the evidence base for using such apps [37,40].

Theoretical and Risk-Related Approaches

Current self-management apps do not appear to have used a theoretical rationale in their development. This aligns with the findings from a 2018 review, which found the development of current self-management mHealth apps for patient education programs lacks the support of underpinning theory or framework [42]. An underpinning theoretical model is widely recognized as a crucial component of health interventions and is important when trying to understand key components of the intervention, how they interact, and the mechanisms of the intervention [43-45]. Systematic reviews of existing evidence demonstrate that interventions underpinned by theory are more effective than those that are not [46,47]. Additionally, as a complex multifactorial condition, the management of LBP should consider theory development and identifying the evidence base in accordance with the MRC framework for complex interventions [48].

Regarding personalized care, almost all included apps provided limited customized service. Tailored communications provide individuals with information that is relevant to them and that fits their situation. This can lead to increased perceived personal relevance, user engagement, more in-depth processing of information, and consequently, more desire and motivation to engage in the health behavior change [49]. The importance of personalization of mHealth apps was emphasized in a qualitative review in 2019, which expressed that mHealth apps should meet patients’ needs since they were created for use [22]. Also, our review suggests that the principle of gradual progression from simple to more advanced levels of intervention is not universal. Providing tools that help the user implement exercise progression ensures that the intervention progressively becomes more challenging to continually stimulate adaptations and maintain interest in the program [24]. The Coventry, Aberdeen, and London—Refined (CALO-RE) taxonomy of behavior change techniques highlighted the importance of setting graded tasks, breaking targeted behaviors into smaller, more manageable tasks, and facilitating progress in small increments [50].

Current apps do not evaluate the age limitation of intervention content for which LBP advice and treatments are appropriate, which may pose some risks for the users. Although age ratings in the App Store are often reported, it is based primarily on the degree to which an app contains sensitive information rather than on the applicability of an intervention to different age groups. It is also worth noting that the current NICE guideline for LBP published in 2016 is based on an adult population. A review from the Lancet showed that the evidence underpinning LBP guidelines is drawn almost exclusively from clinical trials of adults, and there are few trials on the treatment of back pain in children [51]. Thus, adult LBP care pathways may not be suitable for adolescents and children [52,53]. When applying these LBP management apps to nonadults, their scope of application needs to be considered. In addition, most included apps performed poorly in providing safety considerations. Safety check advice, including information on the natural history of the illness, advice on worrying symptoms to watch out for, and specific information on how and when to seek help, as well as advice about the follow-up of investigations and hospital referrals, can effectively address uncertainty in the process of LBP management [54]. A scoping review in 2020 discussed that safety concerns within apps were a primary concern [18]. Thus, such approaches will improve the likelihood of providing users with appropriate care and reduce clinical risks associated with self-management [55].

Evaluation Tool for LBP Apps

Although some self-management apps show relatively high scores on the MARS score, such as “Back Doctor/Pain Relief-1.03.24,” “Perfect Posture&Healthy back-1.5.2,” and “Back pain exercise at home-1.0.99,” they perform poorly on the theoretical and risk-related framework, such as the personalization of care and the age group targeted. Conversely,
some self-management apps (eg, BackTrainer-2.0) adopted a tailored approach and addressed risk-related issues, but MARS’ quality assessment scores were not high. Additionally, MARS has not effectively evaluated or included the biopsychosocial care model commonly used in chronic pain [34]. This indicates that current evaluation tools may be limited in their ability to assess LBP self-management apps comprehensively, and consequently, further study is needed to explore whether a holistic tool to evaluate LBP self-management apps is required.

Barriers to Holistic Digital Apps

Despite the overwhelming evidence recommending the use of holistic multidisciplinary interventions based on a biopsychosocial model for the self-management of LBP, the proliferation of back pain apps on the market that use simple interventions is concerning. It is hard to speculate why this is the case given the strong evidence for the use of a biopsychosocial model in the management of LBP. This may relate to current regulatory approvals. For an app to deliver a complex biopsychosocial intervention it would be classified as a medical device requiring such approvals. Further work is needed to explain why there are so many apps on the market that fail to address the model of care we seek in medical practice, and how current regulatory processes affect this.

Moreover, digital health interventions have been strongly advocated for and implemented in other domains and countries. Notably, draft guidance from NICE has informed digital programs using multidisciplinary models to assist the NHS in delivering specialized services for weight management [56]. Furthermore, the Federal Institute for Drugs and Medical Devices in Germany has authorized the entry of multidisciplinary health care apps with robust trial data into the market, with the overarching objective of empowering clinicians to prescribe health care apps for their patients [57,58]. This illustrates the feasibility of developing multidisciplinary mHealth apps for health care practitioners to recommend. However, it remains imperative to conduct a comprehensive evaluation of the apps using clinical studies to ascertain their efficacy and suitability for widespread prescription.

Limitations

Apps not specifically targeted for the self-management of LBP (eg, chronic pain, pain management, or MSK apps) were excluded from this review, which may result in potentially eligible apps being missed. While many LBP apps in different countries are available in app stores, our search was limited to the UK iTunes and Google Play stores for practical reasons relating to data-capturing capacity. Therefore, this review did not include apps exclusively available in other countries, and therefore may not represent the broader landscape relating to digital approaches for management of LBP. Only 1 independent reviewer assessed app quality using MARS and discussed uncertainties with 2 other research members until a consensus was reached, which may impact the reliability of the assessments.

Conclusions

In this review, we identified 69 apps related to the self-management of LBP and rated them using MARS. Most apps scored poor quality due to their approaches to engagement and information, and many emerged as tools for delivering passive information rather than active management apps. Most apps were aligned with guideline-based care. However, no app offered a holistic self-management intervention approach incorporating a biopsychological model. Most apps underestimated the importance of theoretical and risk-related aspects. Thus, current self-management apps for LBP are limited in what they offer.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Imperial College London is the sponsor for the study and has no influence on the direction or content of the work. We are grateful to Rebecca Jones (Library Manager in Imperial College London) who offered advice and support for the building search strategy and screening process. We also wish to thank our colleagues in the MSK laboratory for their valuable and helpful comments and suggestions. Without these people, studying would have been impossible.

Data Availability

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

Authors’ Contributions

TZ, DS, and AM conceived and designed the study. TZ takes responsibility for the acquisition of the data and the analysis of the data. TZ drafted the paper. DS and AM revised the paper critically for important intellectual content. All authors approved the final version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Characteristics of apps for the self-management of low back pain.
References


Abbreviations

CALO-RE: Coventry, Aberdeen, and London—Refined
LBP: low back pain
MARS: Mobile App Rating Scale
mHealth: mobile health
MRC: Medical Research Council
MSK: musculoskeletal
NICE: National Institute for Health and Care Excellence
Original Paper

Short-Term Effect of a Health Promotion Intervention Based on the Electronic 12-Hour Dietary Recall (e-12HR) Smartphone App on Adherence to the Mediterranean Diet Among Spanish Primary Care Professionals: Randomized Controlled Clinical Trial

Luis María Béjar¹, PhD; Pedro Mesa-Rodríguez², BSc; María Dolores García-Perea³, PhD

¹Department of Preventive Medicine and Public Health, Institute of Anatomy, School of Medicine, University of Seville, Seville, Spain
²Camas Health Center, Seville, Spain
³Virgen Macarena University Hospital, Seville, Spain

Corresponding Author:
Luis María Béjar, PhD
Department of Preventive Medicine and Public Health
Institute of Anatomy
School of Medicine, University of Seville
3rd Floor, Sánchez-Pizjuán Avenue
Seville, 41009
Spain
Phone: 34 954551771
Fax: 34 954556481
Email: lmbprado@us.es

Abstract

Background: The World Health Organization has called for addressing the growing burden of noncommunicable diseases (NCDs) by promoting healthy lifestyles among the population. Regarding patient health, primary care professionals (PCPs) are the first line of care who can positively influence patients’ behavior and lifestyle habits. However, a significant percentage of PCPs do not lead a healthy lifestyle. Therefore, addressing their health behaviors may be the key to substantially increasing health promotion advice in general practice. The Mediterranean diet has been extensively studied, and there is strong evidence of it being a dietary pattern for the prevention of NCDs, in addition to its significant environmental, sociocultural, and local economics benefits.

Objective: This study focused only on the dietary aspect of the PCPs’ lifestyle. The primary objective was to evaluate the effect of using the Electronic 12-Hour Dietary Recall (e-12HR) smartphone app to improve diet, specifically to promote adherence to the Mediterranean diet (AMD), among PCPs. The secondary objectives were to establish the usability of the e-12HR app and to determine AMD among PCPs.

Methods: An individual-level randomized, controlled, and single-blind clinical trial was conducted with 2 parallel groups: a control group (CG), using the nonfeedback version of the e-12HR app, and an intervention group (IG), using the feedback version of the e-12HR app. The level of human involvement was fully automated through the use of the app. There was a 28-day follow-up period. Participants were PCPs (medicine or nursing) recruited offline at one of the selected primary care centers (Andalusia, Spain, Southern Europe), of both sexes, over 18 years old, possessing a smartphone, and having smartphone literacy.

Results: The study response rate was 73% (71 of 97 PCPs), with 27 (38%) women and 44 (62%) men: 40 (56%) PCPs in the CG and 31 (44%) in the IG. At baseline, AMD was medium (mean Mediterranean Diet Serving Score [MDSS] index 9.45, range 0-24), with 47 (66%) PCPs with a medium/high MDSS index. There were significant statistical improvements (CG vs IG, in favor of the IG) at week 4 (no significant statistical differences at baseline): +25.6% for the MDSS index (P=.002) and +213.1% for the percentage with a medium/high MDSS index (P=.001). In relation to specific food groups, there were significant statistical improvements for fruits (+33.8%, P=.02), vegetables (+352%, P=.001), nuts (+184%, P=.02), and legumes (+75.1%, P=.03). The responses to the usability rating questionnaire were satisfactory.
Conclusions: The results support recommending the use of the e-12HR app as a tool to contribute to improving diet and preventing NCDs among PCPs, while positively influencing patient dietary behavior and preventing diet-related NCDs among patients.

Trial Registration: ClinicalTrials.gov NCT05532137; https://clinicaltrials.gov/study/NCT05532137

KEYWORDS: primary care professionals; Mediterranean diet; smartphone applications; smartphone apps; health promotion; Mediterranean diet adherence; food group

Introduction

Major chronic noncommunicable diseases (NCDs; eg, cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes) are responsible for 74% of deaths worldwide, making them the leading cause of preventable mortality [1]; they also reduce patients’ well-being and activity, which contributes to a poor quality of life, disability, and reduced productivity [1]. Notably, these NCDs share common risk factors, such as an unhealthy diet, smoking, harmful alcohol consumption, and physical inactivity [1], all of which are behavioral or lifestyle-related factors that are potentially modifiable [2]. For this reason, the World Health Organization has called for addressing the growing burden of NCDs by promoting healthy lifestyles [1,3].

Health care workers are a subgroup of the population where lifestyle promotion is essential for 3 main reasons. The first reason is their own health: although health care professionals do their best to provide exceptional patient care, they often fail to prioritize their own self-care [4], resulting in unhealthy behaviors associated with a high risk of NCDs. Studies of health care workers’ lifestyles in hospitals, for example, showed high rates of overweight/obesity [5-7], low fruit and vegetable consumption [6,7], low physical activity [5-8], tobacco [5,7] and alcohol consumption [6,7], and high levels of stress and insufficient rest [4-6]. In addition, almost half had more than 2 of these risk factors [5,6].

The second reason is the success of health care organizations: health care workers with better and healthier habits have been shown to have higher personal and job satisfaction and fewer sickness-related absences from work [9]. The third reason is patient health: several studies have described that health care professionals who practice healthy behaviors offer more advice on healthy habits to patients who come to their practice, that they have more assertive attitudes when counselling patients, and they provide more comprehensive and aggressive counselling, which can positively influence patients’ health [10-14].

Regarding patient health, primary care is the cornerstone for preventing NCDs through health education and plays an essential role in the success of therapeutic medicine [11]. Primary care professionals (PCPs) are the first line of care who can positively influence patient behavior and lifestyle habits [15], and indeed, patients perceive PCPs (doctors and nurses) as the most trusted source of health information and advice on healthy lifestyles [13]. However, a significant percentage of PCPs, as with hospital health care professionals, do not lead a healthy lifestyle: there is a high prevalence of NCD risk factors, such as smoking [10,16], alcohol abuse [10,16], overweight/obesity [17], physical inactivity [5,10,17], inadequate fruit and vegetable intake [10,11,17], added salt intake [17], and high consumption of ultraprocessed foods [18], sugars, and fats [10]. Therefore, addressing PCPs’ health behaviors may be the key to substantially increasing health promotion advice in general practice [12].

As the evidence shows, PCPs are exposed to several risk factors of NCDs; this study focused only on the dietary aspect of the PCPs’ lifestyle. As a starting point, the research team posed the following question: How can the dietary habits of PCPs be improved?

The research team has previously evaluated the effectiveness of a smartphone app called the Electronic 12-Hour Dietary Recall (e-12HR) in improving diet in Spanish university students (health science [19,20] and non–health science [20]). The main hypothesis of this study was that the use of this app among PCPs can have an influence on improving their diet, as has already been evidenced in previous research among university students [19,20], the null hypothesis being that the use of the app has no influence on improving the diet of PCPs.

To the best of our knowledge, this study is the first to evaluate the use of a smartphone app to improve diet among PCPs. The primary study objective was to evaluate the effect of using the e-12HR app on improving diet among PCPs, specifically to promote adherence to the Mediterranean diet (AMD). The Mediterranean diet has been extensively studied, and there is strong evidence of it being a dietary pattern for the prevention of NCDs [21-25], in addition to its significant environmental, sociocultural, and local economics benefits [26-28]. In addition, secondary objectives were to establish the usability of the e-12HR app and determine AMD among PCPs.

Methods

Overview of the Study

The study was an individual-level randomized, controlled, and single-blind clinical trial with 2 parallel groups: a control group (CG) and an intervention group (IG). All participants used an app called the e-12HR app, with different versions for the CG and the IG. In the CG, participants used the nonfeedback version of the e-12HR app, and in the IG, participants used the feedback version of the e-12HR app (see the Intervention section). The
level of human involvement was fully automated through the use of the app.

All research was carried out in 3 basic health zones of the Andalusian Health Service (Andalusia, Spain, South of Europe): Camas, Coria del Río, and San Juan de Aznalfarache. Several primary care centers were selected in the 3 zones: the Camas, Santiponce, Valencina de la Concepción, Castilleja de Guzmán, Coca de la Piñera, Carambolo, and Pañoleta health centers in Camas; the Coria del Río health center in Coria del Río; and the San Juan de Aznalfarache and Gelves health centers in San Juan de Aznalfarache. The study ran for 28 days, and participant recruitment took place offline from September to October 2022.

Inclusion criteria for the study were both sexes, age over 18 years, possession of a smartphone (iOS or Android operating system), smartphone literacy, and a PCP (medicine or nursing) at one of the selected primary care centers. Exclusion criteria included food intolerance, chronic disease, or pregnancy (due to the possibility of requiring specific dietary recommendations).

**Ethical Considerations**

The study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving study participants were approved by the Andalusian Biomedical Research Ethics Portal (PEIBA) on March 30, 2022 (identifier: 2813-N-21). The trial was registered at ClinicalTrials (identifier: NCT05532137). Written informed consent was obtained from all participants.

**Participant Enrollment**

The project was publicized in the selected primary care centers by a member of the research team, and individual talks were scheduled with interested health care professionals. In each of the individual talks, the study protocol was explained, including the objectives, risks, and benefits of the research, and an email address for the study was provided.

To participate in the study, it was necessary for interested health care professionals to send an email to the designated address, indicating their “interest in participating in the study” and the primary care center where they worked. After receiving the email, a member of the research group sent candidates a series of documents necessary to be able to participate in the study: (1) an informed consent form; (2) a form with personal information (sex, date of birth, primary care center, weight, height, and smoking status), with documents 1 and 2 to be completed, signed, and returned to the same email address; (3) instructions for downloading the e-12HR app (free to download from the App Store or the Play Store); (4) an image of the Mediterranean diet pyramid (with recommendations for consumption by food group); (5) a personal alphanumeric code; and (6) a user’s guide with detailed information for using the app. Document 6 was the only one that differed depending on whether it was intended for participants in the CG (nonfeedback e-12HR version) or the IG (feedback e-12HR version), and obviously, document 5, which contained the personal code, was unique for each participant.

Throughout the study, participants could contact the research team by email with any questions, including questions to reduce the likelihood of harm.

**Participant Allocation**

In each basic health zone, the recruited participants were randomized into 1 of 2 groups (CG or IG) in a ratio of 1:1 as follows: the participant who sent the first email was assigned to the CG, the participant who sent the second email was assigned to the IG, and so on.

This study was single-blind because, due to the nature of the intervention, the PCPs could not be blinded. However, the investigator who performed the statistical analysis of the data was blinded throughout the study. In addition, each participant only had access to 1 version of the app: the personal codes of the participants assigned to the CG activated only the nonfeedback e-12HR version, while the personal codes of the participants assigned to the IG activated only the feedback e-12HR version.

**Intervention**

The structure and functions of the e-12HR app (nonfeedback and feedback versions) have been described in detail by Béjar et al [19]. In this study, we used e-12HR version 3.0. The e-12HR app did not undergo changes throughout the study. In brief, the nonfeedback e-12HR version allows the user to collect food consumption data; however, this version does not provide any feedback to users to promote the Mediterranean diet (ie, this version of the app presents a single function: diet determination). The feedback e-12HR version allows for the collection of food consumption data, and as an additional automatic function, every 7 days the app issues personalized feedback on how to improve AMD (ie, this version of the app has 2 functions: determining the diet and providing feedback to improve AMD). The feedback provided came in 3 parts: (1) the AMD index score; specifically, the Mediterranean Diet Serving Score (MDSS) index [29] (range 0-24); (2) the image of a traffic light: the MDSS index score was divided into 1 of 3 levels (low: score 0-8, red light; medium: score 9-15, orange light; high: score 16-24, green light) [30]; and (3) recommendations for consumption by food group [19]. The nonfeedback e-12HR version did not provide any of the 3 parts of the feedback, as they were exclusive to the e-12HR feedback version. See Multimedia Appendix 1 for real images of the e-12HR app (nonfeedback and feedback versions).

**Follow-up and Outcome Measures**

To analyze the effect of the intervention (CG nonfeedback e-12HR vs IG feedback e-12HR), 4 follow-up points were established: week 1 (baseline), week 2, week 3, and week 4. At each follow-up point, the research team manually calculated the MDSS index for each of the 2 groups from the data provided by the e-12HR app. The method for calculating the MDSS index has been described in detail elsewhere [19].

The main result variable was the change in the total MDSS index at weeks 2, 3, and 4 of monitoring, while the secondary result variables were the personal information variables, the MDSS index at week 1 (baseline), and the answers to the
usability rating questionnaire for the e-12HR app (see the Usability Rating Questionnaire for the e-12HR section).

The MDSS index at week 1 (baseline) was used to determine AMD among the PCPs (a secondary objective of the study). To relativize the data, the MDSS index of the PCPs was compared to the MDSS index of health sciences students. For a proper comparison, the MDSS index was obtained using the same app (e-12HR) and during the same follow-up period (recruitment period: September-October 2022) for both students and health care professionals.

**Usability Rating Questionnaire for the e-12HR**

After the 4-week study period, a member of the research team sent a new email to each PCP who had completed the follow-up; this new email contained a usability rating questionnaire for the e-12HR app [19] (see Multimedia Appendix 2).

**Statistical Analysis**

The sample size was estimated for the main result variable. Assuming SD=2.7 points, dropout rate=20.6% (from a previous study on use of the e-12HR app among health science university students [19]), α=.05, and β=.20 (bilateral test), 82 participants (n=41, 50%, per group) were needed to detect an increase of 2 points in the MDSS index (CG versus IG). The sample size was calculated using nQuery Advisor Release 7.0 (Statsols).

Quantitative variables were expressed as means (SD), and qualitative variables were displayed as numbers (percentages). The nonparametric Kolmogorov-Smirnov test was used for the test for normality.

For unpaired samples and quantitative variables, the Student t test or the nonparametric Mann-Whitney U test was used, and the chi-square test (or Fisher exact test) was used for the comparison of proportions.

For paired samples, quantitative variables, and two groups, the Student t test or the nonparametric Wilcoxon test was used, penalizing P values with Bonferroni adjustment for multiple comparisons. For 3 or more groups, the ANOVA test or the nonparametric Friedman test was used.

P<.05 was considered significant, except for multiple comparisons using Bonferroni penalization: P<.02 (.05/3).

All statistical analyses were performed using the SPSS statistical software package version 26.0 (SPSS Inc).

**Results**

**Sample and Adherence to the Study**

The sequence of allocating participants to the 2 study groups is detailed in Figure 1. In total, 97 PCPs signed the informed consent form (n=50, 52%, in the CG and n=47, 48%, in the IG). Of them, 26 (27%; n=10, 38%, in the CG and n=16, 62%, in the IG) were considered nonresponsive because they did not complete the study’s 4-week follow-up period (Figure 1). The nonresponsive individuals were not included in the later statistical analysis (ie, in this study, per protocol analysis was applied).
The study response rate was 73% (71 of 97 participants), with 40 of 50 (80%) participants in the CG and 31 of 47 (66%) participants in the IG (Figure 1). Participants did not report any harm or unintended effects throughout the study.

### Personal Information of the Participants

Table 1 shows the personal information of the PCPs who completed the study (CG and IG).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All participants (N=71)</th>
<th>CG (n=40)</th>
<th>IG (n=31)</th>
<th>P value¹</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>43.2 (11.6)</td>
<td>45.1 (11.1)</td>
<td>40.7 (12.1)</td>
<td>.07 ²</td>
</tr>
<tr>
<td>&lt;40, n (%)</td>
<td>34 (47.9)</td>
<td>16 (40.0)</td>
<td>18 (58.1)</td>
<td>.13 ³</td>
</tr>
<tr>
<td>≥40, n (%)</td>
<td>37 (52.1)</td>
<td>24 (60.0)</td>
<td>13 (41.9)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>27 (38.0)</td>
<td>16 (40.0)</td>
<td>11 (35.5)</td>
<td>.70 ³</td>
</tr>
<tr>
<td>Male</td>
<td>44 (62.0)</td>
<td>24 (60.0)</td>
<td>20 (64.5)</td>
<td>-</td>
</tr>
<tr>
<td><strong>BMI (kg/m²)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>25.0 (4.0)</td>
<td>25.3 (4.4)</td>
<td>24.7 (3.4)</td>
<td>.76 ³</td>
</tr>
<tr>
<td>&lt;25, n (%)</td>
<td>41 (57.7)</td>
<td>23 (57.5)</td>
<td>18 (58.1)</td>
<td>.96 ³</td>
</tr>
<tr>
<td>≥25, n (%)</td>
<td>30 (42.3)</td>
<td>17 (42.5)</td>
<td>13 (41.9)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Smoking status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>64 (90.1)</td>
<td>37 (92.5)</td>
<td>27 (87.1)</td>
<td>.45 ³</td>
</tr>
<tr>
<td>Yes</td>
<td>7 (9.9)</td>
<td>3 (7.5)</td>
<td>4 (12.9)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Physical activity status (minutes/week), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥150</td>
<td>45 (63.4)</td>
<td>24 (60.0)</td>
<td>21 (67.7)</td>
<td>.50 ³</td>
</tr>
<tr>
<td>&lt;150</td>
<td>26 (36.6)</td>
<td>16 (40.0)</td>
<td>10 (32.3)</td>
<td>-</td>
</tr>
</tbody>
</table>

¹e-12HR: Electronic 12-Hour Dietary Recall.
²AMD: adherence to the Mediterranean diet.
³PCP: primary care professional.
⁴CG: control group.
⁵IG: intervention group.
⁶P<.05 considered significant.
⁷Evaluated with the Mann-Whitney U test.
⁸Evaluated with the chi-square test.
⁹Not applicable.

MDSS Index

As previously mentioned, the MDSS index was calculated manually by the research team (CG and IG, weeks 1, 2, 3, and 4) [19]. During the process, the research team corrected obvious errors made by the PCPs: for example, when the app asked for the number of standard servings of a certain food group consumed on the current day, if the participant entered a value of 150, it was considered that the data referred to milliliters or grams (instead of standard servings). In any case, only 0.2% (91 of 37,772) of the recorded data were corrected.

At week 1 of the monitoring period (baseline), that is, before the IG received the first feedback from the e-12HR app (only week 1 was considered because the feedback for IG in weeks 2, 3, and 4 could affect the alteration of the usual dietary intake),
the PCPs had a mean MDSS index of 9.45 (SD 2.32), which corresponds to a medium level of adherence [30]; moreover, two-thirds of them (n=47, 66%) had a medium/high MDSS value (≥9) at baseline (week 1) [30].

**Effect of the Intervention With the e-12HR App in Terms of Variation in the MDSS Index and Number of Participants With Medium/High (≥12) MDSS Index**

Tables 2 and 3 show the MDSS index, and Table 2 also shows the number of participants with a medium/high MDSS index (≥12) in the CG and the IG throughout the 4 weeks of follow-up. We decided to use the value of 12 (instead of 9) [30] due to the high percentage of PCPs (n=47, 66%) who at baseline (week 1) already had an MDSS index≥9; therefore, using the value of 9 would have made it difficult to observe statistically significant differences between the CG and the IG.

**Table 2.** MDSS\(^a\) index for the CG\(^b\) and the IG\(^c\) and number of participants with a medium/high (≥12) MDSS index throughout the 4 weeks of follow-up.

<table>
<thead>
<tr>
<th>Variables and week number</th>
<th>CG (n=40)</th>
<th>IG (n=31)</th>
<th>(P) value(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MDSS index, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 1</td>
<td>9.30 (2.40)</td>
<td>9.65 (2.23)</td>
<td>.54(^e)</td>
</tr>
<tr>
<td>Week 2</td>
<td>8.98 (2.84)</td>
<td>10.81 (2.82)</td>
<td>.009(^f)</td>
</tr>
<tr>
<td>Week 3</td>
<td>9.08 (2.45)</td>
<td>10.94 (3.05)</td>
<td>.008(^e)</td>
</tr>
<tr>
<td>Week 4</td>
<td>9.30 (2.59)</td>
<td>11.68 (3.61)</td>
<td>.002(^f)</td>
</tr>
<tr>
<td>(P) value(^e)</td>
<td>.99</td>
<td>.01</td>
<td>_ (^h)</td>
</tr>
<tr>
<td><strong>Participants with a medium/high (≥12) MDSS index, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 1</td>
<td>10 (25.0)</td>
<td>10 (32.3)</td>
<td>.50</td>
</tr>
<tr>
<td>Week 2</td>
<td>7 (17.5)</td>
<td>13 (41.9)</td>
<td>.02</td>
</tr>
<tr>
<td>Week 3</td>
<td>5 (12.5)</td>
<td>13 (41.9)</td>
<td>.005</td>
</tr>
<tr>
<td>Week 4</td>
<td>7 (17.5)</td>
<td>17 (54.8)</td>
<td>.001</td>
</tr>
<tr>
<td>(P) value(^e)</td>
<td>.41</td>
<td>.07</td>
<td>—</td>
</tr>
</tbody>
</table>

\(^a\)MDSS: Mediterranean Diet Serving Score.
\(^b\)CG: control group.
\(^c\)IG: intervention group.
\(^d\)\(P\) value in columns: MDSS index—intragroup differences (CG and IG) throughout the 4 weeks of follow-up in the study, evaluated with the Friedman test; number of participants with a medium/high (≥12) MDSS index—intragroup differences (CG and IG) in week 1 versus week 4, evaluated with the chi-square test. \(P<.05\) was considered significant.
\(^e\)Evaluated with the Student \(t\) test.
\(^f\)Evaluated with the Mann-Whitney \(U\) test.
\(^g\)\(P\) value in rows: MDSS index—intragroup differences (CG versus IG) in each of the 4 study weeks; number of participants with a medium/high (≥12) MDSS index—intragroup differences (CG versus IG) in each of the 4 study weeks, evaluated with the chi-square test. \(P<.05\) was considered significant.
\(^h\)Not applicable.
Table 3. Comparison of the MDSS index in weeks 2, 3, and 4 of follow-up with that in week 1 (baseline) for the CG and the IG.

<table>
<thead>
<tr>
<th>Group and week</th>
<th>MDSS index, mean (SD)</th>
<th>P value&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG Week 1</td>
<td>9.30 (2.40)</td>
<td>Reference</td>
</tr>
<tr>
<td>Week 2</td>
<td>8.98 (2.84)</td>
<td>.34&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Week 3</td>
<td>9.08 (2.45)</td>
<td>.58&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Week 4</td>
<td>9.30 (2.59)</td>
<td>.99&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG Week 1</td>
<td>9.65 (2.23)</td>
<td>Reference</td>
</tr>
<tr>
<td>Week 2</td>
<td>10.81 (2.82)</td>
<td>.005&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Week 3</td>
<td>10.94 (3.05)</td>
<td>.004&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Week 4</td>
<td>11.68 (3.61)</td>
<td>.001&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>MDSS: Mediterranean Diet Serving Score.
<sup>b</sup>CG: control group.
<sup>c</sup>IG: intervention group.
<sup>d</sup>Intragroup differences (CG and IG) in week 1 versus weeks 2, 3, and 4. P < .02 (.05/3) was considered significant (penalizing P values with Bonferroni adjustment for multiple comparisons).
<sup>e</sup>Evaluated with the Wilcoxon test.
<sup>f</sup>Evaluated with the Student t test.

Regarding intragroup modifications, there were no significant statistical differences in the MDSS index in the CG, either throughout the 4 weeks of study, ranging from mean 9.30 (SD 2.59) in weeks 1 and 4 to mean 8.98 (SD 2.84) in week 2 (Table 2) or in weeks 2, 3, and 4 when compared to week 1 (baseline; Table 3). In the IG, there were significant statistical differences in the MDSS index throughout the 4 weeks of study, ranging from mean 9.65 (2.23) in week 1 to mean 11.68 (SD 3.61) in week 4 (Table 2). Compared to week 1 (baseline), the differences were statistically significant from week 2 onward: with 1.16, 1.29, and 2.03 points of improvement at weeks 2, 3, and 4, respectively (Table 3). There were no significant statistical differences in the number of participants with a medium/high (≥12) MDSS index (week 1 vs week 4) in either the CG or the IG.

Regarding intergroup modifications, there were significant statistical differences for both the MDSS index and the number of participants with a medium/high (≥12) MDSS index in the CG versus the IG (in favor of the IG) from week 2 onward (no significant differences in week 1). For the MDSS index, we found 1.83, 1.86, and 2.38 points of improvement at weeks 2, 3, and 4, respectively; for the number of participants with a medium/high (≥12) MDSS index, we found 24.4, 29.4, and 37.3 percentage points of improvement at weeks 2, 3, and 4, respectively (Table 2).

**Effect of the Intervention With the e-12HR App in Terms of Variation in Food Groups**

Table 4 shows the number of participants who met the consumption criteria for each food group [19] in the CG and the IG throughout the 4 weeks of follow-up.
Table 4. Number of participants who met the consumption criteria of the MDSS index for each food group throughout the 4 weeks of follow-up (CG n=40, IG n=31).

<table>
<thead>
<tr>
<th>Food group MDSS index consumption criteria and study group</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
<th>P value&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fruits (1-6 servings/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>26 (65.0)</td>
<td>27 (67.5)</td>
<td>27 (67.5)</td>
<td>27 (67.5)</td>
<td>.81&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>24 (77.4)</td>
<td>27 (87.1)</td>
<td>28 (90.3)</td>
<td>28 (90.3)</td>
<td>.17&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.26&lt;sup&gt;g&lt;/sup&gt;</td>
<td>.06&lt;sup&gt;g&lt;/sup&gt;</td>
<td>.02&lt;sup&gt;g&lt;/sup&gt;</td>
<td>.02&lt;sup&gt;g&lt;/sup&gt;</td>
<td>—&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Vegetables (≥2 servings/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>1 (2.5)</td>
<td>3 (7.5)</td>
<td>3 (7.5)</td>
<td>4 (10.0)</td>
<td>.36&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>3 (9.7)</td>
<td>8 (25.8)</td>
<td>10 (32.3)</td>
<td>14 (45.2)</td>
<td>.002&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.31&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.05&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.01&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.001&lt;sup&gt;j&lt;/sup&gt;</td>
<td>—&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Cereals (1-6 servings/day of breakfast cereals, pasta, rice, and bread)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>40 (100.0)</td>
<td>37 (92.5)</td>
<td>39 (97.5)</td>
<td>37 (92.5)</td>
<td>.241&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>31 (100.0)</td>
<td>30 (96.8)</td>
<td>30 (96.8)</td>
<td>30 (96.8)</td>
<td>.99&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value&lt;sup&gt;j&lt;/sup&gt;</td>
<td>—</td>
<td>.63&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.99&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.63&lt;sup&gt;j&lt;/sup&gt;</td>
<td>—&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Olive oil (1-4 servings/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>31 (77.5)</td>
<td>33 (82.5)</td>
<td>33 (82.5)</td>
<td>34 (85.0)</td>
<td>.390&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>26 (83.9)</td>
<td>26 (83.9)</td>
<td>28 (90.3)</td>
<td>27 (87.1)</td>
<td>.99&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.50&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.88&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.50&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.99&lt;sup&gt;j&lt;/sup&gt;</td>
<td>—&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Milk and dairy products (1-3 servings/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>33 (82.5)</td>
<td>32 (80.0)</td>
<td>32 (80.0)</td>
<td>33 (82.5)</td>
<td>.99&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>26 (83.9)</td>
<td>27 (87.1)</td>
<td>27 (87.1)</td>
<td>28 (90.3)</td>
<td>.71&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.88&lt;sup&gt;e&lt;/sup&gt;</td>
<td>.43&lt;sup&gt;e&lt;/sup&gt;</td>
<td>.43&lt;sup&gt;e&lt;/sup&gt;</td>
<td>.50&lt;sup&gt;j&lt;/sup&gt;</td>
<td>—&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Nuts (1-2 servings/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>3 (7.5)</td>
<td>2 (5.0)</td>
<td>2 (5.0)</td>
<td>5 (12.5)</td>
<td>.71&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>3 (9.7)</td>
<td>5 (16.1)</td>
<td>3 (9.7)</td>
<td>11 (35.5)</td>
<td>.02&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.99&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.23&lt;sup&gt;i&lt;/sup&gt;</td>
<td>.65&lt;sup&gt;i&lt;/sup&gt;</td>
<td>.02&lt;sup&gt;e&lt;/sup&gt;</td>
<td>—&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Fermented beverages (0-2 serving/day of wine and beer)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>36 (90.0)</td>
<td>36 (90.0)</td>
<td>37 (92.5)</td>
<td>37 (92.5)</td>
<td>.99&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>27 (87.1)</td>
<td>27 (87.1)</td>
<td>29 (93.5)</td>
<td>29 (93.5)</td>
<td>.67&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.72&lt;sup&gt;i&lt;/sup&gt;</td>
<td>.72&lt;sup&gt;i&lt;/sup&gt;</td>
<td>.99&lt;sup&gt;i&lt;/sup&gt;</td>
<td>.99&lt;sup&gt;i&lt;/sup&gt;</td>
<td>—&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Potatoes (≥3 servings/week)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>32 (80.0)</td>
<td>27 (67.5)</td>
<td>26 (65.0)</td>
<td>25 (62.5)</td>
<td>.08&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>23 (74.2)</td>
<td>19 (61.3)</td>
<td>17 (54.8)</td>
<td>19 (61.3)</td>
<td>.28&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value&lt;sup&gt;j&lt;/sup&gt;</td>
<td>.56&lt;sup&gt;e&lt;/sup&gt;</td>
<td>.59&lt;sup&gt;e&lt;/sup&gt;</td>
<td>.39&lt;sup&gt;e&lt;/sup&gt;</td>
<td>.92&lt;sup&gt;e&lt;/sup&gt;</td>
<td>—&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Legumes (≥2 servings/week)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>15 (37.5)</td>
<td>13 (32.5)</td>
<td>17 (42.5)</td>
<td>14 (35.0)</td>
<td>.82&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>13 (41.9)</td>
<td>18 (58.1)</td>
<td>22 (71.0)</td>
<td>19 (61.3)</td>
<td>.13&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Regarding intergroup modifications, statistically significant differences (CG vs GI) were observed throughout the study period in 4 food groups: fruits (weeks 3 and 4), vegetables (weeks 2, 3, and 4), nuts (week 4), and legumes (weeks 2, 3, and 4). In these 4 food groups, at week 4, the number of participants meeting the recommendations was higher in the IG compared to the CG, with 22.8% for fruits, 35.2% for vegetables, 23.0% for nuts, and, finally, 26.3% for legumes (Table 4).

Regarding intragroup modifications, statistically significant differences (week 1 vs week 4) were observed in the CG for white meat and in the IG for vegetables and nuts.

Usability Rating Questionnaire for the e-12HR App
Of 71 participants, 45 (63%) returned the completed questionnaire: 25 (63%) of 40 from the CG and 20 (65%) of 31 from the IG. The responses of the 45 PCPs are shown in Table 5.
Table 5. Responses to the usability rating questionnaire for the e-12HR app (CG n=25, IG n=20).

<table>
<thead>
<tr>
<th>Questions and groups</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Easy to complete (strongly agree + agree)</td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>25 (100)</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>20 (100)</td>
</tr>
<tr>
<td>P value(^d)</td>
<td>—(^e)</td>
</tr>
<tr>
<td>2. Understandable questions (strongly agree + agree)</td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>24 (96)</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>20 (100)</td>
</tr>
<tr>
<td>P value</td>
<td>.99(^f)</td>
</tr>
<tr>
<td>3. Understandable feedback only for the IG (strongly agree + agree)</td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>—</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>18 (90)</td>
</tr>
<tr>
<td>P value</td>
<td></td>
</tr>
<tr>
<td>4. I would be willing to complete again (strongly agree + agree)</td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>20 (80)</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>12 (60)</td>
</tr>
<tr>
<td>P value</td>
<td>0.141(^g)</td>
</tr>
<tr>
<td>5. Time to complete (≤3 minutes/day)</td>
<td></td>
</tr>
<tr>
<td>CG, n (%)</td>
<td>19 (76)</td>
</tr>
<tr>
<td>IG, n (%)</td>
<td>15 (75)</td>
</tr>
<tr>
<td>P value</td>
<td>.99(^f)</td>
</tr>
</tbody>
</table>

\(^a\) e-12HR: Electronic 12-Hour Dietary Recall.
\(^b\) CG: control group.
\(^c\) IG: intervention group.
\(^d\) Differences between subgroups. \(P<.05\) was considered significant.
\(^e\) Not applicable.
\(^f\) Evaluated with the Fisher exact test.
\(^g\) Evaluated with the chi-square test.

No statistically significant differences were observed for any of the questions on the questionnaire (CG vs IG). All PCPs indicated that the e-12HR app was easy to complete, and most of them responded that the app contained questions that were understandable (CG n=24, 96%; IG n=20, 100%) and that feedback was understandable (only for the IG, n=18, 90%). Furthermore, a large percentage were willing to use the app again (CG n=20, 80%; IG n=12, 60%). Completion of the app could be considered to have taken 3 minutes or less (CG n=19, 76%; IG n=15, 75%).

Discussion

Principal Findings

In relation to the main objective, there were significant statistical differences between the 2 groups in this study. At week 4 (no significant differences in week 1, baseline), the values were higher in the IG compared to the CG by 25.6% for the MDSS index (Table 2); by 213.1 for the number of participants with a medium/high (≥12) MDSS index (Table 2); and by 33.8% for fruits, 352.0% for vegetables, 184.0% for nuts, and 75.1% for legumes for the number of participants meeting the recommendations for specific food groups (Table 4).

Regarding the secondary objectives, first, the answers to the questions of the usability rating questionnaire for the e-12HR app were satisfactory. According to the questionnaire, the daily use time of the app was about 3 minutes or less per day for most respondents (Table 5). When working with smartphone apps, usability is an important aspect to consider. According to health care professionals, there are 3 principal criteria for selecting a “nutrition and diet” app for clients/patients, which are [31] ease of use (satisfactory data were obtained in this study), free of charge, and validation (the e-12HR app is free to download and has been previously validated [32-36]). Second, at baseline (week 1), AMD for PCPs was medium (mean MDSS index 9.45, SD 2.32) and 66% of participants had a medium/high MDSS index (≥9).
Overview
To begin with, workplace interventions are an excellent strategy to promote a healthy diet, considering that health care professionals spend long hours in their professional activity and often have 1 or more meals during their working day. At the hospital level, interventions have been implemented to facilitate access to and choice of healthy foods during the working day, such as modifying the availability of foods served in the canteen, subsidizing the cost of fresh fruits and vegetables (which are often more expensive than less healthy alternatives) [37], or implementing traffic light labeling (green: healthy; yellow: less healthy; red: unhealthy) [37,38]. In a study by Thorndike et al. [38] (in a hospital in Boston, Massachusetts, USA), their intervention also included personalized automated messages using a platform that automatically generated 2 weekly emails with feedback on previous purchases in the hospital cafeteria and lifestyle advice. Significant statistical increases were observed in green-labeled food purchases and decreases in red-labeled food purchases among the IG compared to the CG in the hospital cafeteria throughout the study period.

Thorndike et al’s [38] intervention was based on information about food eaten only in the hospital cafeteria (without considering other food consumed outside the hospital), so its scope was limited. However, to date, no interventions to promote a healthy diet among PCPs have been implemented; for example, the workplace intervention strategies discussed before would be difficult to implement in Spain because health centers are widely distributed throughout the territory and do not usually have a cafeteria or restaurant. Considering these difficulties as a possible alternative strategy, this study was the first to assess the ability of a smartphone app to improve the dietary habits among health care professionals (specifically, PCPs). Several randomized controlled clinical trials have used an app to improve AMD in Spanish adults, such as patients of health care centers (the EVIDENT II app [39,40] and the SalBi Educa Nutrition app [41]) and patients with type 2 diabetes mellitus (the EVIDENT II app [42]), but not in health care professionals.

Comparison With Prior Work
As previously mentioned, the e-12HR app has also been evaluated among Spanish university students (health sciences and non–health sciences) [19,20]. In relation to the main objective, the results obtained by the e-12HR app among PCPs compared to university students were (1) similar for the MDSS index (the increase among PCPs was 25.6% for the MDSS index, as shown in Table 2, and among university students was 17.4% [19] or 25.7% [20]), (2) more positive for the number of participants with a medium/high MDSS index (the increase among PCPs was 213.1%, as shown in Table 2, and among university student was 61.9% [19] or 74.5% [20]), and (3) less positive regarding the number of participants meeting the recommendations for specific food groups (improvements in 4 food group among PCPs, as shown in Table 4, and in 7 food groups among Spanish university students).

Regarding the secondary objectives, first, similar results were obtained among Spanish university students for the answers to the questions of the usability rating questionnaire for the e-12HR app. Second, at baseline (week 1), the mean MDSS index of 9.45 (SD 2.32) and the number of participants with a medium/high MDSS index (66%) among PCPs (Table 1) were higher compared with the data from health sciences students [20] during the same follow-up period (mean MDSS index 7.59, SD 2.72; percentage of participants with a medium/high MDSS index=33.4%). Significant statistical differences were found (P<.05) for both the MDSS index, which was evaluated with the Mann-Whitney U test, and the number of participants with a medium/high MDSS index, which was evaluated with the chi-square test: PCPs showed an improvement of 24.5% for the MDSS index and 98.2% for the number of participants with a medium/high MDSS index (PCPs vs health science university students). This comparison must be made with caution, since the sample of PCPs was made up of doctors and nurses and the sample of health sciences university students was made up of students from the faculties of medicine and pharmacy. In addition, in a previous study by Sentenach-Carbo et al [43] among Spanish PCPs, the number of participants with medium/high AMD was lower: 55% versus 66.2%. It should, however, be considered that the adherence index used by both studies was different: the MDSS index was used in this study, and the validated 14-point Mediterranean diet adhesion screener was used in the Prevention with Mediterranean Diet (PREDIMED) study.

Limitations
This study presents several limitations, and the first is internal validation. This included the fact that the e-12HR app is a self-reporting method and presents the limitations inherent in this type of tool, which have been amply described [44-50]. Due to the nature of the intervention, on the one hand, only the investigator who performed the statistical analysis of the data was blinded (but not the PCPs) and, on the other hand, it was not possible to guarantee that the participants were not using another nutrition app during the study period.

Regarding external validation, the dietary program was short (4 weeks), and the long-term evolution of the study variables is unknown. In addition, the evaluation of the usability of the app was based on the responses of those participants who completed the study; however, there could be differences in the perception of usability between responsive (those who completed the study) and nonresponsive (those who did not) participants.

Future Research
According to Recio-Rodríguez et al [40], future research related to the effectiveness of apps to improve diet should clarify the possible effects of certain factors (eg, age, gender, or educational level). Therefore, in future studies, the research team intends to evaluate the effectiveness of the e-12HR app in increasing the MDSS index in different strata of PCPs—for example, examining results according to gender, age, occupational category, and the BMI as possible moderating variables and according to technological perception and technological familiarity as possible mediating variables.

Conclusion
At baseline, Spanish PCPs presented medium AMD (measured as the MDSS index and the number of participants with a
medium/high MDSS index). Throughout the study period, in the short term, the use of the e-12HR app (an easy-to-implement and low-cost intervention) showed moderate improvements in the MDSS index and remarkable improvements in the number of participants with a medium/high MDSS index; in addition, PCPs responded positively to questions about the usability of the app. These results support recommending the use of the e-12HR app as a tool to contribute to improving diet and preventing NCDs among PCPs, which, at the same time, could positively influence patient dietary behavior and prevent diet-related NCDs among the patients. From the point of view of health care organizations, the prevention of NCDs among PCPs could, in addition, lead to higher personal and job satisfaction and fewer sickness-related absences from work; for this reason, health organizations themselves should be more involved in the recommendations to use tools such as the one analyzed in this study among their own workers.

Acknowledgments
This work was supported in part by the SAMFyC Foundation (grant number TD180/21). The SAMFyC Foundation had no role in the design, analysis or writing of this article. The research team would like to thank the PCPs who participated in this study.

Authors’ Contributions
LMB performed the conception and design of the study, developed the app, analyzed and interpreted data, and wrote the paper. PMR and MDGP were involved in data collection and interpretation of the data and contributed to drafting the manuscript. The authors/evaluators are the owners and developers of the e-12HR app.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Real images of the e-12HR app (nonfeedback and feedback versions). e-12HR: Electronic 12-Hour Dietary Recall.
[DOCX File, 605 KB - mhealth_v12i1e49302_app1.docx]

Multimedia Appendix 2
Usability rating questionnaire for the e-12HR app. e-12HR: Electronic 12-Hour Dietary Recall.
[DOCX File, 28 KB - mhealth_v12i1e49302_app2.docx]

Multimedia Appendix 3
Consort checklist.
[PDF File (Adobe PDF File), 3262 KB - mhealth_v12i1e49302_app3.pdf]

References


Abbreviations

- **AMD:** adherence to the Mediterranean diet
- **CG:** control group
- **e-12HR:** Electronic 12-Hour Dietary Recall
- **IG:** intervention group
- **MDSS:** Mediterranean Diet Serving Score
- **NCD:** noncommunicable disease
- **PCP:** primary care professional
The Association of eHealth Literacy Skills and mHealth Application Use Among US Adults With Obesity: Analysis of Health Information National Trends Survey Data

George Shaw Jr1, PhD; Bianca A Castro2, MSc; Laura H Gunn1, PhD; Keith Norris3, MD, PhD; Roland J Thorpe Jr4, PhD

1Department of Public Health Sciences, School of Data Science, University of North Carolina at Charlotte, Charlotte, NC, United States
2Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, NC, United States
3The University of California Los Angeles Division of General Internal Medicine and Health Services Research, University of California, Los Angeles, Los Angeles, CA, United States
4Department of Health, Behavior and Society, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
George Shaw Jr, PhD
Department of Public Health Sciences
School of Data Science
University of North Carolina at Charlotte
9201 University City Blvd.
Charlotte, NC, 28223-0001
United States
Phone: 1 7045620021
Email: gshaw11@charlotte.edu

Abstract

Background: Physical inactivity and a poor diet are modifiable behaviors that contribute to obesity. Obesity is a well-recognized risk factor for chronic diseases, including diabetes. Mobile health (mHealth) apps can play an important adjuvant role in preventing and treating chronic diseases and promoting positive health behavior change among people with obesity, and eHealth literacy skills have the potential to impact mHealth app use.

Objective: The purpose of this study was to explore the associations between the 2 dimensions, access and application, of eHealth literacy skills and mHealth app use among US adults (≥18 years of age) with obesity (BMI ≥30 kg/m²).

Methods: Data were obtained from February to June 2020 using the Health Information National Trends Survey 5. A total of 1079 respondents met the inclusion criteria of adults with obesity and owners of smartphones. Individual associations between mHealth app use and sociodemographic variables were explored using weighted chi-square and 2-tailed t tests. A multivariable weighted logistic regression model was fitted, and adjusted odds ratios (ORs) of using mHealth apps with corresponding 95% CIs were reported across multiple sociodemographic variables. An Ising model-weighted network visualization was produced. A receiver operating characteristic curve was calculated, and the area under the curve was reported with the corresponding Delong 95% CI.

Results: A majority of respondents were female (550/923, 59.6%) or non-Hispanic White (543/923, 58.8%). Individuals in households earning less than US $50,000 comprised 41.4% (382/923) of the sample. All sociodemographic variables were found to be univariately significant at the 5% level, except employment and region. Results from the multivariable weighted logistic regression model showed that the adjusted odds of using an mHealth app are 3.13 (95% CI 1.69-5.80) and 2.99 (95% CI 1.67-5.37) times higher among those with an access eHealth literacy skill of using an electronic device to look for health or medical information for themselves and an application eHealth literacy skill of using electronic communications with a doctor or doctor’s office, respectively. Several sociodemographic variables were found to be significant, such as education, where adjusted ORs comparing subgroups to the lowest educational attainment were substantial (ORs ≥7.77). The network visualization demonstrated that all eHealth literacy skills and the mHealth app use variable were positively associated to varying degrees.

Conclusions: This work provides an initial understanding of mHealth app use and eHealth literacy skills among people with obesity, identifying people with obesity subpopulations who are at risk of a digital health divide. Future studies should identify equitable solutions for people with obesity (as well as other groups) and their use of mHealth apps.
Introduction

Overview

Physical inactivity and poor dietary behaviors are modifiable behaviors that contribute to obesity [1,2]. Recent studies show that obesity affects nearly 42% of the US population aged 20 years or older [3], with an associated excess annual estimated medical cost of upwards of US $170 billion in 2019 [4]. Obesity is a well-recognized risk factor for chronic diseases, including diabetes, cardiovascular diseases, and cancer, and a significant cause of premature morbidity and mortality [2,5]. Wang et al [6] demonstrated the importance of reducing the weight of patients classified as obese, which is a major contributor to the increased incidence of type 2 diabetes mellitus. Obesity is a complex and multifaceted disease, extending far beyond the realm of individual behavior or mere lifestyle choices. Genetic, environmental, and socioeconomic factors in the development of obesity [7]. However, mobile health (mHealth) apps can play an important adjunct role in preventing and treating chronic diseases and promoting positive health behavior change among individuals with obesity [8,9], with several factors influencing users’ acceptance of and engagement with mHealth apps [10-12]. Previous intervention studies have used mHealth apps to promote health behavior change among individuals with obesity [13,14]. Participants from the reference studies found that the use of mHealth made the intervention helpful and benefited weight loss when used in conjunction with other weight loss intervention methods. mHealth apps provide access to health information and can extend this access to underserved groups, particularly those at higher risk of chronic diseases [15]. However, many individuals with chronic diseases like obesity fail to engage in mHealth app use [16].

mHealth and eHealth Literacy

Istepanian et al [17] defined mHealth as mobile computing, medical sensors, or communication technologies designed for health care. Recent interventions involving mHealth apps provide evidence of improvement in participants’ self-care and disease self-management [18,19]. It is estimated that more than 350,000 mHealth apps are available on the market [20]. Despite the number of digital health interventions targeting weight management, the level to which users actively and regularly engage with those apps entails user engagement early in the intervention design process [21]. While health literacy is a broader concept associated with the ability of individuals to obtain and understand health information to make rational health decisions, eHealth literacy comprises the complex navigation of health care information from internet sources [22,23]. Kontos et al [24] showed that people with lower levels of education were less likely to use the internet to communicate with a doctor or use health information on their mobile devices. Moreover, national and international studies have shown that weight-management apps are beneficial for improving weight loss [25,26]. Previous studies have recommended focusing on understanding the level of health literacy of recipients who may use these apps, particularly those interested in weight loss interventions. Understanding eHealth literacy skills is critical when evaluating health information from mHealth apps and the subsequent application of the knowledge gained [27,28].

eHealth literacy is defined as seeking, finding, understanding, and appraising health information from electronic sources and applying the knowledge acquired to address or solve a health-related problem [29,30]. More recently, eHealth literacy has been conceptualized in dimensions including access to digital services and the application of services and information that satisfy users [31]. People lacking or with a low level of health literacy or eHealth literacy skills benefit less from digital health information and health informatics interventions [32,33], as low eHealth literacy skills were found to impact mHealth app use when mediated through mHealth app efficacy [23]. By contrast, those with higher levels of health literacy and eHealth literacy skills report a positive connection between mHealth app use and health outcomes [34].

mHealth apps are operated across a diverse group of users. This includes individuals with varying incomes, ages, races, ethnicities, and educations [35]. Curating data that include this and other defining personal characteristics require a significant number of resources. Few studies of eHealth literacy and mHealth app use have included a diverse group of participants. Through a national data set of noninstitutionalized adults, this study uses a diverse national data set. The purpose of this study is to explore the potential associations between the 2 dimensions, access and application, of eHealth literacy skills and mHealth app use among a diverse group of US adults (aged ≥18 years) with obesity (BMI ≥30 kg/m²).

Theoretical Underpinning

Seeking health information has become a typical behavior among people of all ages and health conditions [36]. The information obtained when factoring in a person’s eHealth literacy skills, has the potential to influence health outcomes. This is particularly important for people with obesity. People with obesity have lower self-confidence in managing their health. However, people with obesity view communication with their physician as helpful with self-care weight management [37]. Therefore, people with obesity may improve their confidence over time in managing their weight if they have the eHealth literacy skills to seek information from their physician electronically. There have been numerous information behavior theories and models developed to understand how individuals seek and use information [38]. For example, Zare-Farashbandi and Lalazaryan [39] designed their health information acquisition model based on 6 stages of information seeking. The model acknowledges that the information-seeking process can be iterative and that there is a need for a feedback loop in the search process. However, the model does not consider personal or contextual factors affecting information seeking.
Longo’s model of health information considers the effects of personal and contextual factors on the information-seeking behavior of patients [40]. Focusing on patients with chronic diseases, the model was significant in depicting the output process of information-seeking for patients [39]. These theories and models have also considered various social contexts and population groups, such as older individuals, patients with cancer, prisoners, and diabetics [38,41-43]. However, this study is informed by Lenz’s [41] Information Seeking Model, which is the foundation of many fundamental tenets of recent models and instruments aimed at measuring eHealth literacy skills.

According to this framework, in which information gathering is part of the decision-making process, individuals follow 6 stages to seek health information. First, they receive a stimulus from their previous disease experience or the environment. Second, they establish their informational goals, including sources, time available, and the type of information needed. Third, the person decides whether or not to actively access the information they want. The decision is based on the individual’s previous knowledge, background, and the expected cost-benefit of the action. The fourth stage is of particular interest, characterized by the information-seeking action itself. This stage is correlated to the eHealth literacy dimension of access, and it could be an in-depth search or superficial information gathering, depending on the person’s need and previous attempts. The access dimension corresponds to having the availability of digital services that suit people’s needs and work correctly [44]. The fifth stage corresponds to information achievement and interpretation. This stage is related to the eHealth literacy dimension of application, in which the individual understands and appraises or applies the information obtained. In this final stage, people may have to decide on the adequacy of the acquired information [39]. Understanding health information seeking through the potential associations of eHealth literacy skills and mHealth app use may provide insights into how population groups with health disparities with chronic conditions could access and apply the information they seek [45]. Using this framework, the objective of this study is to explore the potential associations between the 2 dimensions, access and application, of eHealth literacy skills and mHealth app use among a diverse group of US adults with obesity.

Methods

Data

The Health Information National Trends Survey (HINTS) was used to explore the potential association between mHealth app use and eHealth literacy skills. HINTS has been administered every few years by the National Cancer Institute since 2003, and the data sets that have been made publicly available are used for evaluating health information access and use among US adults [24,46,47]. HINTS collects representative data about noninstitutionalized US adults’ knowledge, access, attitudes, and use of cancer- and health-related information. The survey uses a 2-stage stratified random sampling that selects households from residential addresses in the United States and then selects 1 adult within each household [24,46]. This study used the HINTS 5, Cycle 4 data set. The data were collected from February to June 2020 and comprised responses from 3865 participants. Despite the COVID-19 impact on society, the response rate for the survey remained high. The response rate (37%) for the survey remained relatively high and was even higher than pre-pandemic HINTS 5 surveys, which experienced response rates of at most 33% [48]. However, COVID-19 impacted the time frame in which the data are typically collected. Individuals included in this analysis were those who indicated ownership of a tablet, smartphone, or both and self-declared a BMI ≥30 kg/m² (obese). The dependent variable was based on respondents’ answers to the following item: “In the past 12 months, have you used any of these health or wellness apps?” The binary variable derived was used to indicate those who reported using any health or wellness apps within the past 12 months and those who did not.

The main independent variables representing eHealth literacy skill’s access and application dimensions were the following four items pertaining to eHealth information and services, connecting to the common stem of “In the past 12 months, have you used a computer, smartphone, or other electronic means to do any of the following: (1) looked for health or medical information for yourself; (2) used email or the internet to communicate with a doctor or doctor’s office; (3) looked medical test results; and (4) made appointments with a health care provider?” Access within this context is the interpretation and appraisal of information aimed at completing an action. Additional covariates extracted include age in years, health insurance status, sex at birth, employment status, marital status, education, annual household income (in ranges), race and ethnicity, and US Census region. These variables have been used in previous studies to evaluate mHealth app use or can be relevant confounders regarding the associations between eHealth literacy skills and mHealth app use [49,50].

Due to low counts, the following categories were combined: employment status of unemployed across lengths of unemployment; employment status of students and others; marital status of separated and divorced; marital status of married and those living as married or with a romantic partner; education categories below 11 years of education; and race and ethnicity categories of non-Hispanic Native Hawaiian or Other Pacific Islander and American Indian or Alaska Native.

Statistical Analysis

In order to investigate the research objective and hypothesis, a comprehensive statistical analysis was performed on the collected data using univariate and multiple logistic regression modeling. The weights provided by HINTS were used to perform all analyses and adjust for sampling biases [51]. A weighted complete case analysis was performed on the data. Characteristics were summarized using means, SDs, counts, and percentages as appropriate. Weighted chi-square and 2-tailed t tests were used to explore univariate associations between each of the covariates and mHealth app use, with test statistics and corresponding P values tabulated. Visualizations were created to explore associations, including (1) Ising model...
network weighted analysis of the associations between the main independent variables (eHealth literacy skill’s access and application covariates) and the outcome; (2) weighted box plot for the continuous covariate (age) and the dependent variable; and (3) multiple weighted 100% stacked bar charts across the main independent variables and the dependent variable. Additional weighted 100% stacked bar charts were constructed (Multimedia Appendix 1) to visualize the sociodemographic variables and the outcome.

The primary study aim is to assess associations between mHealth app use (binary outcome) and each of the eHealth literacy skills dimensions of access and application (main covariates). Univariate analysis is included to provide a comprehensive description of the individual variables in the study and establish a foundation for more complex multivariable analyses. These were further examined using a multivariable weighted logistic regression adjusted for the aforementioned sociodemographic factors. Adjusted odds ratios (ORs), corresponding 95% CIs, and P values were reported across eHealth literacy skills dimensions and sociodemographic variables. Results were tabulated and highlighted using a significance level of 5%. A pseudo-$R^2$ was calculated. The receiver operating characteristic (ROC) is a common approach used to measure the sensitivity versus specificity of logistic models. Additionally, the area under the curve (AUC) is a single metric for that trade-off, with AUC=1 meaning that the model perfectly fits the data and AUC=0.5 indicating there is a split chance that the model fits the data. Both of these approaches are used to evaluate the performance of logistic models. The ROC curve was estimated, and the corresponding AUC value and Delong 95% CI were reported. R software (version 4.0.3; R Foundation for Statistical Computing) was used for statistical analyses.

Ethical Considerations

This research was approved by the institutional review board (IRB) of the University of North Carolina at Charlotte (study #IRB-22-0585). This data set consisted of deidentified, aggregated data. The IRB approval process did not require additional consent from the respondents representing the data.

Results

A total of 1079 participants were identified as obese and owners of a smartphone, tablet, or both. Fewer than 15% (156/1079) of the responses were removed due to incomplete or incoherent data, resulting in 923 complete observations, with mHealth app use (dependent variable), eHealth literacy skills dimensions (main independent variable), and additional covariates summarized in Table 1. The average age was 53.51 (SD 14.91) years, and most participants were female (550/923, 59.6%) and non-Hispanic White (543/923, 58.8%). A college degree or above was the highest level of education for 43.3% (400/923) of study participants, and they were mainly employed (with a single or multiple employer; 594/923, 64.4%) and covered by health insurance (872/923, 94.5%). The South contained the highest level of education for 43.3% (400/923) of the sample, and 18.9% (174/923) of participants had an annual household income that fell within the range, containing the median annual household income in the United States of US $67,521 in 2020 [53].

The majority of participants (482/923, 52.2%) did not use mHealth apps, resulting in a balanced outcome variable. Within the eHealth literacy skills access dimension, 77.5% (715/923) of respondents used an electronic device to look for health or medical information for themselves within the past 12 months, and approximately half (468/923, 50.7%) used electronic means to look up medical test results, also within the past 12 months. Within the eHealth literacy skills application dimension, 55.6% (513/923) of respondents used email or the internet to communicate with a doctor or doctor’s office within the past 12 months, and 53.3% (429/923) made an appointment with a health care provider through electronic means in that same time period. We also examined the univariate association between mHealth app use, covariates, and main independent covariates.

Table 2 summarizes results from weighted chi-square and t tests for univariate associations between mHealth app use and each of the covariates. Most covariates and all the main covariates were found to be significant at the 5% level. All eHealth literacy skills dimensions were found to be significantly associated with mHealth app use based on univariate weighted chi-square tests (P<.001). Similarly, all demographic factors were found to be significantly associated with mHealth app use except for employment status (P=.20) and Census region (P=.16). Figure 1 displays pairwise weighted 100% stacked bar charts for each of the eHealth literacy skills dimensions versus mHealth app use. Figure 2 portrays a joint network representation of the weighted associations between the eHealth literacy skills dimensions and mHealth app use, which demonstrate strong positive associations both between the skills dimensions as well as between those and the outcome (mHealth app use). Figures S1-S9 in Multimedia Appendix 1 include a weighted box plot (age) and weighted 100% stacked bar charts visualizing the univariate associations with mHealth app use.

Table 3 presents the results of the multivariable weighted logistic regression model. The adjusted odds of using an mHealth app are 3.13 (95% CI 1.69-5.80) times higher among those who responded with an access eHealth literacy skill of using an electronic device to look for health or medical information for themselves within the past 12 months. Similarly, those with an application eHealth literacy skill of using email or the internet to communicate with a doctor or doctor’s office within the past 12 months experience 2.99 (95% CI 1.67-5.37) times higher odds of using an mHealth app compared to those without this skill. Sociodemographic factors found to be significantly associated with mHealth app use include age, disabled or retired status, single or never married or widowed, education, and Hispanic ethnicity. Each additional year of age is associated with a 4% lower odds of using mHealth apps (OR 0.96, 95% CI 0.94-0.98). Disabled and retired participants experienced 4.21 (95% CI 1.28-13.82) and 2.53 (95% CI 1.14-5.60) higher odds, respectively, of using mHealth apps compared to those who were employed. Single or never married and widowed participants experienced 49% and 81% lower odds of mHealth app use, respectively, than those who are married, living as married, or living with romantic partners. Previous work has
indicated that surrogate seekers, those who may seek health information on behalf of others, were more likely to be married or have someone close to them with a chronic illness [54]. Those who received more than 11 years of formal education experienced higher odds of mHealth app use than those with 11 years or less, with OR estimates ranging from 7.77 to 17.24, though with substantially wide CIs. Hispanic participants experienced higher odds of using mHealth apps than non-Hispanic White participants (OR 2.61, 95% CI 1.28-5.33). Insurance status, sex at birth, annual household income, and Census region were not found statistically significant upon adjusting for the other covariates, though there is some level of collinearity present among sociodemographic covariates, as demonstrated in the univariate significance of some of these variables. The multivariable weighted logistic regression adjusted for sociodemographic characteristics showed relatively strong explanatory power with a pseudo-$R^2$ of 0.32 and AUC of 0.7957 (95% CI 0.7671-0.8243). The corresponding ROC is included in Multimedia Appendix 1.
<table>
<thead>
<tr>
<th>Sociodemographic variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>53.51 (14.91)</td>
</tr>
<tr>
<td><strong>Health insurance, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Insured</td>
<td>872 (94.5)</td>
</tr>
<tr>
<td>Uninsured</td>
<td>51 (5.5)</td>
</tr>
<tr>
<td><strong>Sex at birth, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>550 (59.6)</td>
</tr>
<tr>
<td>Male</td>
<td>373 (40.4)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Disabled</td>
<td>67 (7.3)</td>
</tr>
<tr>
<td>Employed</td>
<td>494 (53.5)</td>
</tr>
<tr>
<td>Homemaker</td>
<td>25 (2.7)</td>
</tr>
<tr>
<td>Multiple</td>
<td>100 (10.8)</td>
</tr>
<tr>
<td>Retired</td>
<td>183 (19.8)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>42 (4.6)</td>
</tr>
<tr>
<td>Other</td>
<td>12 (1.3)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Married or living as married or with a romantic partner</td>
<td>518 (56.1)</td>
</tr>
<tr>
<td>Separated or divorced</td>
<td>180 (19.5)</td>
</tr>
<tr>
<td>Single or never married</td>
<td>159 (17.2)</td>
</tr>
<tr>
<td>Widowed</td>
<td>66 (7.2)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>≤11 years</td>
<td>49 (5.3)</td>
</tr>
<tr>
<td>12 years or completed high school</td>
<td>174 (18.9)</td>
</tr>
<tr>
<td>Post–high school training other than college (vocational or technical)</td>
<td>68 (7.4)</td>
</tr>
<tr>
<td>Some college</td>
<td>232 (25.1)</td>
</tr>
<tr>
<td>College graduate</td>
<td>241 (26.1)</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>159 (17.2)</td>
</tr>
<tr>
<td><strong>Annual household income (US $), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>0-9999</td>
<td>51 (5.5)</td>
</tr>
<tr>
<td>10,000-14,999</td>
<td>49 (5.3)</td>
</tr>
<tr>
<td>15,000-19,999</td>
<td>37 (4)</td>
</tr>
<tr>
<td>20,000-34,999</td>
<td>113 (12.2)</td>
</tr>
<tr>
<td>35,000-49,999</td>
<td>132 (14.3)</td>
</tr>
<tr>
<td>50,000-74,999</td>
<td>174 (18.9)</td>
</tr>
<tr>
<td>75,000-99,999</td>
<td>128 (13.9)</td>
</tr>
<tr>
<td>100,000-199,999</td>
<td>198 (21.5)</td>
</tr>
<tr>
<td>≥200,000</td>
<td>41 (4.4)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>148 (16)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>171 (18.5)</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>18 (2)</td>
</tr>
<tr>
<td>Non-Hispanic Native Hawaiian or Other Pacific Islander or American Indian or Alaska Native</td>
<td>9 (1)</td>
</tr>
<tr>
<td>Sociodemographic variables</td>
<td>Values</td>
</tr>
<tr>
<td>-----------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>543 (58.8)</td>
</tr>
<tr>
<td>Non-Hispanic multiple races</td>
<td>34 (3.7)</td>
</tr>
<tr>
<td><strong>Census region, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>155 (16.8)</td>
</tr>
<tr>
<td>Northeast</td>
<td>141 (15.3)</td>
</tr>
<tr>
<td>South</td>
<td>431 (46.7)</td>
</tr>
<tr>
<td>West</td>
<td>196 (21.2)</td>
</tr>
<tr>
<td><strong>Outcome variable, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>mHealth app use</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>482 (52.2)</td>
</tr>
<tr>
<td>Yes</td>
<td>441 (47.8)</td>
</tr>
<tr>
<td><strong>Main covariates, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>eHealth literacy skills access dimension</td>
<td></td>
</tr>
<tr>
<td>Electronic health information for self</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>208 (22.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>715 (77.5)</td>
</tr>
<tr>
<td>Electronic test results</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>455 (49.3)</td>
</tr>
<tr>
<td>Yes</td>
<td>468 (50.7)</td>
</tr>
<tr>
<td>eHealth literacy skills application dimension</td>
<td></td>
</tr>
<tr>
<td>Electronic communication with doctor or doctor’s office</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>410 (44.4)</td>
</tr>
<tr>
<td>Yes</td>
<td>513 (55.6)</td>
</tr>
<tr>
<td>Made provider appointments electronically</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>431 (46.7)</td>
</tr>
<tr>
<td>Yes</td>
<td>492 (53.3)</td>
</tr>
</tbody>
</table>
Table 2. Weighted chi-square and 2-tailed $t$ tests (test statistics and $P$ values) for univariate associations between mHealth app use (dependent variable) and each of the covariates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Chi-square (df)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographic factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>3.77 (921)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Health insurance</td>
<td>6.28 (1)</td>
<td>.01</td>
</tr>
<tr>
<td>Sex at birth</td>
<td>6.25 (1)</td>
<td>.01</td>
</tr>
<tr>
<td>Employment status</td>
<td>8.54 (6)</td>
<td>.20</td>
</tr>
<tr>
<td>Marital status</td>
<td>18.55 (3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Education</td>
<td>78.02 (5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Annual household income</td>
<td>34.38 (8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Race and ethnicity</td>
<td>11.31 (5)</td>
<td>.046</td>
</tr>
<tr>
<td>Census region</td>
<td>5.18 (3)</td>
<td>.16</td>
</tr>
<tr>
<td><strong>Main covariates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>eHealth literacy skills: access dimension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eHealth information for self</td>
<td>95.60 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Electronic test results</td>
<td>97.48 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>eHealth literacy skills: application dimension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic communication with doctor or doctor’s office</td>
<td>127.87 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Made provider appointments electronically</td>
<td>81.48 (1)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*a* $t$ test was used for the univariate analysis.
**Figure 1.** Visualization of weighted 100% stacked bar charts for each of the eHealth literacy skills dimensions (main covariates) against mHealth app use (outcome). A: Electronic health information for self; B: Electronic test results; C: Electronic communication with doctor or doctor's office; D: Made appointments electronically.

**Figure 2.** Ising model-weighted network visualization of eLASSO associations (unadjusted by other covariates and with 0.25 penalization factor) between the eHealth literacy skills dimensions (SHI: self-health information; TD: talk to a doctor or doctor’s office; TR: test results; MA: made appointments; and mH: mHealth app use). Thicker edges (lines) between nodes (circles) represent stronger associations.
Table 3. Adjusted odds ratios (ORs), corresponding 95% CIs, and \( P \) values for the multivariable weighted logistic regression model assessing mHealth app use (n=923). The regression model included adjustments for eHealth literacy, age, insurance status, sex, employment and marital status, education, income, race and ethnicity, and census region.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>OR (95% CI)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Explanatory demographic variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.96 (0.94-0.98)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Insured</td>
<td>2.25 (0.94-5.38)</td>
<td>.07</td>
</tr>
<tr>
<td>Male</td>
<td>0.75 (0.46-1.20)</td>
<td>.22</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed (reference)</td>
<td>N/A(^a)</td>
<td>N/A</td>
</tr>
<tr>
<td>Disabled</td>
<td>4.21 (1.28-13.82)</td>
<td>.02</td>
</tr>
<tr>
<td>Homemaker</td>
<td>2.10 (0.66-6.70)</td>
<td>.21</td>
</tr>
<tr>
<td>Multiple</td>
<td>2.16 (0.92-5.09)</td>
<td>.08</td>
</tr>
<tr>
<td>Retired</td>
<td>2.53 (1.14-5.60)</td>
<td>.02</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1.12 (0.44-2.87)</td>
<td>.81</td>
</tr>
<tr>
<td>Other</td>
<td>0.21 (0.04-1.12)</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married or living as married or with a romantic partner (reference)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Separated or divorced</td>
<td>0.67 (0.34-1.33)</td>
<td>.25</td>
</tr>
<tr>
<td>Single or never married</td>
<td>0.51 (0.27-0.96)</td>
<td>.04</td>
</tr>
<tr>
<td>Widowed</td>
<td>0.19 (0.06-0.57)</td>
<td>.003</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \leq 11 ) years (reference)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>12 years or completed high school</td>
<td>7.77 (2.08-29.01)</td>
<td>.002</td>
</tr>
<tr>
<td>Post–high school training other than college</td>
<td>12.75 (3.18-51.17)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Some college</td>
<td>9.25 (2.60-32.98)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>College graduate</td>
<td>14.01 (3.68-53.26)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>17.24 (4.09-72.64)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Annual household income (US $)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(&lt;10,000 ) (reference)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>10,000-14,999</td>
<td>1.67 (0.38-7.45)</td>
<td>.50</td>
</tr>
<tr>
<td>15,000-19,999</td>
<td>0.81 (0.20-3.18)</td>
<td>.76</td>
</tr>
<tr>
<td>20,000-34,999</td>
<td>1.31 (0.41-4.26)</td>
<td>.65</td>
</tr>
<tr>
<td>35,000-49,999</td>
<td>2.47 (0.73-8.37)</td>
<td>.15</td>
</tr>
<tr>
<td>50,000-74,999</td>
<td>2.27 (0.71-7.27)</td>
<td>.17</td>
</tr>
<tr>
<td>75,000-99,999</td>
<td>3.16 (0.90-11.04)</td>
<td>.07</td>
</tr>
<tr>
<td>100,000-199,999</td>
<td>2.47 (0.72-8.40)</td>
<td>.15</td>
</tr>
<tr>
<td>( \geq 200,000 )</td>
<td>1.81 (0.37-8.93)</td>
<td>.47</td>
</tr>
<tr>
<td><strong>Race and ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>1.05 (0.51-2.15)</td>
<td>.90</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2.61 (1.28-5.33)</td>
<td>.008</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>0.30 (0.06-1.50)</td>
<td>.14</td>
</tr>
<tr>
<td>Non-Hispanic Native Hawaiian or Other Pacific Islander or American Indian or Alaska Native</td>
<td>1.64 (0.17-16.19)</td>
<td>.67</td>
</tr>
<tr>
<td>Non-Hispanic White (reference)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Characteristics</td>
<td>OR (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>Non-Hispanic multiple races</td>
<td>1.21 (0.34-4.22)</td>
<td>.77</td>
</tr>
<tr>
<td><strong>Census region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South (reference)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Midwest</td>
<td>1.42 (0.73-2.75)</td>
<td>.30</td>
</tr>
<tr>
<td>Northeast</td>
<td>0.89 (0.47-1.69)</td>
<td>.72</td>
</tr>
<tr>
<td>West</td>
<td>1.09 (0.57-2.08)</td>
<td>.80</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.03 (0.00-0.19)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>eHealth literacy skills: access dimension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic health information for self (reference: yes)</td>
<td>3.13 (1.69-5.80)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Electronic test results (reference: yes)</td>
<td>1.55 (0.87-2.73)</td>
<td>.13</td>
</tr>
<tr>
<td><strong>eHealth literacy skills: application dimension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic communication with a doctor or doctor’s office (reference: yes)</td>
<td>2.99 (1.67-5.37)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Made appointments electronically (reference: yes)</td>
<td>1.53 (0.91-2.58)</td>
<td>.11</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

**Discussion**

**Overview**

The purpose of this study was to explore the associations between the 2 dimensions, access and application, of eHealth literacy skills and mHealth app use among US adults with obesity. We used the HINTS 2020 data to explore this potential association with a sample of 923 respondents with complete information represented in the data set. We found that the majority of the respondents had health insurance, were female, and were non-Hispanic White, with an average age of 54 years. Also, more than half of the respondents had some level of college or were college graduates. This study highlights the association between eHealth literacy skills for accessing and the application of health information using mHealth apps among people with obesity.

The weighted univariate analyses demonstrated associations between all of the covariates and mHealth app use except employment status and census region. Socioeconomic factors of education and income have been found to be important in the general use of content within digital environments (ie, internet) [44]. However, more specifically to this study, these factors are important in showing the potential relationship they have with mHealth app use among people with obesity. When considering weight management or physical activity interventions using mHealth apps, future interventions should attempt to improve the eHealth literacy of participants by targeting segments of people with obesity identified to be more at risk, such as older individuals with obesity or those in lower income brackets. These initial metrics can be collected through a variety of eHealth literacy assessment tools. The eHealth Literacy Scale, for example, has been studied in diverse languages and populations, and it was designed to convey an estimate of people’s eHealth-related skills. Other instruments, such as the eHealth Literacy Questionnaire, were established to support researchers, designers, and the government in evaluating, developing, and applying effective digital health interventions [54]. Previous research identified that patients with adequate eHealth literacy had more ability to seek health information on the internet and find reliable and high-quality information than patients with inadequate eHealth literacy [55,56].

Accessing health information requires active information-seeking skills. Additionally, context and behaviors to gain information are intertwined in this process. Respondents’ access to web-based health information (seeking health information for themselves) can be informed by the Lenz search behavior stage. The respondents’ access to information requires a search for information from impersonally related sources. There is no indication of the familiarity that respondents have with these resources based on the survey questions. People with obesity who use mHealth devices may exhibit multiple factors in their search for health information, and recent eHealth literacy work supports the nuances involved in seeking health information [57,58].

The information acquired through the information-seeking process impacts an information seeker’s decision-making process. Electronic communication with the doctor’s office can be the result of gathering enough information to move forward based on their original goal or a stop in the information acquisition process that prompts information seeking through a personal connection through digital communication. The results from this work are not intended to model these variables but demonstrate that Lenz’s model, enhanced with recent theories, may help inform studies aimed at understanding active information seeking at the intersection of digital health devices such as mHealth and eHealth literacy skills. Recent models and theories commonly demonstrate that health information–seeking behavior involves the action of seeking out information, irrespective of how or why it is sought [36].

Consistent with Mahmood et al [59], education and age are important sociodemographic factors associated with mHealth app use among people with obesity. As access to health services...
issues through the use of telehealth technology embedded within mHealth apps, it is imperative that this population group be able to benefit from this type of health service [60]. People with lower education levels and older individuals experience more limited eHealth literacy skills and lower mHealth app use, further widening the digital divide gap [18,24]. Additionally, when we examine other sociodemographic factors, marginalized ethnic groups such as Hispanic populations may have access to mHealth apps but experience digital divide issues [30,61].

Issues such as use and knowledge as they relate to using mHealth apps can also contribute to the digital divide [62]. Additional attention is needed to focus on these vulnerable populations. Interventions that can attempt to address this issue are the development of apps and health promotion campaigns that are designed to be culturally relevant [63]. Within the realm of health promotion and wellness, mHealth mindfulness approaches have been used for African American populations [64]. Moreover, studies have described the importance of mHealth interventions with phone features that are familiar to the target population group [62]. There should be consideration of acceptability and efficacy during the developmental phases to support the use of mHealth apps. When considering efficacy, simpler solutions in app design and use should be evaluated. For the older population, features such as 1-click access to a dashboard within health apps that are appropriately displayed in size may be appropriate. Future mHealth apps should also consider health education–related features to support users with low eHealth literacy skills [30,65]. The recent COVID-19 pandemic highlighted the continued digital divide and the disparity in health care services for those who lack sufficient digital literacy skills [66]. mHealth apps benefit people with various chronic conditions, including obesity. People with obesity are less likely to benefit from these mHealth apps if they have low eHealth literacy skills.

mHealth apps facilitate access to health information that has increasingly migrated to web-based spaces [67]. More importantly, mHealth devices assist individuals with seeking health information and decision-making regarding their health [68]. mHealth apps are also advantageous to improve access to health information for personal health data management [15,69]. Since we found that the eHealth literacy skills dimension of access for people with obesity is associated with higher odds of using mHealth apps for seeking health information for themselves, health services should reconsider how they disseminate health information to reach higher proportions of the population. Inevitably, the accessibility of web-based health information has changed the way people engage in health decision-making [70]. This is also evident from our network analysis results, which demonstrate the interconnectivity among all elements relating to eHealth literacy skills and mHealth app use among people with obesity, resulting in the need for holistic solutions to enhance mHealth app use and access to health information. Lenz’s model primarily focuses on the search process and use of the information; however, future studies should consider the nuanced contextual factors for people with obesity and their use of mHealth-related devices.

Accessing health information through mHealth apps streamlines the application of health information for decision-making. Many people with obesity have additional chronic diseases that can benefit from timely communication with their health care provider [71]. Effective communication is important for reported satisfaction and perceived health management outcomes. Face-to-face communication has been the standard for communication among patients and health care providers. However, there are mixed results on the perceived effectiveness of face-to-face communication versus IT-aided communication such as mHealth devices [72]. Recent studies have found that mHealth apps are viewed as useful by patients for improving communication and the accessibility of health data [73]. Therefore, this constant communication creates the potential for a bidirectional channel of communication among people with obesity and their health care providers. An in-depth content analysis of vaccination apps showed that few apps provide the capability for bidirectional communication among users and health care providers [74]. The challenges of bidirectional communication can be attributed to barriers to data integration. Given the numerous mHealth apps available for download, this creates interoperability challenges for electronic health care record systems [75]. For mHealth apps that are designed to improve physician and consumer communication, transdisciplinary scholarship is necessary to overcome these barriers. More importantly, technical and networking policies must be developed to support and incentivize the ability to improve this type of communication.

This study benefits from the use of a nationally representative sample of noninstitutionalized US adults. This study provides an adjusted analysis of the associations between mHealth app use and eHealth literacy skills among people with obesity. New technologies that require eHealth literacy skills are transforming how we receive health care and access health information, but they also highlight new disparities as they relate to digital health services [30]. However, to address the rise of chronic conditions such as obesity, it is essential to empower patients to engage in their own health management. One promising strategy is using mHealth apps as a complementary tool to manage weight loss and track physical activity [26]. We provide evidence of several significant factors that can be informative when designing inclusive mHealth app-based health intervention studies. Our results also have implications for studies aimed at managing weight loss or tracking the physical activity of people with obesity to assist with mHealth app development and uptake.

Concerning limitations, first, there could be additional confounding variables that are not included in the study, which is limited by the survey design questionnaire. Some of these confounding variables may be related to self-care behaviors or use patterns with mHealth apps [25]. Furthermore, a bias in the survey design includes the assumption that apps are used only on tablets or smartphones, such that only individuals who indicated having a tablet or smartphone were asked within the survey about having or using health or wellness apps. Second, respondents were only asked about access to information within the previous 12 months. There is a possibility that users do not access or seek health information between visits with their doctors on a yearly basis. Nevertheless, many patients with low literacy are often left dissatisfied and unsure of the information shared by their doctor and seek third-party sources
such as web-based health communities to fill those gaps [76,77].
Also, respondents who report ownership of a tablet, smartphone,
or both may also use a computer, but they did not indicate that
as part of their response. Third, results from the Ising model
visualization show a strong relationship between respondents
seeking health information for themselves and mHealth app
use. As a result of seeking health information, this may also
explain the strong relationship between mHealth app use and
talking with one’s doctor. Therefore, these correlations may
exist because they are measuring the same events. Also, some
sociodemographic variables used in this study are correlated
(eg, age and retired status), so some multicollinearity may be
present. Fourth, with a small sample of uninsured people
represented in our sample, the statistical significance for health
insurance in our model may have been different with a larger
sample of uninsured people. Additionally, the data did not
provide a distinction between private and public health
insurance, though the information content of such a factor may
already be embedded in the income variable. A study aimed at
self-monitoring of diet, physical activity, and weight among
patients who were underinsured or uninsured demonstrated
higher adherence through the use of 2 mHealth-related apps in
comparison to a paper group [6]. Also, the sample in this study
covers the COVID-19 peak period in early 2020, which may
have represented a crest (and potentially a permanent shift in
behaviors) in electronic access to health information among
people with obesity compared to previous time periods. Lastly,
there is a limitation in the HINTS survey questions as they were
not designed using a web-based health information–seeking
behavior framework, though we were able to detect relevant
associations even with this design limitation.

Future directions of this work should consider this model
structure for people without obesity. A comparative analysis
may identify whether eHealth literacy relevance differs between
individuals with obesity and those without obesity. This work
also considered mHealth app use, but it did not examine
cognitive motivational factors for mHealth use. The
identification of motivational barriers and facilitators can be
analyzed within the context of psychological motivation
frameworks to identify potential intervention targets to leverage
in mHealth intervention–based studies. Also, since the
COVID-19 pandemic may have brought behavioral changes in
the overall population regarding eHealth literacy, a dynamic
study that explores those changes over time could highlight
whether segments of people with obesity may now experience
heightened needs compared to prepandemic stages.

Conclusion
This study estimated the associations between mHealth app use
and eHealth literacy skills. Our findings are consistent with
previous literature, showing that eHealth literacy skills are
associated with accessing digital health information and the
application of digital health services. For example, age is
negatively associated with mHealth app use among people with
obesity, with other sociodemographic factors also showing
strong associations. This highlights substantial uneven access
to eHealth information among people with obesity, potentially
leading to disparities in health outcomes among
sociodemographic groups. It is imperative that this phenomenon
be further investigated as digital health–related services that
involve the use of mHealth apps become more integrated into
health care services and aim to reach wider segments of the
population. A continued challenge is to engage people with
chronic conditions such as obesity to use mHealth apps,
especially older individuals with obesity and those with lower
educational backgrounds. Our work provides evidence of factors
associated with mHealth app use in relation to access and
application. This work provides an initial understanding of
mHealth app use and eHealth literacy skills among people with
obesity, and future studies should identify equitable solutions
for people with obesity (as well as other groups) and their use
of mHealth apps.

Acknowledgments
This project was funded by the National Institute of Health National’s Heart, Lung, and Blood Institute (grant R25HL126145). Dr Thorpe was funded by U54MD000214.

Authors' Contributions
The review concept was designed by GSJ with support from KN and RJTJ. The literature search, screening, analysis, data
collection, and data cleaning were conducted by GSJ with support from BDC. LHG provided the data analysis. GSJ, LHG, and
BDC wrote the original draft, with review and edits by GSJ, LHG, BDC, RJTJ, and KN. The figures were designed by LHG. Additional review was carried out by GSJ, LHG, RJTJ, and KN.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary figures and table.
[DOCX File, 212 KB - mhealth_v12i1e46656_app1.docx ]

References
https://mhealth.jmir.org/2024/11/e46656

JMIR Mhealth Uhealth 2024 | vol. 12 | e46656 | p.421

(page number not for citation purposes)
20. Apps can help keep older folks healthy—but most don't use them. Michigan Medicine-University of Michigan. URL: https://medicine.umich.edu/dept/dgpm/news/archive/202023/apps-can-help-keep-older-folks-healthy-most-dont-use-them [accessed 2023-10-27]


Abbreviations

AUC: area under the curve
HINTS: Health Information National Trends Survey
IRB: institutional review board
mHealth: mobile health
OR: odds ratio
ROC: receiver operating characteristic

©George Shaw Jr, Bianca A Castro, Laura H Gunn, Keith Norris, Roland J Thorpe Jr Jr. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Barriers and Implications of 5G Technology Adoption for Hospitals in Western China: Integrated Interpretive Structural Modeling and Decision-Making Trial and Evaluation Laboratory Analysis

Linyun Zhou¹,², MPH; Minghuan Jiang³, PhD; Ran Duan⁴, PhD; Feng Zuo⁵, MSc; Zongfang Li²,⁶, PhD; Songhua Xu², PhD

1Department of Health Management, The Second Affiliated Hospital of Xi’an Jiaotong University, Xi’an, China
2Institute of Medical Artificial Intelligence, The Second Affiliated Hospital of Xi’an Jiaotong University, Xi’an, China
3Department of Pharmacy Administration and Clinical Pharmacy, School of Pharmacy, Xi’an Jiaotong University, Xi’an, China
4Information Department, Chongqing General Hospital, Chongqing, China
5Information Center, The First Affiliated Hospital of Army Medical University (Southwest Hospital), Chongqing, China
6Clinical Research Center for Hepatic & Splenic Diseases of Shaanxi Province, The Second Affiliated Hospital of Xi’an Jiaotong University, Xi’an, China

Corresponding Author:
Songhua Xu, PhD
Institute of Medical Artificial Intelligence
The Second Affiliated Hospital of Xi’an Jiaotong University
No 5 Jianqiang Road
Xi’an, 710016
China
Phone: 86 029 86320798
Email: songhuaxu@126.com

Abstract

Background: 5G technology is gaining traction in Chinese hospitals for its potential to enhance patient care and internal management. However, various barriers hinder its implementation in clinical settings, and studies on their relevance and importance are scarce.

Objective: This study aimed to identify critical barriers hampering the effective implementation of 5G in hospitals in Western China, to identify interaction relationships and priorities of the above-identified barriers, and to assess the intensity of the relationships and cause-and-effect relations between the adoption barriers.

Methods: This paper uses the Delphi expert consultation method to determine key barriers to 5G adoption in Western China hospitals, the interpretive structural modeling to uncover interaction relationships and priorities, and the decision-making trial and evaluation laboratory method to reveal cause-and-effect relationships and their intensity levels.

Results: In total, 14 barriers were determined by literature review and the Delphi method. Among these, “lack of policies on ethics, rights, and responsibilities in core health care scenarios” emerged as the fundamental influencing factor in the entire system, as it was the only factor at the bottom level of the interpretive structural model. Overall, 8 barriers were classified as the “cause group,” and 6 as the “effect group” by the decision-making trial and evaluation laboratory method. “High expense” and “organizational barriers within hospitals” were determined as the most significant driving barrier (the highest R—C value of 1.361) and the most critical barrier (the highest R+C value of 4.317), respectively.

Conclusions: Promoting the integration of 5G in hospitals in Western China faces multiple complex and interrelated barriers. The study provides valuable quantitative evidence and a comprehensive approach for regulatory authorities, hospitals, and telecom operators, helping them develop strategic pathways for promoting widespread 5G adoption in health care. It is suggested that the stakeholders cooperate to explore and solve the problems in the 5G medical care era, aiming to achieve the coverage of 5G medical care across the country. To our best knowledge, this study is the first academic exploration systematically analyzing factors resisting 5G integration in Chinese hospitals, and it may give subsequent researchers a solid foundation for further studying the application and development of 5G in health care.

(JMIR Mhealth Uhealth 2024;12:e48842) doi:10.2196/48842

https://mhealth.jmir.org/2024/12/e48842
KEYWORDS
5G health care; 5G adoption barriers; 5G adoption strategy; smart health care; Western China hospitals

Introduction

Background
With the advancement of information and communication technology, along with the gradual improvement of China's medical information system construction, China's medical industry is moving away from 1.0 medical informatization to 3.0 medical intelligence [1]. Leveraging the advantages such as ultralow latency, high capacity, high speed, seamless connectivity, high reliability, and low power consumption [2], 5G technology plays an essential role in realizing the interconnection and remote monitoring of medical equipment, patient monitoring, remote consultation, and other telemedicine scenarios. At the same time, 5G technology accelerates the data collection, circulation, analysis, and feedback of various applications in the broad medical and health field. With the advent of 5G technology, medical information can now flow and be shared seamlessly among patients, medical equipment, and hospital information systems. This has paved the way for hospitals to simplify the entire medical treatment and service process, right from prediagnosis to diagnosis and postdiagnosis stages [3].

In recent years, the Chinese government has invested considerably in developing innovative 5G smart hospitals to offer better health care to patients and improve their internal management. For example, in July 2021, a total of 10 departments including the Ministry of Industry and Information Technology, Office of the Central Cyberspace Affairs Commission, National Development and Reform Commission, Ministry of Education, Ministry of Finance, Ministry of Housing and Urban-Rural Development, Ministry of Culture and Tourism, National Health Commission, State-Owned Assets Supervision and Administration Commission of the State Council, and National Energy Administration, jointly released the Sailing Action Plan for 5G Applications (2021-2023). The plan aims to encourage the development of various 5G medical products such as robots, emergency vehicles, medical access gateways, and intelligent medical equipment across the country. The plan also emphasizes the need to strengthen the deployment of 5G medical and health network infrastructure, focusing on improving the coverage of 5G in top-tier national hospitals, disease prevention and control centers, elderly care institutions, and other critical locations. Additionally, the plan aims to build 5G networks and 5G medical edge clouds to enhance in-hospital medical care and telemedicine [4]. In September 2021, the Ministry of Industry and Information Technology, in collaboration with the National Health Commission, released the “Notice on Announcing Pilot Projects for 5G+Medical and Healthcare Applications.” This announcement identified 988 pilot projects aimed at advancing the application of 5G technology in various health care domains, including first aid, telediagnosis, tele-education, tele-intensive care, traditional Chinese medicine diagnosis and treatment, hospital management, intelligent disease control, health management, and other directions. Among the pilot projects, as many as 611 5G smart medical projects are led by hospitals (general hospitals, emergency centers, and specialized hospitals) [5].

More and more hospitals in China are investing in 5G construction. For instance, Guangdong Provincial People's Hospital put into use the 5G hospital in July 2021, aimed to integrate 5G, big data, artificial intelligence, and other new technologies into various medical scenarios such as treatment, teaching, research, management, and service [6]. Shanghai aims to realize 100% 5G in-depth coverage and 5G typical services for all tertiary hospitals and at least 50% 5G in-depth coverage and 5G typical services for other hospitals by 2023 [7]. Sir Run Run Shaw Hospital Affiliated with Zhejiang University School of Medicine successfully performed a cholecystectomy for a patient from Xinjiang Corps Alar Hospital thousands of miles away by leveraging the robotic arm, which achieved a breakthrough in China's 5G ultraromote robot human liver and gallbladder surgery [8]. The Second Affiliated Hospital of Xi'an Jiaotong University has piloted several scenarios, such as 5G+emergency rescue, 5G+mobile computed tomography, 5G+unmanned aerial vehicle medical delivery, and 5G+integrated remote diagnosis.

However, though 5G technology undoubtedly introduces enormous benefits for hospitals if adequately applied, it has yet to be widely used in many health care scenarios. Hospitals are experiencing various challenges during the actual 5G application process. Different problems are met in the implementation process, including expertise, operation, resource, regulation, and market access factors, as described in the innovation resistance theory (IRT) [9]. At the same time, there are still no systematic studies that have explored the barriers to the adoption of 5G applications in hospitals in Western China. This is particularly important given that technological development in the Eastern region of China is more advanced compared to the Western region. The lag in 5G development in Western China may become another factor that increases the economic imbalance between these 2 regions [10]. Hence, more research is essential for Western China to provide a better understanding of the barriers hindering the adoption of 5G in health care.

Objectives
This study addresses the critical research question below: what are the barriers to implementing 5G in hospitals in Western China? Based on the question, the following research objectives have been formulated: (1) to identify critical barriers hampering the effective implementation of 5G in hospitals in Western China, (2) to identify interaction relationships and priorities of the above-identified barriers, and (3) to assess the intensity of the relationships and cause-and-effect relations between the adoption barriers.

Methods

Ethical Considerations
The data were collected through literature review and anonymous questionnaires, which posed no harm to individuals.
and did not involve sensitive personal information or commercial interests. Based on the Regulation for Ethical Review of Life Sciences and Medical Research involving human beings issued by the Chinese National Health Commission, Ministry of Education, Ministry of Science and Technology, and State Administration of Traditional Chinese Medicine (Chapter III, Article 32) [11], ethical review was exempted for this study.

Study Design
To address objectives 1-3, a 3-stage hybrid research methodology was proposed by the authors. As shown in Figure 1, a flowchart of the research procedure is conducted. In the first stage, barriers are identified from the existing literature and discussed with experts for further modification and addition using the Delphi technique. The second stage incorporated the interpretive structural modeling (ISM) to obtain a hierarchical structure and interrelationship between the barriers. ISM has a significant advantage because it displays conclusions in the form of a hierarchical topology diagram that is highly intuitive. The hierarchical topology diagram clearly explains the causal relationship and ladder structure among system factors. However, more is needed to determine the intensity of the relationship between factors. It needs to provide the cause-and-effect relation among barriers, which limits the ISM approach [12]. The decision-making trial and evaluation laboratory (DEMATEL) method, on the other hand, can precisely overcome the limitations of the ISM approach. It can determine the strength of influence between variables within the identified structure, providing a deeper understanding of the causal relationship between influencing factors [13]. Therefore, this paper intends to combine Delphi expert consultation, ISM, and DEMATEL to study the hierarchical structure of driving factors and the causal relationship between them. The procedure is explained in more detail in the following section.

Figure 1. Flowchart of research methodology. DEMATEL: decision-making trial and evaluation laboratory; ISM: interpretive structural modeling; MICMAC: Matriced Impacts Corises-multiplication Appliance Classement.

Stage I: Determination of the Barriers Using Literature Review and Delphi Expert Consultation

Overview
This study conducted a literature review from November 20 to 30, 2022, to gather the resistance factors toward 5G development in health care. After that, a 2-round Delphi expert consultation was implemented to refine the factors identified from the above literature review on December 30, 2022, and February 9, 2023. The Delphi technique was developed by the Rand Corporation in 1953 and used as a multistage self-completed questionnaire with individual feedback [14]. It was initially developed as a method for forecasting but has since been widely applied in other areas, including health research [15,16].

Step 1: Literature Review
Web of Science, PubMed, Google Scholar, Chinese government’s official websites, and gray literature, including industry reports, were searched by referencing keywords including “5G healthcare,” “5G smart healthcare,” “5G in hospitals,” “5G applications in healthcare,” and “digital health in China.”

Step 2: Selection of the Delphi Panel
To ensure the authoritative scoring results of the consulting expert group on the evaluation indicators, experienced professionals with intermediate or senior titles who are willing to cooperate actively and who are interested in this research were selected. Leaders responsible for informatization work in health authority, heads and frontline employees from the
hospital’s information management department, 5G communication suppliers, and scholars in 5G and hospital informatization field were invited to this panel.

**Step 3: Development of Questionnaire**

To determine the importance of the barriers selected from the literature review, consolidated criteria have been designed in questionnaire format following a 5-point Likert-type scale (5=very important, 4=relatively important, 3=intermediate, 2=unimportant, and 1=very unimportant). To determine the degree of expert authority (Cr), the expert’s familiarity with the indicator (Cs) and the judgment basis (Ca) were collected in the questionnaire. The quantitative values for Cs are divided into 5 levels (0.9=very familiar, 0.7=relatively familiar, 0.5=intermediate, 0.3=unfamiliar, and 0.1=very unfamiliar), and the quantitative values for Ca are present in Table 1. The questionnaire is also designed to allow the experts to offer their judgments, with space provided for them to add, remove, and justify their responses.

<table>
<thead>
<tr>
<th>Judgment basis (Ca)</th>
<th>Quantitative value of influence degree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Theoretical analysis</td>
<td>0.3</td>
</tr>
<tr>
<td>Practical experience</td>
<td>0.5</td>
</tr>
<tr>
<td>Learn from domestic and foreign peers</td>
<td>0.1</td>
</tr>
<tr>
<td>Intuition</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Step 4: Data Collection**

The questionnaires were distributed and collected via the WeChat platform, the most widely and frequently used mobile social media in China, which is profoundly integrated into the daily life of Chinese people [17]. It is often used for distributing and collecting questionnaires.

**Step 5: Data Analysis**

An analytical stage followed each round of the Delphi questionnaires. The questionnaire recall rate expresses the degree of positivity of the experts. The degree of expert authority Cr can be calculated from the values of Cs and Ca as follows: 

\[ Cr = \frac{(Cs + Ca)}{2} \]

The degree of coordination of expert opinions is judged by the coefficients of variation (CVs) and Kendall coefficient of concordance (W). In this study, the barrier screening standard is CV \( \leq 0.250 \). Barriers whose CVs are higher than 0.250 will be modified or deleted. CV is calculated by the mean value and SD. For Kendall coefficient of concordance, the larger the value, the better the coordination of expert opinions. After the analysis, experts’ feedback and perspectives will be presented to all participants.

**Stage II: Development of Research Framework Using ISM**

The ISM method originated from structural modeling and was introduced by Warfield [18] for better decision-making when too many factors or constructs exist. It is a qualitative and interpretive method that involves a mutual learning process that uses the experience of experts to identify the relationship between factors, variables, enablers, and barriers [19,20]. Based on the relationship, an overall multilevel structure is extracted from the complex items. It is very suitable for interdisciplinary research of natural science and social science. The ISM method has been widely used in management and new technology research in different industries.

Referring to the above studies, the basic steps of the ISM method in this study are as follows.

**Step 6 involved performing the Matriced Impacts Corises-multiplication Appliance Classement (MICMAC) analysis. The driving power (DP) and dependence power (DEP) of the identified barriers based on the FRM were calculated, and the barriers were classified into 4 clusters, known as an autonomous cluster, dependent cluster, linkage cluster, and independent cluster. The details of these 4 clusters are the following:

- Autonomous clusters: the barriers within the autonomous cluster have low DEP and DP. These barriers have no direct
relation with other barriers and can be considered almost isolated from the system.

- Dependent clusters: the barriers in this group do not have robust DP, but their DEP is strong.
- Linkage clusters: the barriers in this cluster are categorized by high DP and DEP. These factors are unstable, so making any changes to them will significantly affect other barriers and may influence them.
- Independent clusters: the barriers within this cluster have high DP and low DEP. These barriers affect other barriers but are less affected.

Table 2. Conversion rule for IRM\(^a\).

<table>
<thead>
<tr>
<th>(i,j) in SSIM(^b)</th>
<th>(i,j) in IRM</th>
<th>(j,i) in IRM</th>
</tr>
</thead>
<tbody>
<tr>
<td>L(^c)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>M(^d)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N(^e)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>O(^f)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\)IRM: initial reachability matrix.
\(^b\)SSIM: structural self-interaction matrix.
\(^c\)L indicates that barrier i has an impact on barrier j.
\(^d\)M indicates that barrier j has an impact on barrier i.
\(^e\)N indicates that barriers i and j interact with each other.
\(^f\)O indicates that barriers i and j have no interaction with each other.

Stage III: Identification of Cause-and-Effect Group Using DEMATEL

The DEMATEL approach is a system analysis method based on graph theory and matrix tools. It is used to analyze the cause-effect relationship between factors in complex systems and identify the interaction’s intensity \[19,21\]. The basic steps to carry out DEMATEL analysis are as follows:

Step 1 establishes the average “direct relation matrix.” In this step, experts are invited to evaluate each barrier’s influence on another using an integer scale. The designed scale has 5 levels, including integers from 0 to 4, where 0 means no impact, 1 means slight impact, 2 means moderate impact, 3 means high impact, and 4 means extremely high impact. Accordingly, the direct influence matrix of each expert is obtained. Then, the average direct relation matrix is obtained by summarizing and averaging all feedback expert data. Given that \(k\) is the index of experts from a total of \(p\) experts, \(q\) is the index of the barriers, and \(i\) and \(j\) are the indices for 2 barriers, the decision matrix of each expert is given by \(D(i,j)\), and then, the direct impact matrix \(D\) is given by equation (1).

\[
D = ip/jk
\]

Step 2 normalizes the average direct relation matrix. The row and maximum value methods are used. The elements of each row in the average direct influence matrix are summed, and then, the maximum value is obtained by comparison. Finally, each element in the average direct relation matrix is divided by the maximum value. The calculation process can be expressed as equation (2):

\[
D = S/x \quad (2)
\]

Step 3 calculates the “total influence matrix T” by adding all the direct and indirect effects using equation (3).

\[
T = R + C \quad (3)
\]

Step 4 develops the “cause-and-effect diagram” by adding elements of vector \(R\) (row) and vector \(C\) (column) using equations (4) and (5), where, \(R_i\) is the sum of the row and \(C_j\) is the sum of the column of the “total influence matrix.” \((R_i+C_j)\) is called the degree of centrality, and \((R_i-C_j)\) is called the degree of cause.

The horizontal and vertical coordinates can be established according to the values of the degrees of centrality and cause. Among them, the degree of centrality is taken as the abscissa, and the degree of cause is taken as the ordinate. Meanwhile, the cause group and effect group are divided according to the positive and negative values of the \((R_i-C_j)\). If the value of \((R_i-C_j)\) is greater than 0, it indicates that this factor has a more significant influence on other factors in the system, and it is classified as a causal factor. If the value of \((R_i-C_j)\) is less than 0, it indicates that other barriers influence this factor greater and attribute it to the outcome factor.

Results

Literature Review

The literature review identified 15 factors influencing the adoption of 5G in health care. Based on the IRT, we divided the barriers to adopting innovation into 5 primary dimensions: expertise, operation, resource, regulation, and market access.
Compared with the unified theory of acceptance and use of technology, technology acceptance model or technology acceptance model 2, and theory of reasoned action, IRT mentioned above has been verified as an effective and significant alternative for researchers who aim to uncover resistance factors in the health care context [22]. The details are listed in Table 3.

### Table 3. Barriers influencing adoption of 5G in health care: review of literature.

<table>
<thead>
<tr>
<th>Barriers</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Expertise barrier</strong></td>
<td></td>
</tr>
<tr>
<td>A1 Lack of 5G technical talents</td>
<td>5G experts and 5G equipment operators within hospitals are understaffed [23-27].</td>
</tr>
<tr>
<td>A2 Insufficient informatization level</td>
<td>The level of informatization construction of different hospitals is uneven. Significant gaps in equipment networking capabilities, medical data collection, and information integration make it challenging to implement and replicate 5G solutions [28,29].</td>
</tr>
<tr>
<td>A3 Insufficient security verification</td>
<td>Most of the data in the medical field adopt cross-level and multichannel data collection and analysis methods, including hospital management data and private data such as patient psychology, medicine, and behavior data. The security of the data transmitted through 5G network still needs to be verified [28-35].</td>
</tr>
<tr>
<td><strong>B. Operation barrier</strong></td>
<td></td>
</tr>
<tr>
<td>B1 Organizational barriers within hospitals</td>
<td>Not a lot of people understand what 5G is and how it works, and the willingness of traditional hospitals to upgrade and transform 5G networks is relatively low, considering the fact that mature 5G application is mainly concentrated in peripheral medical scenarios such as outpatient guidance and remote consultation [25,28,32,36,37].</td>
</tr>
<tr>
<td>B2 Communication obstacles among hospitals</td>
<td>Communication obstacles exist among hospitals, especially among the higher- and lower-level hospitals [24,28,29,38,39].</td>
</tr>
<tr>
<td><strong>C. Resource barrier</strong></td>
<td></td>
</tr>
<tr>
<td>C1 High expense</td>
<td>Related equipment and communication costs are high, making it difficult for hospitals to afford [25,30,32,40].</td>
</tr>
<tr>
<td>C2 Huge time cost</td>
<td>Installing appropriate equipment and training relevant personnel demand significant time investment [28,40].</td>
</tr>
<tr>
<td>C3 Lack of well-trained medical and technical personnel</td>
<td>Existing medical care and technical personnel are insufficient for 5G integration in medical scenarios [33].</td>
</tr>
<tr>
<td>C4 Lack of mature compatible equipment and systems</td>
<td>It is difficult for 5G network to integrate with existing equipment and systems [31].</td>
</tr>
<tr>
<td><strong>D. Regulation barrier</strong></td>
<td></td>
</tr>
<tr>
<td>D1 Lack of policies related to 5G smart medical integration</td>
<td>Currently, there is no established policy for the integration of 5G smart medical applications [24,28,34].</td>
</tr>
<tr>
<td>D2 Lack of ethics, rights, and responsibilities policies</td>
<td>Lack of policies on ethical controversies, rights, and responsibilities related to the application of 5G [28,31,33,34].</td>
</tr>
<tr>
<td>D3 Lack of standards for corresponding scenarios</td>
<td>There are many 5G smart medical application scenarios; different scenarios have different requirements for network and technical architecture. At present, there is a lack of 5G application standards corresponding to many medical scenarios [24,29,41-43].</td>
</tr>
<tr>
<td><strong>E. Market access barrier</strong></td>
<td></td>
</tr>
<tr>
<td>E1 Lack of unified 5G product standards and listing standards</td>
<td>Emerging 5G smart medical products (such as wearable intelligent terminal equipment and medical instruments) still need unified and perfect listing standards [28,41].</td>
</tr>
<tr>
<td>E2 Lack of complete 5G smart medical product system</td>
<td>5G private network equipment and terminal equipment that meet the customized services of smart medical care still need to be further improved [28].</td>
</tr>
<tr>
<td>E3 Lack of mature business model</td>
<td>There need to be more mechanisms for cross-field cooperation and mature business models [28].</td>
</tr>
</tbody>
</table>

### Delphi Expert Consultation

The Delphi panel in this study comprises 15 members, including practitioners from the health authority, academia, information management departments of the primary, secondary, and tertiary hospitals, and 5G network operators (see Table 4 for panel composition). A 2-round Delphi expert consultation was conducted to explore the views of different experts on the resistance factors toward 5G development in hospitals. The questionnaire was developed based on the literature review.

In the first and second round, we distributed 15 questionnaires each time. In the first round, all the distributed questionnaires were retrieved, while in the second round, 12 questionnaires were collected. The positive coefficient of experts = number of questionnaires returned/number of questionnaires distributed, which can reflect the degree of concern of experts to this study. Thus, in the first round, the positive coefficient of experts was 100%, while in the second round, this coefficient was reduced to 80%.
In the first and second rounds, we obtained data related to the degree of expert authority as follows:

- Degree of familiarity (Cs): 0.670 in the first round and 0.680 in the second round. It shows that the authority of experts in the 2 rounds of consultation is relatively high, and the opinions given are representative to a certain extent.
- Judgment basis (Ca): 0.930 in the first round and 0.920 in the second round.
- Authority coefficient (Cr): 0.800 for both the first and second rounds.

After the first round of expert consultation, the indicator adjustments are as follows, the selection results can be seen in Table 5:

1. Deleted indicators: original C3 (lack of well-trained medical and technical personnel), original C4 (lack of mature compatible equipment and systems), and original E3 (lack of mature business model).
2. Modified indicators: A1 (lack of personnel familiar with 5G within the hospitals) and D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios).
3. Newly added indicators: B3 (lack of cross-unit resource integration channels) and new C3 (lack of means for hospitals to manage their own 5G networks).

The consultation process of the second round is consistent with the first round. According to the expert's scoring, the CV was calculated, and the W test was carried out. As can be seen in Table 6, none of the CVs for the second round of barriers were higher than 0.250.

In the first and second rounds, we obtained the following data on Kendall coefficient of concordance W test: Kendall coefficient of concordance was 0.195 in the first round and 0.258 in the second round, \(\chi^2_{14}=40.854\) in the first round and \(\chi^2_{13}=40.320\) in the second round, and \(P\) value was <.001 for both first and second rounds, which is statistically significant, indicating that the coordination of expert opinions is good. The experts' opinions tended to be unanimous in the second round of consultation, with no modification and new indicators. The final determined barriers can be seen in Figure 2.

**Table 4. Basic information of Delphi panelist.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Experts, n</th>
<th>Constituent ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication technology</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Computer science and technology</td>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td>Health management</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Work experience (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>10-19</td>
<td>10</td>
<td>67</td>
</tr>
<tr>
<td>20-29</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>Professional title</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Vice senior</td>
<td>11</td>
<td>73</td>
</tr>
<tr>
<td>Senior</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>Job description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information technology operations management</td>
<td>10</td>
<td>67</td>
</tr>
<tr>
<td>Academia</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Administrative management</td>
<td>3</td>
<td>20</td>
</tr>
</tbody>
</table>

https://mhealth.jmir.org/2024/1/e48842

© 2024 JMIR Mhealth and Uhealth
### Table 5. Selection results of the first round of expert consultation.

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>3.867 (1.024)</td>
<td>4.267 (0.573)</td>
<td>3.867 (0.806)</td>
<td>4.333 (0.789)</td>
<td>3.200 (0.748)</td>
<td>4.133 (0.884)</td>
<td>3.533 (0.718)</td>
<td>3.267 (1.062)</td>
<td>3.467 (0.957)</td>
<td>3.667 (0.573)</td>
<td>4.267 (0.806)</td>
<td>4.000 (1.247)</td>
<td>4.133 (0.086)</td>
<td>3.667 (1.347)</td>
</tr>
<tr>
<td>CV&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.265</td>
<td>0.134</td>
<td>0.208</td>
<td>0.182</td>
<td>0.234</td>
<td>0.214</td>
<td>0.203</td>
<td>0.325</td>
<td>0.314</td>
<td>0.247</td>
<td>0.276</td>
<td>0.134</td>
<td>0.158</td>
<td>0.195</td>
</tr>
</tbody>
</table>

Selection criterion (CV≤0.250)

<sup>a</sup>CV: coefficient of variation.
<sup>b</sup>The symbol “✓” indicates that the indicator is deleted.
<sup>c</sup>The symbol “✓” indicates that the indicator is modified.
<sup>d</sup>The symbol “✓” indicates that the indicator is retained.

### Table 6. Selection results of the second round of expert consultation.

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>3.917 (0.640)</td>
<td>4.167 (0.553)</td>
<td>3.917 (0.862)</td>
<td>4.750 (0.433)</td>
<td>3.333 (0.745)</td>
<td>4.250 (0.595)</td>
<td>4.167 (0.799)</td>
<td>3.500 (0.866)</td>
<td>4.167 (0.745)</td>
<td>4.167 (0.808)</td>
<td>3.667 (0.759)</td>
<td>4.083 (0.707)</td>
<td>4.000 (0.759)</td>
<td>4.417 (0.759)</td>
</tr>
<tr>
<td>CV&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.163</td>
<td>0.133</td>
<td>0.220</td>
<td>0.091</td>
<td>0.224</td>
<td>0.140</td>
<td>0.192</td>
<td>0.247</td>
<td>0.224</td>
<td>0.215</td>
<td>0.232</td>
<td>0.186</td>
<td>0.177</td>
<td>0.172</td>
</tr>
</tbody>
</table>

Selection criterion (CV≤0.250)

<sup>a</sup>CV: coefficient of variation.
<sup>b</sup>The symbol “✓” indicates that the indicator is retained.

### Figure 2. Adoption barriers of 5G for hospitals in Western China.
Results of ISM

Establish SSIM

In total, 10 experts from the abovementioned 15 were invited to develop the relationships between different barriers using L, M, N, and O. Due to the nonexistence of a special criterion in the literature of ISM concerning what establishes the majority and as in the methodology of ISM [44], this study adopted a 50% criterion to determine the final relationships between different barriers, such as in a relation, if 5 from 10 experts ticked L, the corresponding column is defined as L. However, to be unbiased, for this kind of relationship, the number of specialists answering by M or N or O should be lesser than or equivalent to 3. Fulfilling both requirements, 47 of 91 cases of relations were included to obtain final results, and the remaining cases were all taken as biased and to be taken for more discussion with experts. To analyze the relations of the remaining barriers, this study proceeded for more discussion with 2 senior experts. To eliminate biases, the selected experts were taken out of the given pool of 10 experts. The 2 senior experts were requested to analyze and discuss the remaining barriers until a consensus was reached. Doing so determined a final relationship between different barriers called SSIM, as shown in Table S1 in Multimedia Appendix 1.

Formation of IRM

Transformation of the SSIM into IRM uses the binary rule shown in Table 2. For example, the entry of A1 and D2 in the SSIM is “O,” which is replaced by “0” for A1 and D2 and “0” for D2 and A1 in the IRM (Table S2 in Multimedia Appendix 1).

Formation of FRM

FRM is formed after checking IRM for transitivity. It was done to represent all indirect connections to maintain the consistency of relationships among the barriers. For example, there is a direct relation between A1 and C2 and C2 and D1, but there is no relation between A1 and D1, as shown in the SSIM. Hence, according to the transitivity rule, there is an indirect relation between A1 and D1 corrected during the formation of FRM. It can be observed in Table S3 in Multimedia Appendix 1, where the relation of A1 and D1 is represented by 1*. All of the asterisk signs represent the indirect relation rectified during the formation of the FRM. The FRM calculates each barrier’s DP and DEP. The DP is the summation of the value of all the row elements, while the DEP is the summation of all the column elements corresponding to the respective barrier.

Level Partition

To have a clearer understanding of the relationship between the barriers, a hierarchical structure of the factors is required. Based on FRM, the reachability set, antecedent set, and intersection set for each barrier were developed. Suppose the reachability and intersection set for a specific barrier are identical. In that case, that barrier is deemed at level 1 and assigned the highest position in the ISM hierarchy. After the first iteration, the barriers constituting level 1 are removed, and the previously mentioned procedure is repeated with the remaining barriers until the levels of all barriers have been determined. The results of the different sets and the level iterations are shown in Table S4 in Multimedia Appendix 1.

Formation of ISM

ISM is formulated based on the partition level of barriers. In the first iteration, A2 (insufficient informatization level), A3 (insufficient security verification), B1 (organizational barriers within the hospitals), B2 (communication obstacles among hospitals), B3 (lack of cross-unit resource integration channels), C2 (huge time cost), C3 (lack of means for hospitals to manage their own 5G networks), and E2 (lack of complete 5G smart medical product system) were placed at the top of the ISM. The second iteration resulted in second-level barriers involving A1 (lack of personnel familiar with 5G within hospitals), C1 (high expense), D1 (lack of policies related to 5G smart medical integration), D3 (lack of standards for corresponding scenarios), and E1 (lack of unified 5G product standards and listing standards) placed below the first level. Similarly, in the third iteration, D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) was placed below the second level. The developed framework or ISM of barrier adoption is shown in Figure 3.

Figure 3. Interpretive structural model.
**MICMAC Analysis**

The MICMAC analysis is performed to identify barriers’ DP and DEP and classify them accordingly. As shown in Figures 3 and 4, A2 (insufficient informatization level), A3 (insufficient security verification), B1 (organizational barriers within hospitals), B3 (lack of cross-unit resource integration channels), and C3 (lack of means for hospitals to manage their own 5G networks) were placed at the top of the ISM and fell under the “dependent” cluster. B2 (communication obstacles among hospitals), C2 (huge time cost), and E2 (lack of complete 5G smart medical product system) were categorized under the “linkage” cluster. The barriers under the linkage cluster were volatile due to high DP and DEP. C1 (high expense), D1 (lack of policies related to 5G smart medical integration), D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios), D3 (lack of standards for corresponding scenarios), and E1 (lack of unified 5G product standards and listing standards) were placed in the independent cluster. Considering these barriers as drivers of other barriers in the system, hospitals should prioritize them in their decision-making processes. In addition, D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) has a relatively high driving force and low dependence force, which reveals that it strongly impacts the whole system as displayed in the ISM.

**Figure 4.** Driving and dependence diagram.

The barriers with an R–C value less than 0 were identified as the effect group, while barriers with an R–C value greater than 0 fell under the cause group. As shown in Figure 5, a total of 8 barriers could be classified in the “cause group,” and 6 as the “effect group,” in which C1 (high expense), E1 (lack of unified 5G product standards and listing standards), D1 (lack of policies related to 5G smart medical integration), and D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) took high priority in the causal group, B1 (organizational barriers within hospitals) and A2 (insufficient informatization level) were the most influenced barriers.

**Results of DEMATEL**

The abovementioned 10 experts in the ISM scoring process were also invited to participate in the data collection for the DEMATEL analysis. Experts were invited to evaluate each barrier’s influence on another using a scale of 0–4. After collecting the direct relation matrix of each expert, the average direct relation matrix (Table S5 in Multimedia Appendix 1) was obtained by summarizing and averaging all feedback expert data. Then, the direct relation matrix was converted into a normalized direct relation matrix (Table S6 in Multimedia Appendix 1) using equation (2). Furthermore, the normalized matrix was converted into a total influence matrix (Table S7 in Multimedia Appendix 1) using equation (3). Finally, the degree of influence was calculated using equations (4) and (5). The cause-effect matrix is shown in Table S8 in Multimedia Appendix 1.

The barriers with an R–C value less than 0 were identified as the effect group, while barriers with an R–C value greater than 0 fell under the cause group. As shown in Figure 5, a total of 8 barriers could be classified in the “cause group,” and 6 as the “effect group,” in which C1 (high expense), E1 (lack of unified 5G product standards and listing standards), D1 (lack of policies related to 5G smart medical integration), and D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) took high priority in the causal group, B1 (organizational barriers within hospitals) and A2 (insufficient informatization level) were the most influenced barriers.
Sensitivity Analysis

The sensitivity analysis was performed to verify the consistency of the calculated value and validate the stability of professional judgment. As shown in Table 7, a different weighting was assigned to one expert’s feedback while keeping equal weightings for the other. Four different total relationship matrices and comparable matrices were created by multiplying each weight assigned to the experts, the average relationship matrices were then computed, and the cause-effect relationships among the different barriers were established. As shown in Table 8, the same rank order for cause-effect barriers for each expert was obtained, accepting minor rank order variation. Based on Figure 6, the plots of all barriers during 4 iterations of sensitivity analysis are the same as the base rank. Therefore, it is clear that there was no major change in barrier rankings. The sensitivity analysis confirms the robustness of obtained results.

Table 7. Results of sensitivity analysis.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>R+C</td>
<td>R–C</td>
<td>R+C</td>
<td>R–C</td>
</tr>
<tr>
<td>A1</td>
<td>1.939</td>
<td>0.678</td>
<td>1.939</td>
</tr>
<tr>
<td>A2</td>
<td>2.862</td>
<td>–1.448</td>
<td>2.862</td>
</tr>
<tr>
<td>A3</td>
<td>3.062</td>
<td>–0.864</td>
<td>3.062</td>
</tr>
<tr>
<td>B2</td>
<td>3.982</td>
<td>–0.510</td>
<td>3.982</td>
</tr>
<tr>
<td>B3</td>
<td>2.792</td>
<td>–1.266</td>
<td>2.792</td>
</tr>
<tr>
<td>C2</td>
<td>3.358</td>
<td>0.287</td>
<td>3.358</td>
</tr>
<tr>
<td>C3</td>
<td>1.780</td>
<td>–0.720</td>
<td>1.780</td>
</tr>
<tr>
<td>D1</td>
<td>3.379</td>
<td>1.171</td>
<td>3.379</td>
</tr>
<tr>
<td>D2</td>
<td>1.962</td>
<td>0.955</td>
<td>1.962</td>
</tr>
<tr>
<td>D3</td>
<td>3.843</td>
<td>0.863</td>
<td>3.843</td>
</tr>
<tr>
<td>E1</td>
<td>3.474</td>
<td>1.373</td>
<td>3.474</td>
</tr>
<tr>
<td>E2</td>
<td>3.701</td>
<td>0.316</td>
<td>3.701</td>
</tr>
</tbody>
</table>
Discussion

Principal Results

As shown in Figures 3 and 5, all the barriers in the bottom and middle levels fell into the cause group, indicating those barriers are the primary reasons why 5G technology cannot be adopted in hospitals. In particular, D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) was the only factor at the bottom level, implying it is the underlying influencing factor of the whole system. It also has a relatively high driving force and low dependence force based on the MICMAC analysis, which is consistent with the analysis results of DEMATEL as it took high priority in the causal group with an R–C value of 0.900 and an R+C value of 1.875. The finding aligns with that of [28], which highlights that medical and health care fields are closely linked to people's safety. Additionally, ethical considerations and definitions of responsibilities and risks are the most significant obstacles to the evolution and development of 5G in the core areas of health care. It is suggested that the government should strengthen policy supervision to ensure the safe and ethical use of 5G in the medical field. Specific measures need to be implemented to supervise and regulate the application of 5G in the medical field. In addition, further research on relevant policies and norms is required. At the same time, clarification should be made in policy documents, laws, and regulations.

Another significant obstacle in promoting 5G medical applications is the massive capital investment required, as stated in C1, with the highest R–C value of 1.361, indicating its most significant driving force over other barriers. This finding reflects the facts stated by CN-HEALTHCARE [45]. The overall cost of 5G medical treatment includes the construction of 5G communication networks by operators, procurement and maintenance costs for 5G medical information systems and related medical equipment, purchase of medical care equipment, and services for users. The application of 5G in hospitals also

Table 8. Ranking obtained after sensitivity analysis.

<table>
<thead>
<tr>
<th>Expert</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Expert 3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Expert 4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Figure 6. Sensitivity analysis plots.
requires installing indoor base stations, software support, and computer room upgrades, all of which incur significant expenses. The high cost of 5G applications limits the financing capabilities of hospitals, especially in the Western China region, where hospitals are generally smaller and have limited funding. As a result, only a few large hospitals with telecom operator support have been able to implement 5G medical care. Meanwhile, smaller hospitals are expected to undertake 5G telemedicine with larger tertiary hospitals. Thus, at this stage, it is recommended that smaller hospitals increase their collaboration with larger hospitals to accelerate the adoption of 5G. At the same time, the government is essential to develop affordable solutions and provide financial support for 5G adoption in health care.

With the highest R+C value of 4.317, B1 (organizational barriers within hospitals) is the most closely related barrier to other factors, indicating that it is the most important barrier to adopting 5G in hospitals in Western China. It also has the lowest R–C value of −2.107, which means it is in the effect group and the most influenced barriers by other factors. These results echo the findings of CN-HEALTHCARE [46], which identified a lack of mature 5G applications in medical scenarios, varying levels of acceptance of new technology, concerns about 5G security, and limitations in human, financial, and material resources as factors influencing attitudes toward 5G adoption among different hospitals. Organizational barriers may be addressed through effective management and leadership, clear communication, and collaboration among different departments.

E1 (lack of unified 5G product standards and listing standards) has the second-highest R–C value of 1.316 and a relatively high R+C value of 3.300. The high R–C value specifies its driving force over the other barriers, while the high R+C value indicates its strong impact on the adoption process. This outcome is akin to the findings of Bruer and Doug [47], who mentioned that 5G standards play a crucial role in hardware infrastructure to software running on top of components, and unified 5G standards help to ensure that a range of devices and equipment can operate in a shared system. Therefore, developing unified 5G product standards and listing standards for corresponding scenarios are essential. It is recommended that the Chinese government, 5G network operators, and hospital administrators accelerate cooperation to establish unified 5G product standards and list standards to jointly promote the large-scale development of 5G in the health care sector.

B2 (communication obstacles among hospitals) is a part of the linkage cluster with the second-higherest R+C value of 3.782, indicating it is one of the crucial factors in the whole system. This finding is consistent with that of Wang et al [39], who found that effective communication is critical for the success of 5G adoption in health care. Communication obstacles can lead to misunderstandings and a lack of trust. It is urgent to break through the communication obstacles among hospitals, especially between the higher- and lower-level hospitals, which need to accelerate top-level design, formulate policy documents, and improve relevant legislation and management mechanisms to promote the opening and sharing of 5G medical data and ensure the deep integration of the 5G medical industry.

**Concluding Remarks**

The paper comprehensively analyzes barriers to 5G adoption in hospitals in Western China. Experts from different stakeholders validated 14 resistance factors. Next, an integrated ISM-DEMATEL approach was applied to model the barriers as a network of factors and alternatives categorized into clusters. All barriers were related to human expertise, resource allocation, operational procedures, laws and regulations, and market access capability. Overall, the study shows that promoting the integration of 5G in hospitals in Western China faces multiple complex and interrelated barriers. It constructs a framework for the main barriers to 5G adoption in the hospital context and provides regulatory authorities, hospital managers, and telecom operators with theoretical and managerial insights into the strategic pathways.

**Theoretical and Managerial Implications**

- The barrier at the bottom level of the ISM should be emphasized for short-term strategy. The middle-level barriers can be considered for medium- and long-term strategies. The barriers at the top of the ISM can be a long-term strategic focus.
- The effect group can easily be influenced by the cause group, and therefore, managers should give the most attention to causal barriers when implementing 5G practices in hospitals.
- The ranking of cause-effect group barriers can assist regulatory authorities, hospital managers, and telecom operators in developing strategic policy during 5G implementation.
- To overcome these barriers, hospital managers should formulate a 5G adoption strategy that considers the specific needs of the institution and the local market. Specific measures include increasing investment in information infrastructure and human resources, establishing supplier communication channels, and promoting cross-unit resource integration.
- To expand the use of 5G in health care scenarios, it is recommended that the government accelerates the construction of an innovation system consisting of regulators, hospitals, telecom operators, academic researchers, and patient representatives.

**Outlook of 5G Health Care**

As the infrastructure of intelligent medical care, 5G allows the transmission of vast amounts of data and information, making the information superhighway a reality. Furthermore, with the full deployment of 5G medical care, especially the integration with big data, artificial intelligence, internet, internet of things, and blockchain technology, 5G is expected to bring significant changes to the current medical and health system and promote the evolution of the entire medical ecology, including hospital operation and management. In the long run, 5G health care promotes the sinking of high-quality medical resources and the development of China's “primary health care” and “family doctor” systems. It can improve the population's overall health, reduce medical expenses for ordinary people, and relieve medical insurance burdens. Therefore, no matter how barriers are faced, the benefits of 5G medical care are expected to
outweigh the costs, making it a worthy investment. In the development of 5G medical care, various technical, economic, institutional, interest, and ethical problems will inevitably be encountered. The regulatory authorities, hospitals, telecom operators, and the public must cooperate to explore and solve the problems in the 5G medical care era, aiming to achieve the coverage of 5G medical care across the country.

Limitations
Although this study suggests some useful implications, there are some limitations that could be considered for future research. First, due to the challenges in reaching out to health care professionals and telecom operators from all regions of China, most of the experts invited for this research were from representative cities in Northwest and Southwest China, including Xi'an, Chongqing, and Chengdu. This may not fully represent the entire country, as Eastern China is generally more developed than Western China. Therefore, there is a lack of balance among the groups of participants in this research. In future research, we plan to invite experts from Eastern China as a complement study. Second, the relations established among barriers might be biased because they are selected and analyzed based on expert opinions that are context-dependent and depend on their organization’s culture and experience. Third, the outcome of this study is valid for the Chinese health care field and cannot be generalized for other sectors without modifications. It can be extended from the Chinese context to a broader coverage by selecting experts from different countries for benchmarking studies. Finally, only 4 groups of experts, namely, government information department staff, managers from the hospital information technology departments, telecom operators, and scholars, are involved in the research process. Other vital stakeholders, such as patients, can also provide crucial information and insights related to the development of 5G health care.

Acknowledgments
The authors acknowledged the support of the Key Research and Development Projects of Shaanxi Province, China (grant 2021SF-188), and the Youth Fund of the Second Affiliated Hospital of Xi'an Jiaotong University, China (grant YJ(QN)202018). The authors also acknowledged the support of Mr Duwei Dai, Ms Caixia Dong, and Ms Chunyan Zhang during the investigation and data analysis process.

Authors’ Contributions
LZ was responsible for the initial concept development, formal analysis, and writing the first draft of the study. MJ contributed to the development of the research methods and edited the writing. RD and FZ were involved in data collection and curation. ZL and SX provided supervision and reviewed the writing. All authors contributed to the final version of the paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Interpretive structural modeling model and decision-making trial and evaluation laboratory method operation process.

References
7. Shanghai: By 2023, 100% of all tertiary hospitals will achieve 5G deep coverage and 5G typical services online. Shanghai Communications Administration. URL: https://www.cn-healthcare.com/articlewm/20220331/content-1332938.html [accessed 2022-11-22]


27. Zhou et al. JMIR Mhealth Uhealth 2024 | vol. 12 | e48842 | p.440 [https://mhealth.jmir.org/2024/1/e48842](https://mhealth.jmir.org/2024/1/e48842)


45. 5G healthcare: pain points, problems, applications and prospects. CN-HEALTHCARE. URL: https://www.sohu.com/a/426797130_139908 [accessed 2023-03-22]

46. Bottom line "5G+medical care" craze: just a need for hospitals or a commercial gimmick? CN-HEALTHCARE. URL: http://tinishurl.com/3v5r72uy [accessed 2023-03-25]


**Abbreviations**

CV: coefficient of variation  
DEMATEL: decision-making trial and evaluation laboratory  
DEP: dependence power  
DP: driving power  
FRM: final reachability matrix  
IRM: initial reachability matrix  
IRT: innovation resistance theory  
ISM: interpretive structural modeling  
MICMAC: Matriced Impacts Corises-multiplication Appiance Classemement  
SSIM: structural self-interaction matrix

©Linyun Zhou, Minghuan Jiang, Ran Duan, Feng Zuo, Zongfang Li, Songhua Xu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.01.2024. This is an open-access article distributed under the terms of the Creative Commons
Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Trial Participants’ Perceptions of the Impact of Ecological Momentary Assessment on Smoking Behaviors: Qualitative Analysis

Elizabeth R Stevens¹, MPH, PhD; Rina Li¹; Grace Xiang¹, BA; Rachel Wisniewski¹, BA; Sidney Rojas¹, BA; Katherine O'Connor¹, MS; Olivia Wilker¹, MPH; Mahathi Vojjala¹², MPH; Omar El-Shahawy¹², MPH, MD, PhD; Scott E Sherman¹, MPH, MD

Corresponding Author:
Elizabeth R Stevens, MPH, PhD

Abstract

Background: Ecological momentary assessment (EMA) is an increasingly used tool for data collection in behavioral research, including smoking cessation studies. As previous addiction research suggests, EMA has the potential to elicit cue reactivity by triggering craving and increasing behavioral awareness. However, there has been limited evaluation of its potential influence on behavior.

Objective: By examining the perspectives of research participants enrolled in a tobacco treatment intervention trial, this qualitative analysis aims to understand the potential impact that EMA use may have had on smoking behaviors that may not have otherwise been captured through other study measures.

Methods: We performed a qualitative analysis of in-depth interviews with participants enrolled in a pilot randomized controlled trial of a tobacco treatment intervention that used SMS text messaging to collect EMA data on smoking behaviors. In the pilot randomized controlled trial, combustible cigarette and e-cigarette use and smoking-related cravings were measured as part of an EMA protocol, in which SMS text messaging served as a smoking diary. SMS text messaging was intended for data collection only and not designed to serve as part of the intervention. After a baseline assessment, participants were asked to record daily nicotine use for 12 weeks by responding to text message prompts that they received 4 times per day. Participants were prompted to share their experiences with the EMA text messaging component of the trial but were not directly asked about the influence of EMA on their behaviors. Transcripts were coded according to the principles of the framework for applied research. The codes were then examined, summarized, and grouped into themes based on the principles of grounded theory.

Results: Interviews were analyzed for 26 participants. The themes developed from the analysis suggested the potential for EMA, in the form of an SMS text messaging smoking diary, to influence participants’ smoking behaviors. The perceived impacts of EMA text messaging on smoking behaviors were polarized; some participants emphasized the positive impacts of text messages on their efforts to reduce smoking, while others stressed the ways that text messaging negatively impacted their smoking reduction efforts. These contrasting experiences were captured by themes reflecting the positive impacts on smoking behaviors, including increased awareness of smoking behaviors and a sense of accountability, and the negative impacts on emotions and smoking behaviors, including provoking a sense of guilt and triggering smoking behaviors.

Conclusions: The collection of EMA smoking behavior data via SMS text messaging may influence the behaviors and perceptions of participants in tobacco treatment interventions. More research is needed to determine the magnitude of impact and mechanisms, to account for the potential effects of EMA. A broader discussion of the unintended effects introduced by EMA use is warranted among the research community.

(JMIR Mhealth Uhealth 2024;12:e52122) doi:10.2196/52122

KEYWORDS

smoking behavior; ecological momentary assessment; bias; behavioral impact; smoking; smoker; qualitative analysis; pilot study; tool; data collection; tobacco; text message; accountability; mHealth; mobile health; message; trigger; cigarette

https://mhealth.jmir.org/2024/1/e52122

JMIR Mhealth Uhealth 2024 | vol. 12 | e52122 | p.443

(page number not for citation purposes)
Introduction

Ecological momentary assessment (EMA) is a data collection method that is increasingly being used in health and behavioral sciences [1,2]. EMA has been shown to be a useful tool for measuring behaviors associated with substance use [3]. Consequently, there has been a strong interest in the use of EMA in smoking cessation studies [4]. However, as indicated in previous addiction research, EMA has the potential to elicit cue reactivity by triggering craving and increasing behavioral awareness [5-8]. As such, there is potential for the use of EMA to create an assessment effect and inadvertently influence behaviors in some settings.

Assessment effects refer to the phenomenon in which the outcome of interest (eg, a behavior) is modified simply by assessing it, arising due to the assessment method or its interaction with the intervention [9,10]. The frequent prompts for self-reporting that are often required by EMA might inadvertently influence participants’ behaviors by, among other mechanisms, increasing their self-awareness, altering an emotional response, or serving as reminders of the behavior [2,8,11,12]. Depending on the behavior of interest, these EMA consequences may differentially impact outcomes. These assessment effects can have a significant impact on the interpretation of trial results but are rarely considered in trial design [13].

Despite its potential importance to the interpretation of research results, there are a limited number of studies investigating the potential for EMA to produce an assessment effect. Within the research that does exist, there have been mixed results reported, with some studies reporting no impact on behavior [11,14,15] and others indicating that EMA likely has an impact [2,5-8,11,12,14]. For this reason, it is important to understand if and how the use of EMA data collection in smoking research could impact smoking behaviors.

To better understand the potential impact of using EMA to measure smoking behaviors, we performed a qualitative analysis of in-depth interviews with participants in a randomized controlled trial (RCT) pilot study of a smoking intervention that used SMS text messaging–based EMA data collection as a data collection strategy [16]. By examining the perspectives of the trial participants, this qualitative analysis aims to understand the potential impact that the use of EMA may have had on smoking behaviors that may not have otherwise been captured through other means of data collection.

Methods

Study Design

For the purposes of assessing the acceptability of a smoking intervention and determining points for potential program improvement, we performed a qualitative analysis of in-depth interviews conducted with participants of an RCT pilot study that compared the effectiveness of behavioral counseling and the use of e-cigarettes on smoking outcomes to that of behavioral counseling and nicotine replacement therapy (NRT).

Ethical Considerations

The interviews and analysis procedures were reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1), as applicable [17]. The study protocol was approved by the New York University Langone Health Institutional Review Board (approval number i20-00839), and written documentation of informed consent was received prior to starting data collection. Participants were provided with a US $20 incentive for their participation. Participant data were maintained on a secure server. After transcription, all participant data were deidentified prior to data analysis.

Setting and Participants

Interview participants were recruited upon completion of the intervention phase of the RCT at the 12-week follow-up study visit. All RCT participants were invited to participate in an in-depth interview to discuss their experiences with the intervention and other aspects of the trial. Interview recruitment ended once thematic saturation was reached. Interviews lasted approximately 30 minutes and were performed between April 2021 and November 2022.

The pilot RCT was performed to determine the feasibility and acceptability of an e-cigarette–based smoking intervention and to compare the effectiveness of counseling and e-cigarette use on smoking outcomes to that of counseling and NRT [16]. Text message–based EMA data collection was used to record smoking patterns. Patients from the electronic health record system of New York University Langone Health—a private hospital system serving New York, New Jersey, and Connecticut—were recruited as RCT participants. The RCT participant sample was initially restricted to patients with a diagnosis of chronic obstructive pulmonary disease, but the scope was later expanded to include patients with a diagnosis of coronary artery disease, peripheral artery disease, or asthma. In addition, to be eligible, RCT participants were required to smoke ≥4 days per week, with at least 5 cigarettes smoked on the days that participants did smoke; be motivated to quit smoking; and possess a phone with SMS text messaging capabilities. A total of 121 participants were recruited into the pilot RCT.

RCT EMA Protocol

In the pilot RCT, combustible cigarette use and e-cigarette use, as well as smoking-related cravings, were measured as part of an EMA protocol, in which SMS text messaging served as a smoking diary. The SMS text messaging was intended for data collection only and not designed to serve as part of the intervention. After a baseline assessment, participants were asked to record their daily nicotine use for 12 weeks by responding to text message prompts that they received 4 times per day. The coverage design prompted participants to provide brief check-in reports via text message over the course of each day, wherein they were asked to report combustible cigarette and e-cigarette use based on their study arm (Figure 1). Responses to 1-item measures of cigarette craving and satisfaction were also collected from each report.
Data Collection

Semistructured, in-depth telephone interviews, which lasted approximately 30 minutes, were conducted by research staff (OW and MV). Interviews were audio-recorded, transcribed, and imported into Dedoose software (SocioCultural Research Consultants LLC) for qualitative data analyses.

The interview guide covered topics that were designed for the following goal: gaining a deeper understanding of the participants’ experiences, intervention satisfaction, attitudes toward e-cigarettes, and intentions to quit. The interviews were intended to assist with the further adaptation of the e-cigarette smoking intervention and behavioral counseling manual used in the RCT. The interview guide was developed based on the pilot RCT procedures to provide additional insights about the barriers and facilitators of e-cigarette use and how to refine the current approach to enhance program retention and outcomes. Interviews covered topics such as program aspects that the participants liked or disliked, features of the intervention that should be modified, participants’ experiences with using e-cigarettes or NRT, intentions of using e-cigarettes after the intervention, and whether participants’ health symptoms interfered with their ability to engage in the intervention. The participants were prompted to share their opinions on the EMA texting component of the trial but were not directly asked about the influence of the EMA on their behaviors. Participants were prompted to discuss the texting in the interview, as follows: “During the program, you answered questions over text on daily
basis—What was that experience like for you?” This was followed by probes, including “How may texting have affected your overall experience with the program?” A copy of the interview script is available in Multimedia Appendix 1.

Data Analysis
Interview transcripts were coded by using procedures that were designed to ensure thoroughness and reliability. We used Dedoose software to manage the data and coded the data according to the principles of the framework for applied research [18], which consists of the following 5-stage process: familiarization, identifying themes, indexing, charting, and interpretation. Codes were primarily developed a priori based on intervention components and the quality improvement goals of this study. Additional codes were developed by reviewing a random sample of interviews and via discussion with the coding team. The general development of themes arose from the data, using the principles of grounded theory [19]. To enhance reliability, 2 researchers took part in the coding and analysis process for each interview. Prior to full coding, a random sample of interviews was double coded, and intercoder reliability was assessed based on percent agreement (range 79.5%-87.9%). Afterward, all disagreements were discussed as a group to improve concordance among coders. All interviews were then independently coded by 5 coders (GX, KO, RL, RW, and SR) who worked in various pair combinations; each pair met with the other coders and a coauthor (ERS) to resolve discrepancies. When coding was completed, the quotations for each code were examined, summarized, and grouped together into themes.

Results
Participants
Interviews were performed with a total of 27 participants. Due to an audio malfunction, 1 interview was not included in the analysis. As such, 14 participants were in the e-cigarette study arm and 12 were in the NRT study arm. The average age of the participants included in the analysis was 57.1 (range 28-74) years; 54% (14/26) of participants identified as female, 46% (12/26) were White, 31% (8/26) were Black, 15% (4/26) were Hispanic, and 4% (1/26) were Asian. The majority (17/26, 65%) of participants had completed at least some college; 35% had a high school education or less. Around two-thirds (17/26, 65%) of participants had a diagnosis of chronic obstructive pulmonary disease, 15% (4/26) of participants were diagnosed with coronary artery disease or peripheral artery disease, and 19% (5/26) of participants were diagnosed with asthma.

Themes
Overview of Themes
Without being directly prompted, 18 of the 26 participants described EMA impacting their behaviors or emotions, and several themes indicating a potential, inadvertent influence of EMA on smoking behaviors emerged. The perceived impacts of EMA texting on smoking behaviors were polarized; some participants emphasized the positive impacts of the text messages on their efforts to reduce smoking, while others stressed the ways in which the texting negatively impacted their smoking reduction efforts. These contrasting experiences were captured by 2 major themes and their subthemes, reflecting the positive impacts on smoking behaviors, including increased awareness of smoking habits and a sense of accountability, and the negative impacts on emotions and smoking behaviors, including provoking a sense of guilt and triggering smoking behaviors (Table 1).

### Table 1. Major themes and subthemes.

<table>
<thead>
<tr>
<th>Themes and subthemes</th>
<th>Frequency, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EMA(^a) texting has a positive impact on smoking behaviors</strong></td>
<td></td>
</tr>
<tr>
<td>EMA texting serves as a source of accountability</td>
<td>16</td>
</tr>
<tr>
<td>Anticipation of the next text message serves as a deterrent to impulsive smoking</td>
<td>13</td>
</tr>
<tr>
<td>Texting prompts increase awareness of smoking habits</td>
<td>6</td>
</tr>
<tr>
<td><strong>Reminders of the goal to reduce cigarette smoking</strong></td>
<td>11</td>
</tr>
<tr>
<td>Check-ins serve as markers of progress made toward quitting</td>
<td>5</td>
</tr>
<tr>
<td><strong>Negative impact of EMA texting on emotions and smoking behaviors</strong></td>
<td>7</td>
</tr>
<tr>
<td>Repeated text messages asking about smoking behaviors produce negative emotions</td>
<td>4</td>
</tr>
<tr>
<td>Text messages inquiring about cigarette use may have a triggering effect</td>
<td>5</td>
</tr>
</tbody>
</table>

\(^a\)EMA: ecological momentary assessment.
**Perceived Positive Impacts of EMA Texting Procedures**

**Overview of Positive Impacts**

A major theme that arose was the perceived positive impacts of the EMA methods on efforts to reduce smoking. Many (n=16) participants perceived the SMS text messaging smoking diary as an important component of the tobacco treatment intervention and their experience during their efforts to reduce their cigarette smoking. Participants described the EMA text messages as helpful because “you could use [them] for yourself as a tool” (Participant E09), and while stating that “the text messages are a pain in the butt,” a participant thought that “they were very useful….A useful pain in the butt” (Participant E04). Within this overarching theme—the positive influence of EMA—four subthemes that highlighted the potential roles of EMA text messages emerged: EMA text messages (1) increase awareness of smoking habits, (2) serve as reminders of smoking goals, (3) mark progress made, and (4) provide a sense of accountability.

**Texting Prompts Increase Participants’ Awareness of Their Smoking Habits**

Many (n=16) participants noted that the EMA text messages helped them with their smoking behaviors because the act of recording the number of cigarettes smoked increased their awareness of their smoking behaviors. This theme—increased smoking habit awareness—encompassed the following three layers: (1) awareness of the number of cigarettes smoked and smoking habits, (2) reminders of the goal to reduce cigarette smoking, and (3) markers of progress made.

The predictable and repeated EMA text message requests for participants to report the number of cigarettes smoked prompted reflection on their smoking habits. By engaging in the conscious effort of quantifying the number of cigarettes smoked, the participants heightened their self-awareness and gained a clearer understanding of the frequency of their cigarette use. One participant remarked:

> [The text message would say] “...have you smoked? When was the last time you smoked your last cigarette?” So, it helped me to be aware of how many cigarettes I was smoking per day. [Participant N24]

With regard to the act of recording cigarette use, a participant noted:

> ...you’re so aware with the...text messages...of just how often you smoke....Because with the text, I was physically writing it and seeing it. [Participant E27]

**Reminders of the Goal to Reduce Smoking**

For some participants (n=11), the EMA text messages also served as reminders of their goal to reduce or quit smoking and heightened their sense of purpose and determination. One participant said, “It [the text messaging] does remind you what you’re supposed to be focused on” (Participant N20), and in this way, the messages served as frequent reminders of participants’ intentions to change their smoking behaviors. The messages also invigorated their commitment to achieving these goals; one participant said:

> It [the text messages] kept me going. It kept me wanting to quit, you know, and to keep doing it, to keep at the program. [Participant N36]

The text messages also kept participants feeling involved:

> ...[questions] like, “how many cigarettes have you smoked?”...kind of keep you involved in it instead of letting it go on the way side. Like it would kind of keep reminding you that...this is what you’re working on, you know? [Participant N50]

**Markers of Progress Made Toward Smoking Reduction**

The EMA text messages served as markers of progress made toward reducing smoking (n=5). By regularly quantifying and reporting their smoking behaviors, participants saw tangible evidence and took note of the accomplishments resulting from their efforts. By tracking progress over time, participants observed patterns of improvement, which reinforced their motivation to continue cutting back on smoking. When describing why they found the text messages useful, one participant said:

> ...it’s like a progress type thing. So, I enjoyed the texts...It kind of gave me a reminder [of my progress] because as I went on, I had less and less craving. [Participant N58]

**EMA Texting Serves as a Source of Accountability for Progress Toward Reducing Cigarette Smoking**

Many participants (n=13) believed that the EMA text messages had a positive effect on their sense of accountability in their efforts to reduce smoking. Knowing that they would receive inquiries about their smoking habits increased participants’ mindfulness in reducing their smoking. One participant stated:

> I loved it [the text messaging] because it kept me...accountable...because it kept asking the questions over and over again. In the beginning, it was like a little stressing because I was like, “Oh, my God, these messages, I don’t want to deal with it.” But, it kept me accountable. And it was good. [Participant N36]

Similarly, another participant remarked:

> I had to answer for all the cigarettes I smoked today and hold myself accountable. I thought that was ingenious. [Participant N06]

In some instances (n=6), the text messages served as deterrents to impulsive smoking. One participant shared:

> [There were a couple of times when they [cigarettes] were right on hand...I’m upstairs, I’m getting ready to light and then boom [sic], oh, “did you smoke today? How many times did you smoke since we last communicated?” [Participant N31]

Similarly, the anticipation of upcoming EMA messages served as a motivator for refraining from smoking or delaying the next cigarette. The participant went on to describe their morning routine:

> ...while you’re trying to wake up and organize yourself, you pick up the cigarette...[but] you know
you’re going to get a text at 9 o’clock saying, “did you smoke?” right? So that notion pops in your head, so you don’t smoke right away. [Participant N31]

Having the knowledge that there was a regular time for reporting their smoking activities encouraged participants to make more deliberate choices regarding their smoking behaviors. One participant discussed how they used EMA check-ins to consciously reduce their cigarette consumption during the time leading up to the EMA prompt:

I knew at a certain time, I was going to get this text. So, when I went to the text, I wanted to have everything in line. I wanted it to be right. So therefore, I would only smoke four cigarettes because that was the allotted cigarettes that I was supposed to smoke at the time. During that time, I would only smoke three cigarettes because I had cut down to three during that period of time and I looked forward to doing it. [Participant N04]

EMA Text Messages Can Negatively Impact Participants’ Emotions and Smoking Behaviors

Overview of Negative Impacts

A second major theme that emerged from the data was the potential negative effects of EMA texting could have on participants. Although many participants reported experiencing positive effects of EMA on their smoking behaviors, some participants (n=7) reflected on the potential negative impacts of the EMA text messages. This theme—the negative effects of EMA—was further distilled into the following two subthemes: the potential roles of EMA in (1) producing negative emotions and (2) triggering cigarette cravings.

Repeated Text Messages Asking About Smoking Behaviors Produce Negative Emotions

Some participants (n=4) mentioned feeling “bad” when they had to admit to smoking or experiencing relapses via the text messages. However, some respondents admitted that the guilt experienced as a result of the text messages, while being a negative emotion, reinforced their determination and prompted them to make renewed efforts, with one participant stating that the text messages “helped me because when I wrote how…I smoked a cigarette…I kind of felt bad” (Participant N58). Moreover, other participants emphasized the negative emotions and guilt experienced when a spotlight was focused on their perceived failures. One participant described the experience by saying, “being able to tell about my progress, or lack thereof, I would have felt bad if I had a slip up” (Participant N21).

Text Messages Inquiring About Cigarette Use May Have a Triggering Effect

Several participants (n=5) expressed their desire to remove the texting component of the program due to its triggering effect on their smoking. Some participants reported that the text messages acted as triggers for cigarette use, as the text messages reminded them of smoking, thereby eliciting an urge to smoke. One participant said that the EMA “was a reminder, actually,” and “…with the reminder came in the struggles” (Participant N20). Notably, another participant said:

[the text messages] happened so often, and you knew they were coming, and they started to almost act like a trigger because you sometimes weren’t even thinking about [smoking], but then they would ask you about cigarettes and suddenly you’re thinking about it...they were more triggers to smoke than to prevent smoking. [Participant N31]

Discussion

Principal Findings

This study indicates that smoking intervention participants perceive the collection of EMA smoking behavior data via SMS text messaging as a potential influence on their smoking behaviors. The themes developed from the analysis revealed that EMA, in the form of an SMS text messaging smoking diary, may be perceived as a source of accountability for smoking reduction but may also be a trigger for cigarette use among some people. The results of this study emphasize the need to examine the potential influence of EMA data collection techniques on participants’ behaviors within smoking interventions, as well as in other behavioral research.

The perceived impacts of EMA on smoking behaviors are consistent with previously made observations that the act of receiving EMA prompts can increase behavioral awareness and act as a trigger for craving [6-8], in addition to altering participants’ moods [2,12]. This suggests that EMA for data collection purposes has the potential to unintentionally create an assessment and intervention effects in itself. Although a lack of EMA impact on behavior has been reported in some studies [11,14,15], these studies may be limited by the choice of measures used. As seen in suicide research, EMA prompts have been observed to have an effect on some measures, such as mood, but a minimal effect on other measures, such as suicidal ideation [2]. Therefore, when designing a study, it is important for researchers to reflect on the various factors that may influence the behaviors of interest and consider the potential effect that EMA may have on these factors, in addition to the primary outcomes of interest. Due to its potential effect on participants, investigators should consider and discuss the potential for a behavioral influence to be introduced into a study through the use of EMA data collection.

There is a need for further investigation into the ways that and the degree to which EMA affects participants. Within the EMA literature, there is a general lack of discussion around the effects of EMA on participants’ behaviors. When seeking to improve EMA methods, focus is often placed on participant retention and the validity of the data collection method [20-23], with little to no discussion on the potential behavioral impact of EMA. Indeed, when discussing strengths and limitations of EMA, a large portion of the EMA study literature discusses and reports measures of adherence to and reliability of EMA data collection [20-23], with few studies exploring the potential limitation of EMA in which the data collection itself may affect behaviors of interest [2,5-8,11,12,14,15]. A potential effect from EMA may influence the interpretation of the results; therefore, investigators ought to be encouraged to report considerations related to EMA when designing and publishing a study. Future
research may benefit from randomizing a subset of participants to receive one EMA modality (eg, texting) while observing behaviors among all participants with another measurement modality (eg, Bluetooth e-cigarette monitor).

When the potential for EMA to influence study outcomes is identified, less obtrusive EMA methods could be considered, when available. The participants in this study expressed the omnipresent awareness and anticipation of the SMS text messaging–based EMA. This awareness altered participants’ behaviors and resulted in negative emotions that likely would not have been emphasized had the EMA not been used or had been subtler. There are numerous types of EMA strategies used in smoking research [4]. Future intervention research studies could consider less frequent SMS text messaging or EMA data collection methods outside of SMS text messaging that may have a more minor impact on smoking behaviors, such as the use of biosensors [24], Bluetooth-enabled devices [25], or puff counters [26].

This study had a few limitations. First, the interview guide was not designed to investigate the impact of EMA on participants’ smoking behaviors. Therefore, further details on the effects of EMA were not deeply explored, limiting the scope of this analysis. The unprompted nature of the participants’ observations of behavioral impact, however, strengthens the conclusion that the SMS text messaging–based EMA had a meaningful impact on the trial participants. Second, the sample of interview participants was not randomized, and the interview was not required; rather, it was offered to all participants sequentially as an optional component. This potentially introduced selection bias, as those with stronger opinions on the program may have been more likely to participate. Third, as EMA data were collected as part of a smoking reduction trial, it is difficult to completely disentangle the effects of the intervention on changes in behavior from the effects of EMA. Finally, the impact of EMA on behavior change was based on the self-reported perceptions of interview participants, and behavior changes were not directly observed. Therefore, this study cannot be used as conclusive evidence that the EMA had a significant impact on smoking behaviors, and further research is needed.

**Conclusion**

The collection of EMA smoking behavior data via SMS text messaging may influence the behaviors and perceptions of participants in smoking interventions. More research is needed to determine the magnitude of impact and mechanisms, to account for the potential effects of EMA on behavior change. Furthermore, a broader discussion of the behavioral influence introduced by the use of EMA may be warranted among the EMA research community.

**Acknowledgments**

Funding was received from internal funds provided by NYU Grossman School of Medicine (principal investigator: SES).

**Data Availability**

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

**Authors’ Contributions**

ERS conceived the study, contributed to methods development, participated in data collection and analysis, and led manuscript writing. RL contributed to methods development, participated in data analysis, and contributed to manuscript writing. GX and RW contributed to data collection and analysis and provided manuscript revisions. SR and KO contributed to data analysis and provided manuscript revisions. OW contributed to methods development and data collection and provided manuscript revisions. MV contributed to data collection and provided manuscript revisions. OES contributed to methods development and provided manuscript revisions. SES obtained funding, contributed to methods development, and provided manuscript revisions. All authors reviewed and approved the final manuscript.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Study in-depth interview guide.

[DOCX File, 28 KB - mhealth_v12i1e52122_app1.docx]

Checklist 1

COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.

[DOCX File, 17 KB - mhealth_v12i1e52122_app2.docx]

**References**


Abbreviations

- COREQ: Consolidated Criteria for Reporting Qualitative Research
- EMA: ecological momentary assessment
- NRT: nicotine replacement therapy
- RCT: randomized controlled trial

© Elizabeth R Stevens, Rina Li, Grace Xiang, Rachel Wisniewski, Sidney Rojas, Katherine O’Connor, Olivia Wilker, Mahathi Vojjala, Omar El-Shahawy, Scott E Sherman. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.1.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.